Proceedings of the Focused Research Program on Spectral Theory and Boundary Value Problems

Linear Differential Equations and Systems

Argonne National Laboratory, Argonne, Illinois 60439
operated by The University of Chicago
for the United States Department of Energy under Contract W-31-109-Eng-38
Argonne National Laboratory, with facilities in the states of Illinois and Idaho, is owned by the United States government, and operated by The University of Chicago under the provisions of a contract with the Department of Energy.

**DISCLAIMER**

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

This report has been reproduced from the best available copy.

Available from the
National Technical Information Service
NTIS Energy Distribution Center
P.O. Box 1300
Oak Ridge, TN 37831

Price: Printed Copy A08
Microfiche A01
ANL--87-26--Vol. 3

DE89 011499

ARGONNE NATIONAL LABORATORY
9700 South Cass Avenue
Argonne, Illinois 60439-4801

PROCEEDINGS OF THE FOCUSED RESEARCH PROGRAM ON
SPECTRAL THEORY AND BOUNDARY VALUE PROBLEMS

VOL. 3: LINEAR DIFFERENTIAL EQUATIONS AND SYSTEMS

Hans G. Kaper, Man Kam Kwong, and Anton Zettl, organizers

Gail W. Pieper, technical editor

Mathematics and Computer Science Division

April 1989

This work was supported in part by the Applied Mathematical Sciences subprogram of the Office
A major purpose of the Technical Information Center is to provide the broadest dissemination possible of information contained in DOE's Research and Development Reports to business, industry, the academic community, and federal, state and local governments.

Although a small portion of this report is not reproducible, it is being made available to expedite the availability of information on the research discussed herein.
Asymptotic Linearity of Solutions of Second-Order Systems of Differential Equations — Shaozhu Chen

Abstract ........................................................................................................................................ 49
1. Introduction ......................................................................................................................... 49
2. Basic Hypotheses .................................................................................................................. 51
3. Riccati Techniques and Preliminary Lemmas .................................................................. 52
4. Main Results ......................................................................................................................... 53
5. Proof of Theorem 1 ............................................................................................................... 57
6. Proof of Theorem 2 ............................................................................................................... 59
7. Proof of Theorem 3 ............................................................................................................... 60
8. Nonhomogeneous Systems .................................................................................................. 63
9. Terminal Value Problems ................................................................................................. 64
10. Nonlinear Perturbation ....................................................................................................... 64
Acknowledgment ....................................................................................................................... 66
References .................................................................................................................................... 66


Abstract ........................................................................................................................................ 69
1. Introduction ......................................................................................................................... 69
2. Main Results ......................................................................................................................... 70
References .................................................................................................................................... 75


Abstract ........................................................................................................................................ 77
1. Introduction ......................................................................................................................... 77
2. The Basic Problem .............................................................................................................. 78
3. How to Use the Code .......................................................................................................... 80
4. Details of Subprograms ...................................................................................................... 84
5. Computational Examples and Timings ............................................................................... 91
Acknowledgment ....................................................................................................................... 95
References .................................................................................................................................... 95
Appendix - Sample Driver Program for Subroutine SPDNSF .................................................. 99

$L^2$-Norm Inequalities for the Differentiation Operator on Compact Intervals — David W. Gillman, Hans G. Kaper, and Man Kam Kwong

Abstract ........................................................................................................................................ 103
1. Introduction ......................................................................................................................... 103
2. Reformulation of the Problem ........................................................................................... 104
3. The Variational Problem .................................................................................................... 107
4. The Best Constant and the Extremal Function .................................................................. 108
References .................................................................................................................................... 111
Preface

This is the third volume of a series of reports containing the proceedings of the Focused Research Program on “Spectral Theory and Boundary Value Problems,” which was held at Argonne National Laboratory during the period 1986-1987. The program was organized by the Mathematics and Computer Science (MCS) Division as part of its activities in applied analysis. Members of the organizing committee were F. V. Atkinson, H. G. Kaper (chairman), M. K. Kwong, A. M. Krall, and A. Zettl.

The objective of the program was to provide an opportunity for research and exchange of views, problems, and ideas in three main areas of investigation: (1) the theory of singular Sturm-Liouville equations, (2) the asymptotic analysis of the Titchmarsh-Weyl $m(\lambda)$-coefficient, and (3) the qualitative theory of nonlinear differential equations. The program had five full-time participants, who were joined by five more participants for periods of several months. Twenty-four mathematicians from the United States, Canada, and Europe visited for shorter periods for seminars and technical discussions. These proceedings are the permanent record of the research stimulated by the year-long program.

The MCS Division generously supported the activities of the Focused Research Program. A grant for the visitors program was provided by the Argonne Universities Association Trust Fund.

Following this preface is a list of all participants and visitors with their current affiliations and addresses. Also included is a schedule of the talks presented as part of the research program. We express our gratitude to our colleagues and especially to those who contributed manuscripts to the proceedings.

Hans G. Kaper
Man Kam Kwong
Anton Zettl
Argonne National Laboratory
Mathematics and Computer Science Division

1986-87 Focused Research Program
“Spectral Theory and Boundary Value Problems”

Participants

Full-time

F. V. Atkinson
Department of Mathematics
University of Toronto
Toronto M5S 1A1, Ontario
Canada
October 1986 - July 1987

Hans G. Kaper
Mathematics and Computer Science Div.
Argonne National Laboratory
9700 South Cass Avenue
Argonne, IL 60439-4844
September 1986 - September 1987

Allan M. Krall
Department of Mathematics
Pennsylvania State University
215 McAllister Building
University Park, PA 16802
September 1986 - May 1987

Man Kam Kwong
Mathematics and Computer Science Division
Argonne National Laboratory
Argonne, IL 60439-4844
September 1986 - September 1987

Anton Zetti
Department of Mathematical Sciences
Northern Illinois University
DeKalb, IL 60115
September 1986 - June 1987

Part-time

Alfonso Castro
Department of Mathematics
North Texas State University
Denton, TX 76203-5116
May - July 1987

C. Y. Chan
Department of Mathematics
University of Southwestern Louisiana
Lafayette, LA 70504-1010
May - July 1987

Charles T. Fulton
Department of Applied Mathematics
Florida Institute of Technology
Melbourne, FL 32901
April - June 1987

Marc Garbey
Department of Mathematics
U. de Valenciennes
Le Mont Houy
59326 Valenciennes
France
June - July 1987

Eduardo Socolovsky
Department of Mathematics
University of Pittsburgh
Pittsburgh, PA 15260
June - September 1987

Visitors

W. Allegretto
Department of Mathematics
University of Alberta
Edmonton, Alberta T6G 2G1
Canada
Dates of visit: April 28-30, 1987

Paul B. Bailey
Numerical Mathematics Division
Sandia National Laboratories
Albuquerque, NM 87185
Dates of visit: April 20-24, 1987

Chr. Bennewitz
Department of Mathematics
University of Uppsala
Sweden
Dates of visit: March 17-31, 1987

H. Benzinger
Department of Mathematics
University of Illinois
273 Altgeld Hall
Urbana, IL 61801
Dates of visit: March 16-17, 1987
R. C. Brown  
Mathematics Department  
University of Alabama  
Tuscaloosa, AL 35487-1416  
Dates of visit: April 13-18, 1987

S. Chen  
Department of Mathematics  
Shandong University  
Jinan, Shandong  
People's Republic of China  
Dates of visit: March 17-20, 1987

P. Concus  
50A-2129  
Lawrence Berkeley Laboratory  
Berkeley, CA 94720  
Dates of visit: January 30-31, 1987

L. Erbe  
Department of Mathematics  
University of Alberta  
Edmonton, Alberta T6G 2G1  
Canada  
Dates of visit: April 27-30, 1987

W. N. Everitt  
Department of Mathematics  
The University of Birmingham  
P. O. Box 363  
Birmingham B15 2TT  
United Kingdom  
Dates of visit: April 16-30, 1987

J. Goldstein  
Department of Mathematics  
Tulane University  
New Orleans, LA 70118  
Dates of visit: May 13-14, 1987

G. Halvorsen  
Institute for Energy Technology  
Department KRS, Box 40  
2007 Kjeller  
Norway  
Dates of visit: May 18-26, 1987

B. J. Harris  
Dept. of Mathematical Sciences  
Northern Illinois University  
DeKalb, IL 60115-2888  
Dates of visit: June-August, 1987

D. Hinton  
Department of Mathematics  
University of Tennessee  
Knoxville, TN 37996-1300  
Dates of visit: April 13-18, 1987

V. Jurdjevic  
Department of Mathematics  
University of Toronto  
Toronto, Ontario M5S 1A1  
Canada  
Date of visit: July 30, 1987

A. B. Mingarelli  
Department of Mathematics  
University of Ottawa  
585 King Edward  
Ottawa K1N 6N5  
Canada  

J. Neuberger  
Department of Mathematics  
North Texas State University  
P. O. Box 5116  
Denton, TX 76203-5116  
Dates of visit: April 1-4, 1987

S. Pruess  
Mathematics Department  
Colorado School of Mines  
Golden, CO 80401  
Dates of visit: June 15-19, 1987

T. Read  
Department of Mathematics  
Western Washington University  
Bellingham, Washington 98225  
Dates of visit: April 13-18, 1987

J. Ridenhour  
Department of Mathematics  
Utah State University  
Logan, UT 84322  
Dates of visit: May 13-20, 1987

Bernd Schultze  
Universitaet Gesmathochshule Essen  
Fachbereich 6, Mathematik  
Postfach 103 764  
4300 Essen 1  
West Germany  
Dates of visit: May 31-June 8, 1987

G. Sell  
Inst. for Mathematics and Its Applications  
University of Minnesota  
206 Church Street  
Minneapolis, MN 55455  
Date of visit: April 23, 1987

J. Serrin  
Department of Mathematics  
University of Minnesota  
206 Church Street  
Minneapolis, MN 55455  
Date of visit: June 18, 1987

J. K. Shaw  
Department of Mathematics  
Virginia Polytechnic Institute and State University  
Blacksburg, VA 24061  
Dates of visit: April 13-18, 1987
### Schedule of Talks

<table>
<thead>
<tr>
<th>Date</th>
<th>Speaker(s)</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>October 15</td>
<td>Allan Krall</td>
<td>&quot;Orthogonal Polynomials and Boundary Value Problems&quot;</td>
</tr>
<tr>
<td>October 22</td>
<td>Allan Krall</td>
<td>&quot;Orthogonal Polynomials and Boundary Value Problems&quot;</td>
</tr>
<tr>
<td>October 28</td>
<td>Allan Krall</td>
<td>&quot;M (λ)-Theory for Singular Hamiltonian Systems&quot;</td>
</tr>
<tr>
<td>November 7</td>
<td>Derick Atkinson</td>
<td>&quot;Prüfer Transformation for Systems of Second-Order Differential Equations&quot;</td>
</tr>
<tr>
<td>November 13</td>
<td>Derick Atkinson</td>
<td>&quot;Prüfer Transformations for Systems of Second-Order Differential Equations, II&quot;</td>
</tr>
<tr>
<td>January 14</td>
<td>Allan Krall</td>
<td>&quot;Singular Hamiltonian Systems&quot;</td>
</tr>
<tr>
<td>January 15</td>
<td>Allan Krall</td>
<td>&quot;The Titchmarsh-Weyl M-Function for Singular Hamiltonian Systems&quot;</td>
</tr>
<tr>
<td>January 16</td>
<td>Allan Krall</td>
<td>&quot;The Titchmarsh-Weyl M-Function for Singular Hamiltonian Systems, II&quot;</td>
</tr>
<tr>
<td>January 21</td>
<td>Derick Atkinson</td>
<td>&quot;Asymptotics of the Titchmarsh-Weyl M-Function for Singular Hamiltonian Systems&quot;</td>
</tr>
<tr>
<td>January 30</td>
<td>Bert Pelletier</td>
<td>&quot;The Initial Development of Dead Core in a Reaction Diffusion Equation&quot;</td>
</tr>
<tr>
<td>March 17</td>
<td>Hal Benzinger</td>
<td>&quot;Chaotic Dynamical Systems&quot;</td>
</tr>
<tr>
<td>March 18</td>
<td>Shaozhu Chen</td>
<td>&quot;Asymptotic Linearity of the Solutions of Second-order Linear Differential Equations&quot;</td>
</tr>
<tr>
<td>April 1</td>
<td>John Neuberger</td>
<td>&quot;Numerical Computation of Eigenvalues of the Schroedinger Equation&quot;</td>
</tr>
<tr>
<td>April 2</td>
<td>Michael Jolly</td>
<td>&quot;The Geometry of the Global Attractor for a Reaction-Diffusion Equation&quot;</td>
</tr>
<tr>
<td>April 15</td>
<td>Allan Krall</td>
<td>&quot;Characterization of Singular Boundary Conditions&quot;</td>
</tr>
<tr>
<td>April 15</td>
<td>Tony Zettl</td>
<td>&quot;Norm Inequalities for Differential and Difference Operators&quot;</td>
</tr>
<tr>
<td>April 16</td>
<td>Don Hinton</td>
<td>&quot;One Variable Weighted Interpolation Inequalities&quot;</td>
</tr>
<tr>
<td>April 16</td>
<td>Ken Shaw</td>
<td>&quot;Extensions of Levinson's Theorem to Dirac Systems&quot;</td>
</tr>
<tr>
<td>April 17</td>
<td>Tom Read</td>
<td>&quot;Sturm-Liouville Problems with Large Leading Coefficients&quot;</td>
</tr>
</tbody>
</table>
April 20  Paul Bailey, "Computation of Eigenvalues and Eigenfunctions of Sturm-Liouville Equations using SLEIGN"

April 21  Paul Bailey, "Computation of Eigenvalues and Eigenfunctions of Sturm-Liouville Equations using SLEIGN, II"

April 22  Norrie Everitt, "The Laplace Tidal Wave Equation"

April 23  George Sell, "The Principle of Spatial Averaging and Inertial Manifolds"

April 23  Paul Bailey, "Computation of Eigenvalues and Eigenfunctions of Sturm-Liouville Equations using SLEIGN, III"

April 28  Lynn Erbe, "Oscillation Theory for Systems of Second-Order Differential Equations"

April 29  Walter Allegretto, "Spectral Analysis of Second-Order Boundary Problems with Indefinite Weight Functions"

April 30  Charles Fulton, "Asymptotics of $m(\lambda)$ for Singular Potentials"

May 14  Jerry Goldstein, "Recent Developments in Thomas-Fermi Theory"

May 15  Charles Fulton, "Singular Hamiltonian Systems"

May 18  Jerry Ridenhour, "Zeros of Solutions of n-th Order Differential Equations"

May 20  Charles Fulton, "The Bessel-squared Operator in the lim-2, lim-3, and lim-4 Cases"

May 21  Gotskalk Halvorsen, "Oscillation Results for Second-Order Equations"

May 22  Derick Atkinson, "Estimation of $m(\lambda)$ in a Case with an Oscillating Leading Coefficient"

May 27  Hans Kaper, "A Non-oscillation Theorem for an Emden-Fowler Equation"

June 1  C.Y. Chan, "A Generalization of the Thomas-Fermi Equation"

June 3  Bernd Schultz, "Spectral Properties of Nonselfadjoint Differential Operators"

June 5  Alfonso Castro, "Superlinear Boundary Value Problems"

June 5  Man Kam Kwong, "Concavity of Solutions of Certain Emden-Fowler Equations"

June 15  Charles Fulton, "Convergence of Spectral Functions"

June 16  Bernie Matkowsky, "Introduction to Bifurcation Theory"

June 18  James Serrin, "Asymptotics of the Emden-Fowler Equation"

June 19  Steve Pruess, "SPDNSF: A Code to Compute the Spectral Density Function"

June 25  Bernie Matkowsky, "Stability Analysis and Bifurcation Theory"

July 17  Marc Garbey, "A Quasilinear Prabolic-hyperbolic Singular Perturbation Problem"

July 30  Val Jurdjevic, "Differential Equations of Control Theory"

July 31  Bernie Harris, "Asymptotics of the Titchmarsh-Weyl $m(\lambda)$-coefficient"
OSCIILATION OF SUPERLINEAR MATRIX DIFFERENTIAL EQUATIONS

Calvin D. Ahlbrandt
Department of Mathematics
University of Missouri-Columbia
Columbia, MO 65201

Jerry Ridenhour and Russell C. Thompson
Department of Mathematics
Utah State University
Logan, UT 84321

Abstract

The main theorems extend to matrix differential equations Atkinson’s classic theorem giving necessary and sufficient conditions for the oscillation of superlinear second-order scalar differential equations. The theorems improve recent results of Kura and of Butler and Erbe by removing the very restrictive hypothesis that solutions be symmetric.

1. Introduction

In his classic paper [1955], Atkinson showed that when \( f (t) \) is positive and continuous for \( t \) in \( [0, \infty) \), a necessary and sufficient condition for the second-order superlinear scalar equation

\[
y'' + f(t)y^{2n+1} = 0
\]

(1)

to be oscillatory is that \( \int_0^\infty t f(t)dt = \infty \). The papers of Kura [1982] and Butler and Erbe [1985] extend Atkinson’s theorem to the case of superlinear matrix equations; however, their theorems apply only to symmetric matrix solutions. This symmetry assumption is very restrictive, as noted at the end of Section 2. The main results of this paper also extend Atkinson’s theorem to superlinear matrix equations with the principal advantage of our theorems being that they apply to solutions that may or may not be symmetric.

Atkinson’s paper and the corresponding work of Belohorec [1961] on sublinear equations provided the impetus for a very large body of research on the oscillation theory of nonlinear scalar differential equations. The survey article by Kartsatos [1977] contains over 300 such references, some of which are themselves survey articles containing additional references. In the past ten years, the oscillation theory for systems of ordinary differential equations, both linear and nonlinear, has received considerable attention. To give the reader a brief and very incomplete sampling of the research in these directions, we cite Kartsatos and Walters [1980], Kreith [1970], Kura [1982], Noussair and Swanson [1971], and Tomastik [1968] as indicative of the work done on nonlinear systems and Ahmad [1983], Ahmad and Lazer [1978], Ahmad and Travis [1978],
2. Statement of Results

In passing from the scalar equation (1) to a superlinear matrix differential equation, there are many ways to write the superlinear matrix term corresponding to the term $f(t)y^{2n+1}$ in equation (1). Because of the noncommutativity of matrix multiplication, the analysis (as we shall see below) may vary depending on the way in which this nonlinear term is expressed. To begin, we suppose throughout that $F(t)$ is a continuous $n \times m$ complex matrix-valued function on $[0, \infty)$ such that $F(t)$ is Hermitian and positive definite for each $t$ in $[0, \infty)$. We introduce an equation with a cubic nonlinearity in two forms:

$$Y^{(2)} + (YF(t)Y^*)Y = 0$$

(2)

and

$$Y^{(2)} + (Y^*F(t)Y)Y = 0.$$  (3)

Here, as elsewhere, $Y^*$ denotes the adjoint of $Y$.

For a positive integer $n$, let $H_n$ represent any product of the form

$$H_n = H_n(Y) = Z_1Z_2...Z_n,$$

where, for each $i = 1, 2, ..., n$, either $Z_i = Y$ or $Z_i = Y^*$ (that is, $H_n$ is a general matrix analogue of the scalar power $y^n$). To unify notation, we let $H_0$ be the $m \times m$ identity matrix. The matrix equations analogous to equation (1) are then

$$Y^{(2)} + (YH_nF(t)H_n^*Y^*)Y = 0$$

(4)

and

$$Y^{(2)} + (Y^*H_nF(t)H_n^*Y)Y = 0.$$  (5)

Although (2) and (3) are special cases (with $n = 0$) of (4) and (5), they are considered separately since the proofs amount to first establishing results for (2) and (3) and then extending to (4) and (5). It will be important that the expressions in parentheses in (2)-(5) are Hermitian; however, these expressions could be commuted with $Y$ in the nonlinear term, and analogous results would follow.

We consider only those $m \times m$ matrix solutions $Y(t)$ that extend to infinity (i.e., exist on some ray of the form $[a, \infty)$, $a \geq 0$). Following the lead of others (e.g., Kreith [1970], Noussair and Swanson [1971], and Tomasik [1968]), we call such a solution prepared (or self-conjoined) if $Y^*(t)Y(t) - Y^*(t)Y(t)$ is the zero matrix for all $t$. Differentiation shows that, for any solution $Y(t)$ of one of equations (2)-(5), the matrix $Y^*(t)Y(t) - Y^*(t)Y(t)$ is necessarily constant; hence,
$Y(t)$ is prepared when $Y^*(t_0)Y(t_0) - Y^*(t_0)Y(t_0) = 0$ for some $t_0$. By a zero of a matrix solution $Y$, we mean a point $t$ where the $m \times m$ matrix $Y(t)$ is singular. A matrix solution is called oscillatory if it extends to infinity and has arbitrarily large zeros. Differential equations of the form (2)-(5) are said to be oscillatory if all prepared solutions that extend to infinity are oscillatory. Noussair and Swanson [1971] have produced an example showing that, without the hypothesis of preparedness, it is possible to have a mixture of oscillatory and non-oscillatory solutions.

We denote the maximum and minimum eigenvalues of a Hermitian matrix $H$ by $\lambda_{\max}(H)$ and $\lambda_{\min}(H)$, respectively. Our main result follows.

**THEOREM 1.** Suppose $F(t)$ is a continuous $m \times m$ complex matrix-valued function on $[0, \infty)$ such that $F(t)$ is Hermitian and positive definite for each $t$ in $[0, \infty)$. Then equation (4) is oscillatory provided that

$$\int_0^\infty \lambda_{\max}(F(t)) \, dt = \infty$$

(6)

holds, and equation (5) is oscillatory provided that

$$\int_0^\infty \lambda_{\min}(F(t)) \, dt = \infty$$

(7)

holds.

We conjecture that equation (5) is also oscillatory with the weaker hypothesis (6) replacing (7); however, the proof that we give below does not yield this result. We have undertaken a numerical study of this question in which, for the same $F(t)$, solutions of (2) and solutions of (3) satisfying the same initial conditions at some point $t_0$ were compared to each other. In all examples, solutions of (3) either oscillated significantly more quickly or oscillated as quickly as did solutions of (2). These examples seem to indicate that (3) is also oscillatory with the weaker assumption. Of course, in the scalar case, both (6) and (7) reduce to Atkinson's condition.

Theorem 1 provides sufficient conditions for equations (4) and (5) to be oscillatory. As noted in a remark of Kura [1982: p. 225], the fact that (6) is a necessary condition for oscillation follows easily from a result of Kartsatos and Walters [1980]. We summarize that fact in a theorem.

**THEOREM 2.** If $F(t)$ is a continuous $m \times m$ complex matrix-valued function on $[0, \infty)$ which is also Hermitian and positive definite for each $t$ in $[0, \infty)$, then (6) is a necessary condition for equations (4) and (5) to be oscillatory.

Combining Theorems 1 and 2, we see that (6) is both necessary and sufficient for equation (4) to be oscillatory.
We remark on the consequences of the assumption in [Butler and Erbe 1985] and [Kura 1982] that the solutions \( Y(t) \) are both symmetric and prepared. For symmetric solutions, equations (4) and (5) both reduce to

\[
Y'' + Y^n F(t) Y^{n+1} = 0. \tag{8}
\]

At any point where a symmetric solution \( Y(t) \) is nonsingular, it follows from taking adjoints in (8) and multiplying appropriately by \( Y^{-1}(t) \) that \( F(t) \) commutes with \( Y(t) \). The additional assumption that \( Y(t) \) is prepared implies that \( Y(t) \) commutes with \( Y'(t) \). Dieudonné [1974] has shown that continuously differentiable matrices that commute with their derivatives must be functionally commutative (i.e., \( Y(t)Y(s) = Y(s)Y(t) \) for all \( t \) and \( s \)). Furthermore, as proved by Goff [1981], analytic Hermitian matrices that commute with their derivatives must be functionally commutative.

Hence, at points where a solution \( Y(t) \) satisfies \( \det(Y(t)) \neq 0 \), the assumption that \( Y(t) \) is both symmetric and prepared implies that \( T(t) \) commutes both with \( F(t) \) and with \( Y'(t) \). This is clearly a very restrictive assumption. On the other hand, our results apply when \( Y(t) \) is any prepared solution, a condition that can be guaranteed by correct assignment of initial conditions at a single point \( t_0 \).

3. Proofs of the Theorems

As a tool in the proof of Theorem 1, we will need the following lemma from matrix theory.

**Lemma 1.** If \( P \) is a positive definite \( m \times m \) Hermitian matrix and \( Q \) is any \( m \times m \) matrix, then the following inequalities hold:

\[
\lambda_{\max}(Q^*PQ) \geq \lambda_{\max}(Q^*Q)\lambda_{\min}(P), \tag{9}
\]

\[
\lambda_{\max}(Q^*PQ) \geq \lambda_{\min}(Q^*Q)\lambda_{\max}(P), \tag{10}
\]

and

\[
\lambda_{\min}(Q^*PQ) \geq \lambda_{\min}(Q^*Q)\lambda_{\min}(P). \tag{11}
\]

**Proof.** These inequalities follow from Rayleigh's principle. The least direct is (10), so we prove (10) and leave the proofs of (9) and (11) to the reader. Let \( \| \cdot \| \) denote the Euclidean norm, and let \( \langle \cdot, \cdot \rangle \) denote the inner product in \( \mathbb{R}^n \). If \( Q \) is singular, then (10) follows immediately since the righthand side is then zero; hence, we assume \( Q \) is nonsingular.

Choose a vector \( y \) with \( \| y \| = 1 \) so that

\[
\langle Py, y \rangle = \lambda_{\max}(P).
\]

Since \( Q \) is nonsingular, we choose \( x \) with \( Qx = y \). Then
\[ \lambda_{\text{max}}(Q^*PQ) \geq \frac{\langle P^*PQx,x \rangle}{\langle x,x \rangle} \]
\[ = \frac{\langle PQx,Qx \rangle}{\langle x,x \rangle} \]
\[ = \frac{\lambda_{\text{max}}(P)}{\langle x,x \rangle} \]
\[ = \lambda_{\text{max}}(P) \frac{\langle Ax,Ax \rangle}{\langle x,x \rangle} \]
\[ = \lambda_{\text{max}}(P) \frac{\langle A^*Ax,x \rangle}{\langle x,x \rangle} \]
\[ \geq \lambda_{\text{max}}(P) \lambda_{\text{min}}(A^*A). \]

This completes the proof.

**Proof of Theorem 1.** As a first step, we show that equation (2) is oscillatory provided that (6) holds. Suppose to the contrary that (6) holds but there is a prepared solution \( Y \) of (2) that is non-singular on some interval \([a,\infty)\), where \( a \geq 0 \). We make the Riccati substitution

\[ R(t) = Y(t)Y^{-1}(t), \quad t \in [a,\infty). \quad (12) \]

The \( R(t) \) defined by (12) is Hermitian on \([a,\infty)\) since \( Y(t) \) is prepared. We now show that \( R(t) \) is eventually positive definite. Differentiating and using (2), we see that

\[ R' = -YFY^* - R^2 \]

and consequently

\[ (R^{-1})' = R^{-1}YFY^*R^{-1} + I \]

are valid on \([a,\infty)\). Hence, all eigenvalues of \( R^{-1}(t) \) tend to infinity as \( t \to \infty \), and \( R(t) \) is eventually positive definite as claimed.

Solving for \( tF(t) \) in (2) yields

\[ tF(t) = -tY^{-1}Y'Y^{-1}Y'Y^{-1}, \quad t \in [a,\infty). \quad (13) \]

where in (13), as below, the argument \( t \) has been suppressed in several places to shorten the notation. We integrate both sides of (13) using integration by parts on the righthand side and the preparedness of \( Y \) to obtain
\[
\int_a^t x F(x) \, dx = -x Y^{-1} Y^{-1} Y^* + \int_a^t Y^{-1} Y^{-1} x Y^* \, dx - 3 \int_a^t x Y^{-1} R^2 Y^* \, dx.
\]  
(14)

Since \(-\frac{1}{2} Y^{-1} Y^* \) is an antiderivative of \(Y^{-1} Y^{-1} Y^* \), one can rearrange (14) to find that
\[
\int_a^t x F(x) \, dx = -\frac{1}{2}(Y^* Y)^{-1} - 3 \int_a^t x Y^{-1} R^2 Y^* \, dx - \int_a^t x F(x) \, dx + K
\]
for some constant Hermitian matrix \(K\).

As noted by Tomastik [1968: p. 1428], when \(A(t)\) is positive semidefinite, the conditions
\[
\lambda_{\text{max}} \left[ \int_a^\infty A(s) \, ds \right] = \infty \quad \text{and} \quad \int_a^\infty \lambda_{\text{max}} [A(s)] \, ds = \infty
\]
are equivalent. Using this fact and (6), we see that the minimum eigenvalue of the righthand side of (15) approaches \(-\infty\) as \(t \to \infty\). This contradicts \(R(t)\) being eventually positive definite and completes this part of the proof.

As a second step in the proof of Theorem 1, we show that (3) is oscillatory provided that (7) holds. Suppose not. Let (7) hold, and let \(Y(t)\) be a prepared solution of (3) that is nonsingular on \([a, \infty)\). We again let \(R(t) = Y(t) Y^{-1}(t)\) for \(t \geq a\) and use the same argument as above to show that \(R(t)\) is eventually positive definite. However, the remainder of the proof is different since isolating \(t F(t)\) and integrating by parts does not easily lead to a nice analogue of equation (15). Alternatively, we solve equation (3) for \(t Y^{-1} Y^{-1} Y^* \) which results in
\[
\int_a^t x \, dx = -t Y^{-1} Y^* F(t) Y Y^* \quad \text{valid for } t \in [a, \infty).
\]
(16)

We integrate (16) to obtain
\[
\int_a^t x \, dx = -\frac{1}{2}(Y^* Y)^{-1} - 3 \int_a^t x Y^{-1} R^2 Y^* \, dx - \int_a^t x Y^{-1} F(x) Y Y^* \, dx + K
\]
(17)

as the analogue to (15) above. From (9), it follows that
\[
\lambda_{\text{max}} (Y^{-1} Y^* F(x) Y Y^* \quad \text{valid for } t \in [a, \infty).
\]
(16)

Hence, letting \(t \to \infty\), we see that (7) and (17) contradict that \(R(t)\) is eventually positive definite. This completes the proof that equation (3) is oscillatory when (7) holds.

Suppose again that (6) holds. We now use the fact that (2) is oscillatory to prove that (4) is also oscillatory. Suppose not, and let \(Y_0\) be a prepared solution of (4) that is nonsingular on \([a, \infty)\). Again let \(R(t) = Y_0(t) Y_0^{-1}(t)\) for \(t \geq a\). The same calculations as before show that \(R(t)\) is
eventually positive definite. Also, since $Y_0$ is prepared,

$$
(Y_0^* Y_0)' = 2Y_0^* Y_0 = 2Y_0^* Y_0 Y_0^{-1} Y_0 = 2Y_0^* R Y_0.
$$

Hence, the eigenvalues of $Y_0^* Y_0$ are eventually increasing and therefore eventually bounded below by a positive constant. Moreover, $Y^* Y$ and $Y Y^*$ have the same eigenvalues.

Let $F_0(t) = H_n(Y_0) F(t) H_n^*(Y_0)$. Then the equation

$$
Y'' + (Y F_0(t) Y^*) Y = 0 \tag{18}
$$

has $Y_0$ as a prepared and eventually nonsingular solution. By (10) and (11), we see

$$
\lambda_{\max}(F_0(t)) \geq \lambda_{\min}(H_n H_n^*) \lambda_{\max}(F(t)) \geq (\lambda_{\min}(Y_0 Y_0^*))^{\nu} \lambda_{\max}(F(t)).
$$

Since the eigenvalues of $Y_0 Y_0^*$ are eventually bounded below by a positive constant, it follows that

$$
\int_0^\infty \lambda_{\max}(F_0(t)) \, dt = \infty.
$$

But this contradicts what we have already proved, since (15) is an equation of the same form as (2) for which (6) holds, and yet there is an eventually nonsingular prepared solution.

The proof that equation (5) is oscillatory when (7) holds follows in similar fashion. This proves Theorem 1.

**Proof of Theorem 2.** To show that (6) is a necessary condition for (4) to be oscillatory, we assume

$$
\int_0^\infty \lambda_{\max}(F(t)) \, dt < \infty \tag{19}
$$

and prove the existence of a (non-oscillatory) prepared solution of $Y(t)$ of (4) that converges to $I$, the $m \times m$ identity matrix, as $t \to \infty$. As noted by Kura [1982], this follows easily from Theorem 1 of Kartsatos and Walters [1980]. To match the notation of Kartsatos and Walters, let

$$
P(t, X) = X H_n(X) F(t) H_n^*(X) X^* X,
$$

$$
Q(t) = 0, \quad V(t) = 0,
$$

and

$$
q(t, \| Y \|) = \lambda_{\max}(F(t)) \| Y \|^{2s+1}.
$$

Since (19) holds, Theorem 1 of Kartsatos and Walters [1980] applies directly with $\lambda = 1$ to show that (4) has a solution $Y(t)$ with $\lim_{t \to \infty} Y(t) = I$. Also, the proof of Theorem 1 of Kartsatos and
Walters gives that the derivative of the solution $Y(t)$ satisfies

$$\| Y'(t) \| \leq \int_{t}^{\infty} s \lambda_{\max}(F(s)) \, ds,$$

where $\| Y'(t) \|$ denotes the matrix norm induced by the Euclidean vector norm. Hence, it follows that

$$\lim_{t \to \infty} [Y^*(t)Y'(t) - Y^{*'}(t)Y(t)] = 0$$

and $Y(t)$ is prepared.

This completes the proof in the case of equation (4). The proof that condition (6) is necessary for equation (5) to be oscillatory follows in the same way.

4. Note

Most of the results in this paper predate the appearance of the papers of Kura [1982] and Butler and Erbe [1985]. An unpublished paper written in 1971 by Ahlbrandt contained Theorem 1 as it applies to Equation (2). Lemma 1 and Theorem 1, exactly as they appear in this paper, were presented by Ridenhour in a talk given at the Eighth Annual Midwest Differential Equations Conference held October 18-19, 1979, at Oklahoma State University.

References


SPECTRAL ANALYSIS OF SECOND-ORDER BOUNDARY PROBLEMS WITH INDEFINITE WEIGHT FUNCTIONS

W. Allegretto
Department of Mathematics
University of Alberta
Edmonton, Canada T6G 2G1

Abstract

We study the qualitative behavior of solutions and the spectrum of boundary problems of the form $Au = \lambda Bu$, where $A$ is an ordinary/partial differential operator and $B$ is a multiplication operator by a generally (sign) indefinite weight function.

Let $\Omega$ be a bounded domain in $\mathbb{R}^n$, and consider the Dirichlet problem

$$Lu = \lambda wu$$

in $\Omega$ with $u \in H^{1,2}(\Omega)$ and $L$ either $-\Delta + q$ ($n > 1$) or $-\frac{d}{dx}(\rho \frac{d}{dx}) + q$ ($n = 1$). In (1) we assume that $w$ changes sign in $\Omega$ and that $L$ has some negative spectrum. For technical simplicity of presentation we assume that $q, w \in L^\infty(\Omega)$ and that $\mu(x)w(x) = 0$. For any subset $S \subset H^{1,2}(\Omega)$ we let $|S|$ denote the largest number of linearly independent vectors in $S$, while $\#S^+$ (resp. $\#S^-$) denotes the largest number of vectors in $S$ such that $(wu, u) = \delta_{ij}$ (resp. $-\delta_{ij}$).

Since $L$ and the operator generated by $w$ have both positive and negative spectrum, problem (1.1) is called indefinite (or nondefinite). It has been known for some time that such problems may admit complex eigenvalues $\lambda$. We recall that $u$ is termed a ghost state iff $(wu, u) = 0$. Clearly any complex $\lambda$ must associated with a ghost state. For a survey of boundary problems of this type we refer to [Mingarelli 1986] from which much of our terminology is also taken.

The results we present here are selected from those in [Allegretto and Mingarelli 1987], where detailed proofs, more results, and references can be found.

We are interested in locating the eigenvalues $\lambda$ of (1), ensuring sufficient conditions for the existence of complex $\lambda$ and, for $n = 1$, estimating the Haupt and Richardson indices associated with (1).

Our first result is given below.

*Research supported by National Science and Engineering Research Council of Canada.
THEOREM 1. (a) Equation (1) admits two positive eigenfunctions (resp. one) if \( L - \xi w > (\text{resp.} \geq) 0 \) for some real \( \xi \).

(b) If (1) admits a positive eigenfunction \( u \) with eigenvalue \( \lambda \), then

(i) \( \lambda \) is simple (i.e., \( \dim(\text{eigenspace}) = 1 \));

(ii) there are no ghost states except, possibly, \( u \) itself;

(iii) if \( \mu \) is any other eigenvalue with eigenvector \( v \), then \( (\mu - \lambda)(wv, v) > 0 \).

(c) Let \( S = \{ u \mid \lambda(\lambda, u) \leq 0, \ u \text{ eigenfunction of (1) with eigenvalue } \lambda \} \). Then \( |S| \) is finite.

(d) For any fixed \( \lambda \), let \( S = \{ u \mid Lu = \lambda wu \} \). Then \( |S| \) is finite.

(e) \( L, w \) do not have nontrivial common null space.

(f) If \( L^{-1} \) exists and \( \mu, \lambda \neq 0 \), then root vectors \( u, v \) corresponding to eigenvalues \( \lambda, \mu (\lambda \neq \bar{\mu}) \) satisfy \( (wv, v) = 0 \).

(g) Eigenvectors corresponding to different eigenvalues of (2.2) are linearly independent.

Theorem 1 indicates that positive eigenfunctions cannot be used to investigate the spectrum of truly indefinite problems. Indeed, such problems may be characterized by the absence of (positive) ground states. This is in sharp contrast to the definite cases (see, e.g., Manes and Micheletti [1973] Reed and Simon [1978], and Hess and Kato [1980]), where there are eigenvalue(s) with positive eigenvector(s) at the onset of the spectrum.

Select a constant \( k_0 > 0 \) such that \( L_1 = L + k_0 > 0 \). It will be useful to consider in the sequel the problem

\[
L_1 u = \lambda wu + ku
\]

with eigenpair \((\lambda, k)\). For fixed \( \lambda \in \mathbb{R} \) we let \( k_j = k_j(\lambda) \) denote the \( j \)th eigenvalue of (2) as given by the Courant min./max. principle. Observe that, in particular, the real eigenvalues of the original problem (1) correspond to the set

\[
U_{j=1}^\infty [k_j^{-1}(k_0)] .
\]

We first show that the set of \( \lambda \)'s associated with ghost states must be bounded uniformly for bounded \( k \). Throughout our discussion while \( \lambda \) may be complex, \( k \) will always be taken real unless otherwise specified.
THEOREM 2. If $(\lambda, k)$ is associated with an eigenvector $u$ such that $\lambda(wu,u) \leq 0$ and $L_1u = \lambda wu + ku$ with $|k| < \beta$, then there exists a constant $\gamma = \gamma(\beta) > 0$ such that $|\lambda| < \gamma$.

We also recall the following properties of equation (2).

THEOREM 3. Let $y = k_j(\lambda)$ denote the real eigencurves of (2) as given by the Courant min./max. principle.

(a) $k_j(\lambda)$ is Lipschitz continuous (indeed, $|k_j(\lambda) - k_j(\mu)| \leq |\lambda - \mu| |w|_{L^\infty}$);

(b) $k_j(\lambda) \to +\infty$ as $j \to \infty$, uniformly for $\lambda$ in compact sets;

(c) $y = k_1(\lambda)$ is concave;

(d) $k_j(\lambda) \to -\infty$ as $|\lambda| \to \infty$;

(c) Problem (1) has infinitely many eigenvalues $\{\lambda_1^\pm\}$, $\lambda_1^+ \to \pm\infty$.

As a consequence, we have the following corollary.

COROLLARY 4. For any real $k$ there exists a countable set $Z$ such that $L_1 - k - tw$ does not have ordinary eigenvalue zero for any complex $\tau \notin Z$.

We now proceed to estimate the eigenvalues of (1). For this it is convenient to introduce the following notation: Let $(\lambda, k)$ be an eigenpair for (2). We decompose the eigenspace associated with the pair into the following sets, the last three of which are in particular the positive, neutral, and negative parts (see [Bognar 1974: Chapter 1]):

$S_0 = \{v \mid (v,v) = 1\}$

$S_1 = \{v \mid (wv, v) > 0\}$

$S_2 = \{v \mid (wv, v) = 0, v \text{ nontrivial}\}$

$S_3 = \{v \mid (wv, v) < 0\}$.

Observe that if $S_1 \neq \phi$ and $S_3 \neq \phi$, then $S_2 \neq \phi$: see, e.g., Bognar [1974: Chapter 1]. Let $n_i = |S_i|$ for $i = 0, 2$, $n_1 = \#S_1^+$, $n_3 = \#S_3^-$. We also observe the following relationships between the $n_i$. A short proof is given for convenience.

LEMMA 5. $n_i$ is finite for $i = 0, \ldots, 3$. Assume $S_2 = \phi$. Then either $S_1 = \phi$ or $S_3 = \phi$. Furthermore, in this case, if $S_1$ (resp. $S_3$) $\neq \phi$, then $n_1$ (resp. $n_3$) $= n_0$. 

13
Proof. That $n_0, n_2$ are finite is an immediate consequence of the Sobolev compact embedding theorems. We show $n_1, n_3$ are finite by showing $n_1 \leq n_0$ and $n_3 \leq n_0$. Indeed, let $\{u_i\}_i$ be given, $(wu_i, u_j) = \delta_{ij}$. We orthonormalize $\{u_i\}$ in $L^2$ and thus create an orthonormal set $\{v_i\}_i$, whence $n_1 \leq n_0$. The proof that $n_3 \leq n_0$ is identical. Finally, assume $S_2 = \phi$, and without loss of generality that $S_1 = \phi$. Given $\{v_i\}_j$ eigenvectors such that $(v_i, v_j) = \delta_{ij}$, we now orthonormalize with respect to $(w, \cdot)$ to obtain the set $\{u_i\}_j$ such that $(wu_i, u_j) = 0$ if $i \neq j$. Since $S_2 = S_3 = \phi$, we conclude $\{u_i\} \subset S_1$, i.e., $(wu_i, u_i) > 0$, whence $n_0 \leq n_1$.

If $S_2 = \phi$, we thus define the multiplicity of $(\lambda, k)$ to be $n_0$. If $S_2 \neq \phi$, then the multiplicity of $(\lambda, k)$ is given by the four vector $(n_0, n_1, n_2, n_3)$. Observe that in such a case there are $n_2$ linearly independent ghost states associated with $(\lambda, k)$. The same terminology will, in particular, be used for complex $\lambda$.

Again motivated by Bognar [1974], we term $(\lambda, k)$ a positive (resp. negative) eigenpair iff $S_2 = S_3 = \phi$ (resp. $S_1 = S_2 = \phi$). An associated eigenvector $u$ will be termed positive (resp. negative). Note that this definition does not imply any actual sign conditions on $(\lambda, k)$ or the function $u$.

Fix $k$ and let $(\lambda, k)$ be a positive eigenpair of multiplicity $m$. We order the eigenvectors associated with $(\lambda, k)$ by the procedure of Lemma 5. Specifically, we specify and fix $\{v_1, \ldots, v_m\}$, orthonormal in $L^2$, by the Courant min./max. principle. We then construct $\{u_1, \ldots, u_m\}$ by the Gram-Schmidt procedure of the lemma. Assume that, for a fixed $k$, the positive eigenpairs $(\mu, k)$ with $\mu < \lambda$ have been designated as the first $j$th pairs. Then $\lambda$ is termed the $i + j$th positive eigenvalue and is denoted by $\lambda_{i+j}$ iff it is associated with $u_i$. An identical procedure is followed for negative eigenpairs, while the precise ordering of multiple (real or complex) eigenvalues with associated ghost states is left to convenience. Observe that $\lambda_i$ is well defined, since it can be shown that $(\lambda, k)$ is negative (resp. positive)—for fixed $k$—if $\lambda \ll \lambda (\text{resp. } \gg) 0$. We thus construct for each $k$ two sequences $\{u_\lambda^\pm\}_i$ of positive (resp. negative) eigenvectors with associated eigenvalues $\lambda_i^\pm \to \pm \infty$ as $i \to \infty$, and such that $(wu_i^+, u_j^-) = \pm \delta_{ij}$. These sequences constitute possibly different sets from those given in Theorem 3(e). Finally, observe that in determining that $\lambda$ is the $i$th positive eigenvalue, we do not count the negative eigenvalues nor those associated with ghost states.

Henceforth, $(\lambda, k)$ will be termed the $(i,j)$th positive eigenpair iff $(\lambda, k) = (\lambda_i, k_j(\lambda_i))$ where $\lambda_i$ is the $i$th positive eigenvalue of (2). In view of the symmetry present in the results, we shall explicitly consider only positive eigenpairs.

Fix $k$, and let $G$ denote any linear space generated by a linearly independent set of root vectors $\{g_i\}_m=1$ such that $(wg_i, g_j) = ((L_1 - k)g_i, g_j) = 0$ for all $i, j$. Note that any such $G$ is finite dimensional, and we let $M$ denote the set of possible $G$. If no such vectors exist, we set $G = \{0\}$ and note that $|G|$ is then 0. Observe that the set of all non-real root vectors may not be a member of $M$. For example, if $\{u, \bar{u}\}$ are a non-real pair of eigenvalues of (1) with eigenvectors
If \( u, \bar{u} \), then \( sp\{u, \bar{u}\} \in M \) iff \( (wu, \bar{u}) = 0 \). We also observe that if \( g \in G \in M \) and \( h \) is a positive eigenvector, then \( (wg, h) = 0 \) by Theorem 1(f).

We now state our first estimate.

**Theorem 6.** Let \((\lambda, k)\) be the \((i, j)\)th positive eigenpair. Then

\[
\sup_{G \in M} \left[ \sup_{V, \dim V = j-1} \left[ \inf_{u \in V^\perp} \frac{(L_1-k)u, u}{(wu, u)} \right] \right] \leq \lambda.
\]

\[
\leq \inf_{G \in M} \left[ \sup_{S, \dim S = i} \left[ \inf_{u \in S^\perp} \frac{(L_1-k)u, u}{(wu, u)} \right] \right],
\]

where \( S, V \) are subspaces in \( L^2 \) with \( \dim(S) = j-1 \), \( \dim V = i-1 + \|G\|; S^\perp, V^\perp \) are understood in the \( L^2 \) sense; the member of each subspace or set are allowed to be complex; and \( u \in H^{1,2}(\Omega) \).

**Theorem 7.** Let \( k_0 \) be as above, and assume that for any \( k \leq k_0 \), all eigenvalues of \( L_1u - ku = \lambda wu \) have one-dimensional eigenspaces and root spaces of dimension \( \leq 2 \). Let \( L_1u - k_0u = \lambda wu \) with \( \lambda > 0 \) sufficiently large. Then

\[
\lambda = \sup_{G \in M} \left[ \sup_{G \in M} \left[ \inf_{u \in V^\perp} \frac{(L_1u - k_0u, u)}{(wu, u)} \right] \right],
\]

where \( \dim(V) \) depends on \( \lambda \). Conversely, for \( \dim V \) sufficiently large, the right-hand side of (4) gives an eigenvalue of \( L_1u - k_0u = \lambda wu \).

We now remark on the relation of our results to previously known estimates for the cases of "no negative squares" (see, e.g., Textorius [1974], Phillips [1970], and Curgus and Langer [1984]). An example is known (see Atkinson and Jabon [1984]) where the (1,2) and (2,2) positive eigenpairs exist and there are no ghost states. Consequently, even if \( M \) consists only of \( \{0\} \), we cannot conclude \( i = j \) for all \( i, j \). In the same example, it can be shown that setting \( \dim S = 0 \) in (3) (i.e., \( S = \{0\} \)) yields \(-\infty\), while \( \dim S = 1 \) gives the (2,2) eigenpair. The details of these remarks are given by Allegretto and Mingarelli [1987]. What this shows is that, even if no ghost states are present, we cannot hope to precisely locate all eigenpairs \((i, j)\), for fixed \( k \), by a min./max. argument of the above type, and the preceding results cannot be strengthened in this direction.

These results furnish us with the means of constructing sufficient conditions for the existence of complex eigenvalues of (1). As an example, we give the following corollary.
COROLLARY 8. Let \( \Omega = (-a, a) \times (0, a) \) in \( \mathbb{R}^2 \), and assume \( w(x,y) = -w(-x,y) \). If \( k \in \left( \frac{5}{4} \frac{\pi^2}{a^2}, \frac{2\pi^2}{a^2} \right) \), then \( Lu = (-\Delta - k)u = \lambda w(u) \) has exactly two complex (purely imaginary) eigenvalues. If \( k = \frac{5}{4} \frac{\pi^2}{a^2} \) or \( k = \frac{2\pi^2}{a^2} \), then this operator has a real ghost state. For \( k \) in the given interval, the min./max. equation (3) holds unchanged for positive eigenpairs, with smallest trial space \( S \) of dimension \( 2, S^\perp \cap G = \{0\} \).

We now pass to considerations of (1) for \( n = 1 \) but with more general coefficients. We also point out that it is inconvenient for what follows to term \( (\lambda, k) \) positive iff \( (w, u) > 0 \). Instead we shall henceforth term \( (\lambda, k) \) positive iff \( \lambda > 0 \) unless otherwise specified. Furthermore, in this section we shall assume the more general conditions: \( 1/p, q, w \in L(a,b), p(x) > 0 \) a.e. on \( (a,b) \) and \( w \) not a.e. zero. Note that here (unlike Section 2) we allow \( p \) to vanish on some sets of measure zero. If the coefficients are regular, many of the results follow immediately from above. We thus now consider

\[
y = -(p(x)y)' + q(x)y = \lambda w(x)y.
\]

The Richardson index, \( n_R \), of (5), \( (\lambda > 0) \), is defined as that smallest non-negative integer for which (5) has no real eigenfunctions with precisely \( n \) zeros in \( (a,b) \) for \( n < n_R \), whereas for \( n \geq n_R \), there is at least one real eigenfunction having exactly \( n \) zeros in \( (a,b) \). (For negative eigenvalues, the Richardson index, \( n_R \), is defined accordingly.)

The Haupt index, \( n_H \), of (5), \( (\lambda > 0) \), is defined as that smallest non-negative integer for which there is precisely one (independent) real eigenfunction with \( n \) zeros in \( (a,b) \) for each \( n \geq n_H \). (A similar definition applies for the negative eigenvalues.)

We recall that we may assume that \( \lambda = 0 \) is not an eigenvalue of (5) and that the problem

\[
y = \lambda y
\]

\[
y(a) = y(b) = 0
\]

has exactly \( N \) negative eigenvalues \( \{\lambda_i\}_{i=1}^N \) (with \( \lambda_i < \lambda_j \) if \( j > i \)) with \( N \geq 1 \).

We denote the first positive eigenvalue of (5) by \( \lambda^* \). The number \( N(\lambda^*) \) denotes the number of zeros of a corresponding eigenfunction in \( (a,b) \).

THEOREM 9.

(a) For \( N \geq 1 \), \( N(\lambda^*) \) is equal to either \( N \) or \( N - 1 \).

(b) If \( N(\lambda^*) = N, N \geq 1 \), we have \( 0 \leq n_R \leq n_H \).
(c) If $N(\lambda^*) = N-1, N \geq 2$, we have $0 \leq n_1 \leq N-1 \leq n_2$.

All the estimates are precise. (Similar estimates hold if $\lambda^* < 0$.)

We conclude with some remarks on mixed problems for the case $n = 1$; i.e., $p$ need no longer be of fixed sign. We know of no counterpart of such results for the case $n > 1$. As usual, however, we do assume $1/p, q, w \in L(a,b)$ and $w$ not a.e. zero.

We show that most of the results above cannot be improved upon by neglecting a sign condition on $p(x)$ by means of the following example of independent interest.

**Example.** Let $q \equiv 0$ in (5), $w \equiv p$, and consider the boundary problem
\[-(p(x)y)' = \lambda p(x)y \]
\[y(-1) = 0 = y(1), \]
where $p(x) = \text{sgn}(x)$. We recall that solutions $y \in AC[-1,1]$ along with $py'$.

Now for each $\lambda \in \mathbb{C} - \{0\}$, fixing a determination for the root, the function
\[y(x,\lambda) = \frac{\sin[\sqrt{\lambda}(1-1x\mathbb{I})]}{\sqrt{\lambda}}, \quad -1 \leq x \leq +1, \]
satisfies the differential equation and the boundary conditions. Thus $y$ is an eigenfunction for each $\lambda \neq 0$. If $\lambda = 0$, we may define an eigenfunction by
\[y(x,0) = -\int_{-1}^{x} \frac{ds}{p(s)}. \]

Thus every $\lambda \in \mathbb{C}$ is an eigenvalue of this problem. This example, first reported in passing by Atkinson and Mingarelli [1987] in the context of spectral asymptotics and oscillation theory, has the following features:

1. There is an interval $I \subset \mathbb{R}$ (e.g., $I = (0,1)$) corresponding to which each $\lambda \in I$ generates a positive eigenfunction (in $(a,b)$) in sharp contrast to the results of Section 2 (see also Allegretto and Mingarelli [1987]).

2. The operator $T$ defined by
\[(Tf)(x) = -(p(x)f')' \]
on $D(T) = \{f \in L^2(-1,1): f, pf' \in AC(-1,1), (pf')' \in L^2(-1,1), \text{ and } f(-1) = 0 = f(1)\}$ has infinitely many negative eigenvalues, and this in turn puts no restriction on the number of pairs of non-real eigenvalues; cf. Mingarelli [1983]. When $p(x) > 0$ a.e., $T$, as defined above, is semibounded from below.
3. The Richardson index $n_{R}^{+} = 0$ while the Haupt index does not even exist. If $\lambda < 0$, all the eigenfunctions are non-zero in $(-1,1)$. Thus $n_{R}^{-}$ does not exist.

H. Langer gave the apparently only other known example of a symmetric linear operator on a Krein space whose spectrum fills all of $C$. In his case, however, the spectrum is purely continuous.

Acknowledgment

We thank the organizers at Argonne National Laboratory for the opportunity to participate in the 1986-87 focused research program on "Spectral Theory and Boundary Value Problems."

References


ON THE NUMERICAL COMPUTATION OF EIGENVALUES OF STURM-LIOUVILLE PROBLEMS WITH MATRIX COEFFICIENTS

F. V. Atkinson*
Department of Mathematics
University of Toronto
Toronto M5S 1A1
Ontario, Canada

A. M. Krall**
Department of Mathematics
Pennsylvania State University
McAllister Building
University Park, PA 16802

G. K. Leaf
Mathematics and Computer Science Division
Argonne National Laboratory
Argonne, IL 60439

A. Zett i
Department of Mathematics
Northern Illinois University
DeKalb, IL 60115

Abstract

We describe a method for the numerical computation of the eigenvalues of Sturm-Liouville differential equations with matrix function coefficients. The numerical procedure is based on unitary matrices constructed from the original S-L problem and the monitoring of the eigenvalues of these associated unitary matrices on the unit circle \( w_j(x, \lambda), \ a \leq x \leq b. \) Our algorithm uses a search procedure based on tracking these eigenvalues. Our search procedure, described below, relies entirely on the behavior of the angles of certain points \( w_j(b, \lambda) \) on the unit circle as functions of \( \lambda \) only. As functions of \( \lambda \) these move around the unit circle in a counterclockwise direction as \( \lambda \) increases. These angles are functions of both \( x \) and \( \lambda. \) For most problems these points on the unit circle also move counterclockwise as \( x \) increases from the left end point \( a \) to the right end point \( b \) of the basic interval. However, in some problems these points move counterclockwise as \( x \) increases through values in a subinterval of \([a, b]\). It may

---

*Senior Mathematician Emeritus at the Mathematics and Computer Science Division, Argonne National Laboratory, October 1, 1986 - July 17, 1987.

**Participant in Faculty Research Leave at Argonne program, Mathematics and Computer Science Division, Argonne National Laboratory, September 1986 - May 1987.

†Participant in Faculty Research Leave at Argonne program, Mathematics and Computer Science Division, Argonne National Laboratory, September 1986 - June 1987.
be these problems that cause our procedure to miss some eigenvalues. In the near future, we plan to alter our search algorithm by monitoring the behavior of these certain points on the unit circle in the complex plane as functions of both \(x\) and \(\lambda\). In this way we hope to overcome the above-mentioned problem of missing eigenvalues in certain circumstances. The difficulty we encounter can be illustrated in the scalar case. It can occur in situations when the Prüfer angle \(\Theta(x, \lambda)\), for fixed \(\lambda\), decreases (rather than increases) as a function of \(x\) on part or all of the basic interval \([a,b]\). A significant feature of our algorithm is the fact that it computes the eigenvalues \(\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \cdots\) in order counting multiplicity. For instance, if \(\lambda_1\) is a double but not a triple eigenvalue, the procedure will produce \(\lambda_2 = \lambda_1\) for the second eigenvalue; \(\lambda_3\) would be the second distinct eigenvalue and so on. With a minor adjustment the search routine could be altered to produce the pair \((\lambda_n, \text{multiplicity of } \lambda_n)\).

1. Introduction

We describe a method for the numerical computation of the eigenvalues of Sturm-Liouville (S-L) differential equations with matrix function coefficients. The basic theory underlying this method is developed in Atkinson's book [Atkinson 1964].

Standard approaches to the numerical computation of eigenvalues for Sturm-Liouville problems involve a discrete approximation to the problem which reduces the continuous problem to a matrix eigenvalue problem. Commonly-used approximations are finite differences, collocations, and finite elements. Under the assumption that the spectrum of the continuous problem is bounded below, one expects that the eigenvalues of the matrix problem will approximate the eigenvalues of the continuum problem. However, the quality of the approximation deteriorates rapidly as one passes from the lowest few modes to the higher modes. This property is independent of what method is used to solve the matrix eigenvalue problem. The solution methods have their own attendant problems. For example, if a direct method is used, as exemplified in the software package EISPACK, storage of full matrices is required and substantial arithmetic is needed. Moreover, all the matrix eigenvalues are produced even though only the lowest few have any approximating value. An alternative to a direct method would be an iterative technique such as inverse power iterations. Such a method requires less storage and less arithmetic; however, the computational difficulties increase very rapidly and the accuracy decreases sharply beyond the first few lowest eigenvalues.

For scalar Sturm-Liouville problems, numerical methods based on the Prüfer transformation [Bailey, Gordon, and Shampine 1978] avoid many of the above-mentioned difficulties. In principle, any mode can be determined as accurately as any other mode in sharp contrast with the matrix eigenvalue approach where only the lowest modes are well approximated. This method has been implemented in such codes as SLEIGN [Bailey, Gordon, and Shampine 1978]. For Sturm-Liouville equation with matrix coefficients, the numerical method described herein is not an extension of SLEIGN since the Prüfer transformation in its original form does not extend to problems with matrix coefficients. The general approach is similar; however, in the matrix
context a certain class of unitary matrices plays a central role in the theory and the numerical approach. The scalar analog of this approach would be to study \( \exp(i\theta) \) in place of the Prüfer angle \( \theta \).

2. The Problem

We are interested in the numerical computation of the eigenvalues of the following matrix-vector Sturm-Liouville system:

\[
-(P(x)y')' + Q(x)y = \lambda W(x)y, \quad a \leq x \leq b,
\]

\[
A_1y(a) + A_2(Py')(a) = 0,
\]

\[
B_1y(b) + B_2(Py')(b) = 0, \quad -\infty < a < b < \infty.
\]

Here \( P, Q, W \) are \( n \times n \) matrix functions from \([a,b]\) into the reals satisfying the following conditions:

\[ P^* = P, \ W^* = W, \ P \text{ and } W \text{ positive definite}, \]

and

\[ P, Q, W \text{ are continuous.} \]

The adjoint of a matrix \( A \) is denoted by \( A^* \). In (2.1) \( \lambda \) is a real valued eigenvalue parameter, the unknown function \( y \) is an \( n \)-dimensional column vector, and \( A_j, B_j, j = 1,2 \) are \( n \times : \) real constant matrices. Although the equation (2.1a) is formally self-adjoint, some conditions are needed on the boundary conditions (2.1b) and (2.1c) to make the problem (2.1) self-adjoint. We choose to restrict the boundary condition matrices \( A_j, B_j \) such that (i) the problem (2.1) is self-adjoint, and (ii) to avoid having to compute certain inverses of matrices arising below. These restrictions are as follows: Let \( (A:B) \) denote the \( n \times 2n \) matrix obtained by placing the \( n \times n \) matrix \( B \) to the right of the \( n \times n \) matrix \( A \).

\[
\text{rank}(A_1:A_2) = n = \text{rank}(B_1:B_2)
\]

\[
A_1^*A_1 + A_2^*A_2 = I, \quad A_1^*A_2 - A_2^*A_1 = 0
\]

\[
B_1^*B_1 + B_2^*B_2 = I, \quad B_1^*B_2 - B_2^*B_1 = 0.
\]

Observe that the conditions (1.2b) are equivalent to

\[
A_1^*A_1 + A_2^*A_2 = I, \quad A_1^*A_2 - A_2^*A_1 = 0,
\]

\[
B_1^*B_1 + B_2^*B_2 = I, \quad B_1^*B_2 - B_2^*B_1 = 0.
\]
As mentioned above, conditions (1.3) can be used to avoid having to compute inverses of certain matrices. This is significant numerically and follows from the observation that

\[ AA^* + BB^* = I \implies (A + iB)^{-1} = (A^* - iB^*) \text{ with } i = \sqrt{-1}. \]  

(2.4)

**DEFINITION.** An eigenvalue of problem (2.1) is a value of the parameter \( \lambda \) for which the equation (2.1a) has a solution \( y \) which is nontrivial, i.e., not identically zero on \([a, b]\) and which satisfies both boundary conditions (2.1b) and (2.1c).

**REMARK.** Even though the boundary conditions (2.1b) and (2.1c) are separated, there exist, in general, eigenvalues of multiplicity greater than 1 when \( n > 1 \). This is in contrast to the scalar case \( n = 1 \) when the eigenvalues for separated boundary conditions are all simple. Our code, to be described below, can be used to detect the multiplicities of the computed eigenvalues.

Under the conditions (2.2) the boundary value problem (2.1) has only real eigenvalues; these are countably infinite in number and can be indexed so that they satisfy \(-\infty < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \ldots; \lambda_n \to \infty \text{ as } n \to \infty.\)

Important special cases of the boundary conditions above are the Dirichlet conditions

\[ y(a) = 0 = y(b), \ (A_1 = I = B_1, \ A_2 = 0 = B_2) \]

and the Neumann conditions

\[ (Py')(a) = 0 = (Py')(b), \ (A_1 = 0 = B_1, \ A_2 = I = B_2). \]

These can be mixed, that is, one can have Dirichlet conditions at one end point and Neumann at the other.

3. **Theoretical Foundations**

In this section we summarize the basic theory used by our code. For details, the reader is referred to [Atkinson 1964], especially Chapter 10.

Let \( Y = Y(x, \lambda) \) be a fundamental matrix solution of the differential equation (2.1a) and following [Atkinson 1964, Ch. 10] set

\[ U = Y, \ V = PY', \ R = R(x, \lambda) = Q(x) - \lambda W(x). \]

Then (2.1a) is equivalent to the first-order system

\[ U' = P^{-1}V, \ V' = RU. \]  

(3.1)

The boundary conditions (2.1b) expressed in terms of \( U \) and \( V \) are
For any fixed value of $\lambda$, the first-order system (3.1) with the initial condition (3.2) determines a unique solution $U(x)$, $V(x)$ for $x$ in $[a,b]$. In particular, the final values $U(b)$, $V(b)$ are uniquely determined. The parameter $\lambda$ is an eigenvalue of the S-L problem (2.1) if and only if the final values $U(b)$, $V(b)$ are such that the corresponding solution $y$ satisfies the boundary condition (2.1c). This occurs precisely when

$$B_1 U(b) + B_2 V(b) \text{ is singular.} \quad (3.3)$$

Clearly, a sufficient but not necessary condition for (3.3) to hold is for $U(b) = -B_2^*$ and $V(b) = B_1^*$.

**Definition.**

$$\Theta(x) = (V(x) + iU(x))(V(x) - iU(x))^{-1}, \quad a \leq x \leq b, \quad i = \sqrt{-1}. \quad (3.6)$$

**Theorem 3.1.** Let $U$, $V$ be defined as above with $U(a)$, $V(a)$ given by (3.2). Then

(i) $(V(x) - iU(x))^{-1}$ exists for all $x$ in $[a,b]$,

(ii) $\Theta(x)$ is unitary for all $x$ in $[a,b]$.

(iii) $\Theta$ satisfies the following equation

$$\Theta' = i\Theta \Omega, \quad (3.4)$$

where

$$2\Omega = (\Theta + I)^*P^{-1}(\Theta + I) - (\Theta - I)^*R(\lambda)(\Theta - I). \quad (3.5)$$

**Proof.** This is taken from [1, Ch. 10]. Consider

$$(V^* - iU^*)(V + iU) = V^*V + U^*U = (V^* + iU)(V - iU).$$

Claim: $V^*(x)V(x) + U^*(x)U(x) > 0$ for all $x$ in $[a,b]$.

If not, then for some $x_0$ there exists a vector $c \neq 0$ such that $c^*(V^*V + U^*U)(x_0)c = 0$. Then $c^*(V^*V)(x_0)c = 0 = c^* (U^*U)(x_0)c$. Thus $U(x_0)c = 0 = V(x_0)c$. Then $u = Uc$, $v = Vc$ is a vector solution of (1) satisfying $u(x_0) = 0 = v(x_0)$. This implies that $U(x)c = 0 = V(x)c$ for all $x$ in $[a,b]$ and therefore that

$$(V - iU)(x)c = 0 = (V - iU)(x_0)c.$$ 

This implies that $c = 0$ which is a contradiction completing the proof of (i).

Now (ii) follows from
\[ \Theta^* \Theta = (V-iU)^{-1} (V+iU)^*(V+iU)(V-iU)^{-1} \]
\[ = (V^*+iU^*)^{-1} (V^*-iU^*)(V+iU)(V-iU)^{-1} \]
\[ = (V^*+iU^*)^{-1} (V^*+iU^*)(V-iU)(V-iU)^{-1} \]
\[ = I . \]

To prove (iii), we proceed as follows:

\[ \Theta = (V+iU)(V-iU)^{-1} \]
\[ \Theta^* = (V^*+iU^*)(V-iU)^{-1} + (V+iU)(V-iU)^{-1} . \]

\[ (V-iU)^{-1} (V-iU)^{-1} = (V+iU)(V-iU)^{-1} - \Theta (V-iU)(V-iU)^{-1} . \]

Multiplying on the left by \( \Theta^* \) and using \( \Theta^* \Theta = I \) we get

\[ \Theta^* \Theta = (V^*+iU^*)^{-1} (V^*-iU^*)(V^*+iU^*)(V-iU)^{-1} - (V-iU)(V-iU)^{-1} \]
\[ = (V^*+iU^*)^{-1} (2iV^*U^* - 2iU^*V^*)(V-iU)^{-1} \]
\[ = (V^*+iU^*)^{-1} [2iV^* (P^{-1} V) - 2iU^* (R(\lambda) U)] (V-iU)^{-1} \]
\[ = 2i (V^*+iU^*)^{-1} [V^* P^{-1} V - U^* R(\lambda) U] (V-iU)^{-1} \]
\[ = 2i ((V^*+iU^*)^{-1} V^* P^{-1} V (V-iU)^{-1} - (V^*+iU^*)^{-1} U^* R(\lambda) U (V-iU)^{-1}) . \]

From the definition of \( \Theta \) we have

\[ \Theta + I = 2V (V-iU)^{-1} , \]
\[ \Theta - I = 2iU (V-iU)^{-1} . \]

From this and the above calculation we get

\[ \Theta^{-1} \Theta^* = 2i \left[ \frac{1}{4} (\Theta+I)^* P^{-1} (\Theta+I) - \frac{1}{2i} \left( (\Theta-I)^* R(\lambda) \frac{1}{2i} (\Theta-I) \right) \right] \]

or

\[ 2\Theta^* = i \Theta [ (\Theta+I)^* P^{-1} (\Theta+I) - (\Theta-I)^* R(\lambda) (\Theta-I) ] . \]

giving (iii). This completes the proof of Theorem 3.1.

Next we express the boundary conditions in terms of the matrix \( \Theta \). From (3.2) using (2.3) and (2.4) we get
\[ \Theta(a) = (A_1^*-iA_2^*)(A_1-iA_2). \]  

(3.6)

For a given value of \( \lambda \), equations (3.4), (3.5), and (3.6) determine a unique solution \( \Theta(x, \lambda) \) defined for \( x \) in \([a,b] \). Our goal is to use \( \Theta \) to determine the eigenvalues of the original matrix S-L problem (2.1). In particular, we will examine the eigenvalues of a matrix constructed from \( \Theta(b, \lambda) \) and \( B_1 \) and \( B_2 \). The behavior of these associated eigenvalues as functions of \( \lambda \) is used to compute the eigenvalues of the original matrix S-L problem (2.1).

Let

\[ B(\lambda) = (B_1^*+iB_2^*)(B_1+iB_2)\Theta(b, \lambda). \]  

(3.7)

It is this matrix \( B(\lambda) \) whose eigenvalues we monitor, as a function of \( \lambda \), in order to compute the eigenvalues of the S-L problem (2.1). The relationship between the eigenvalues of \( B(\lambda) \) and those of the original S-L problem is given by the following theorem.

THEOREM 3.2. The matrix \( B(\lambda) \), given by (3.7), is unitary for each \( \lambda \). A number \( \lambda \) is an eigenvalue of the matrix Sturm-Liouville problem (2.1) if and only if the number 1 is an eigenvalue of \( B(\lambda) \). Furthermore, the multiplicity of \( \lambda \) as an eigenvalue of (2.1) is the same as the multiplicity of the number 1 as an eigenvalue of \( B(\lambda) \).

Proof. A straightforward computation, using conditions (2.2b), shows that the matrix \( (B_1^*+iB_2^*)(B_1+iB_2) \) is unitary. From Theorem 3.1 we have that \( \Theta(b, \lambda) \) is unitary. Since the product of unitary matrices is unitary, it follows that \( B(\lambda) \) is unitary.

Suppose 1 is an eigenvalue of \( B(\lambda) \). Then there exists a nonzero vector \( w \) such that

\[ B(\lambda)w = w, \text{ i.e.,} \]

\[ (B_1^*+iB_2^*)(B_1+iB_2)(V(b, \lambda)+iU(b, \lambda))(V(b, \lambda)-iU(b, \lambda))^{-1}w = w. \]

Let \( z = (V(b, \lambda)-iU(b, \lambda))^{-1}w \). Then \( z \neq 0 \) and

\[ (B_1^*+iB_2^*)(B_1+iB_2)(V(b, \lambda)+iU(b, \lambda))z = (V(b, \lambda)-iU(b, \lambda))z. \]

From (1.4) we have

\[ (B_1^*+iB_2^*)^{-1} = B_1-iB_2. \]

Hence

\[ (B_1+iB_2)(V(b, \lambda)+iU(b, \lambda))z = (B_1-iB_2)(V(b, \lambda)-iU(b, \lambda))z. \]

This implies that

\[ 2i(B_1U(b, \lambda) + B_2 V(b, \lambda))z = 0. \]

Therefore, \( B_1U(b, \lambda) + B_2 V(b, \lambda) \) is singular and \( \lambda \) is an eigenvalue by (3.3).
The converse follows by reversing the above steps. The furthermore part also follows from the above computations. This completes the proof of Theorem 3.2.

**THEOREM 3.3.** The eigenvalues of \( \Theta(b, \lambda) \) move positively, i.e., counterclockwise on the unit circle as \( \lambda \) increases.

Proof. See [1, p. 308].

From Theorems 3.2 and 3.3 we get the following.

**PROPOSITION.** Let \( \omega_1(b, \lambda), \omega_2(b, \lambda), \ldots, \omega_n(b, \lambda) \) be the arguments of the eigenvalues of the unitary matrix \( B(\lambda) \) defined such that \( \omega_j(b, \lambda) \) is a continuous function of \( \lambda \). Let \( \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \cdots \) denote the eigenvalues of the original matrix Sturm-Liouville problem.

Fundamental to this numerical procedure is the following observation: Let \( I_j \) denote the number of times that \( \omega_j(b, \lambda) \) has passed through the point \((1,0)\) on the unit circle. Then if \( \lambda_k < \lambda < \lambda_{k+1} \) we have

\[
\sum_{j=1}^{n} I_j(\lambda) = k.
\]

The code is based on monitoring the arguments of the eigenvalues of \( B(\lambda) \) and counting the number of times these go through a multiple of 360°. In the numerical procedure, this monitoring of the movement of the eigenvalues on the unit circle of the unitary matrix \( B(\lambda) = B(x, \lambda) \) is done with respect to \( x \) as \( x \) varies from \( x = a \) to \( x = b \) in the interval \([a,b]\).

4. Numerical Procedure

As seen in the earlier sections, we have an \( n \times n \) complex valued matrix differential equation defined on the interval \([a,b]\).

\[
\Theta'(x) = i\Theta(x)\Omega(x, \Theta(x)) \tag{4.1a}
\]

\[
\Theta(x=a) = \Theta_a \tag{4.1b}
\]

where

\[
\Omega(x, \Theta(x)) = \frac{1}{2} (\Theta+I)^* P^{-1}(x)(\Theta+I) - \frac{1}{2} (\Theta-I)^* R(x, \lambda)(\Theta-I), \tag{4.2}
\]

\[
R(x, \lambda) = Q(x) - \lambda W(x). \tag{4.3}
\]

On the interval \([a,b]\), the matrices \( P, Q, W \) are real hermitian with \( P \) and \( W \) positive definite, \( \Omega \) is hermitian, and \( \Theta \) is unitary. In addition, the constant matrix \( \Theta_a \) is unitary.
We start the approximation procedure by imposing a uniform mesh
\[ a = x_0 < x_1 < x_2 < \cdots < x_N = b \]
on the interval with mesh spacing \( h = (b-a)/N \). Consider a subinterval \([x_j, x_j+h]\) with \( \Theta(x_j) \) known. On this subinterval we make the following approximations:
\[ P(x) \approx P(x_j), \quad Q(x) \approx Q(x_j), \quad W(x) \approx W(x_j), \quad (4.4a) \]
and
\[ \Theta(x) = \Theta(x_j). \quad (4.4b) \]

With these approximations \( \Omega_j = \Omega(x_j, \Theta(x_j)) \) is a constant matrix on the subinterval \([x_j, x_j+h]\). Thus on this interval we can solve the approximate differential equation
\[ \Theta'(x) = i\Theta \Omega_j \quad (4.5) \]
with \( \Theta(x=x_j) \) known, by the expression
\[ \Theta(x) = \Theta(x_j) \exp(i(x-x_j)\Omega_j). \]
In particular, we have the approximation
\[ \Theta(x_j+h) = \Theta(x_j) \exp(ih \Omega_j). \quad (4.6) \]

The next task is to evaluate the exponential of the matrix \( ih \Omega_j \). Our basic approach is to make sure that any approximation to \( \Theta(x) \) is also a unitary matrix. That is, all approximations will be unitary so that this property is preserved. This goal is very important for the following reason. Recall that \( \lambda \) is an eigenvalue of the Sturm-Liouville system iff the point \((1,0)\) is in the spectrum of the unitary matrix \( B(\lambda) \) defined in (3.7). Since \( B(\lambda) \) is unitary; its spectrum lies on the unit circle. Thus the eigenvalues are functions of a single parameter such as the phase angle rather than two parameters, e.g. Cartesian coordinates, which would be the case if the matrix \( B(\lambda) \) were not unitary. Hence, in order to determine whether the point \((1,0)\) lies in the spectrum, one has only to monitor the phase of each eigenvalue for multiples of \( 2\pi \). Since this one-dimensional monitoring is much simpler than a two-dimensional situation, our goal is to ensure that our approximate matrices are unitary. To this end, we approximate the exponential of the matrix appearing in (4.6) by numerically determining the complete set of eigenvalues and eigenvectors of the matrix \( \Omega_j \). Thus we have
\[ \Omega_j = \Phi_j^* \omega_j \Phi_j, \]
where \( \Phi_j \) denotes the matrix formed by the column of the eigenvectors and \( \omega_j \) is the diagonal matrix formed by the eigenvalues. Since \( \Omega_j \) is hermitian, we can be assured that \( \omega_j \) has real entries and that the eigenvectors are orthogonal. Thus we can normalize so that \( \Phi_j^* \Phi_j = I \).

With this spectral decomposition, (3.6) takes the form
\[ \Theta(x_{j+1}) = \Theta(x_j) \Phi_j e^{i \omega_0 \Phi_j^*}, \quad j = 0, 1, \ldots, N. \] (4.7)

We observe that when \( \Theta(x_j) \) is unitary, then \( \Theta(x_{j+1}) \) is unitary as desired. Equation (4.7) forms the basis for the numerical approximation of (4.1). For each \( j \), we also calculate the spectrum of \( B(\lambda) \) and monitor the phase associated with each of these eigenvalues. These phases are used in the search for the desired eigenvalues as described in the next section.

5. Search Procedure

For a given value of \( \lambda \), the numerical solution procedure calculates a unitary matrix \( B(\lambda) \) and its corresponding eigenvalue \( \{\mu_k(\lambda)\}^n_{k=1} \). Since each eigenvalue \( \mu_k(\lambda) \) lies on the unit circle, let \( w_k(\lambda) \) denote the associated positive angle, \( n_k(\lambda) \) the greatest integer in \( w_k(\lambda)/2\pi \), and

\[ I(\lambda) = \sum_{k=1}^n n_k(\lambda). \]

The search procedure is based on the following observation (cf. section 2). Let the eigenvalues be ordered \( \lambda_1 < \lambda_2 < \ldots \) and suppose that \( \lambda < \lambda_m \) for some integer \( m \), then \( I(\lambda) < m \); moreover, if \( \lambda > \lambda_m \) then \( I(\lambda) \geq m \). Hence, the search is based on solving the equation

\[ I(\lambda) = m \]

for \( \lambda = \lambda_m \).

The search for the \( m \)-th eigenvalue proceeds in two stages. The first stage is to establish lower and upper bounds for \( \lambda_m \). This is done by starting with an arbitrary estimate \( \lambda^{(0)} \), then comparing \( I(\lambda^{(0)}) \) with \( m \). If \( I(\lambda^{(0)}) < m \) we have a lower bound \( \lambda_L = \lambda^{(0)} \) and we increase \( \lambda^{(0)} \) in order to find either an upper bound or a tighter lower bound. If \( I(\lambda^{(0)}) \geq m \), we have an upper bound \( \lambda_U = \lambda^{(0)} \) and we decrease \( \lambda^{(0)} \) in order to find either a lower bound or a tighter upper bound. Having established a lower and an upper bound for \( \lambda_m \), the second stage estimates \( \lambda_m \) by successively tightening the bounds by bisection of the bounds interval using the inequalities \( I(\lambda) \leq m \) to establish new lower/upper bounds. The bisection continues until the relative size of the bounds interval is less than a specified error tolerance.

The numerical procedure described above is not in final form. It seems to work well in most cases but in certain circumstances it may miss some eigenvalue, particularly the first one or possibly even the first few. The "thick barrier" option can be used to try to get all eigenvalues followed by the thin barrier option to obtain more accuracy.

Experience has shown that this procedure is very sensitive to step size. Thus it is advisable to check for stability with respect to step size. When the step size is not fine enough, the computed numbers can be way off the mark — nowhere near the correct answer.
6. Examples

In this section we illustrate the numerical procedure described in the previous sections using various simple problems.

EXAMPLE I. A scalar Airy equation which demonstrates a basic capability with standard boundary conditions.

Governing equation:

\[-Y'' = \lambda x Y \text{ on } [0,1]\]

\[Y(0) = Y(1) = 0\]

The eigenvalues for this problem were computed independently by means of series solutions. We will use this example to illustrate the effect of mesh size and corrector iterations on the accuracy of the numerical eigenvalues. The effects will be considered for the first and fourth eigenvalues of the above Airy equation. From the series solution method we take the benchmark values as

\[\lambda_1 = 18.95621569\]

\[\lambda_4 = 340.966959\]

In Table 6.1 we present the results for the numerical calculation of \(\lambda_1\) for various combinations of mesh subdivisions and correction iterations.

<table>
<thead>
<tr>
<th>Number of correction iterations</th>
<th>Number of Mesh Subdivisions</th>
<th>100</th>
<th>200</th>
<th>400</th>
<th>800</th>
<th>1600</th>
<th>3200</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>18.7124</td>
<td>18.8365</td>
<td>18.89695</td>
<td>10.9267</td>
<td>18.9415</td>
<td>18.9489</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.01286</td>
<td>0.00632</td>
<td>0.00313</td>
<td>0.00156</td>
<td>0.00078</td>
<td>0.00039</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.78)</td>
<td>(11.32)</td>
<td>(22.68)</td>
<td>(45.08)</td>
<td>(88.60)</td>
<td>(73.12)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>18.9803</td>
<td>18.9622</td>
<td>18.9577</td>
<td>18.9566</td>
<td>18.95636</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.00127</td>
<td>0.000316</td>
<td>0.00078</td>
<td>0.000020</td>
<td>0.000008</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(9.93)</td>
<td>(19.83)</td>
<td>(39.40)</td>
<td>(78.70)</td>
<td>(155.32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>18.9437</td>
<td>18.9533</td>
<td>18.95552</td>
<td>18.9561</td>
<td>18.95621569</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.00066</td>
<td>0.000153</td>
<td>0.000036</td>
<td>0.000061</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(14.12)</td>
<td>(28.08)</td>
<td>(55.25)</td>
<td>(2.162)</td>
<td>(221.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>18.9447</td>
<td>18.9534</td>
<td>18.95554</td>
<td>18.95609</td>
<td>18.95621569</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.00061</td>
<td>0.000149</td>
<td>0.000035</td>
<td>0.000059</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(18.32)</td>
<td>(36.63)</td>
<td>(72.92)</td>
<td>(144.37)</td>
<td>(287.5)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The number of mesh subdivisions varied from 100 to 3200 while the number of correction
iterations varied from 0 to 3. For each calculation we present the numerical value, the relative
error, and the cpu time for the computations. Note these calculations were done on an Alliant
computer and the times should be interpreted as indicating relative cost of each computation.
This information is also presented graphically in Figures 6.1 and 6.2. In Figure 6.1 we show a
plot of the log of the relative error against the log of the number of mesh points for various
correction iterations. This graph shows that it may be worthwhile performing at least one itera-
tion, and it is probably not worthwhile performing more than three iterations. Figure 6.2 confirms
these conjectures. This figure shows a plot of the log of the relative error against the log of the
cpu time. This clearly shows that one or two correction iterations are worthwhile in terms of cpu
time.

For the fourth eigenvalue, Table 6.2 presents a similar set of data.

<table>
<thead>
<tr>
<th>Number of correction iterations</th>
<th>400</th>
<th>800</th>
<th>1600</th>
<th>3200</th>
<th>6400</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>336.9337</td>
<td>339.4954</td>
<td>340.3552</td>
<td>340.6913</td>
<td>340.83663</td>
</tr>
<tr>
<td></td>
<td>0.0118</td>
<td>0.0043</td>
<td>0.0018</td>
<td>0.0008</td>
<td>0.00038</td>
</tr>
<tr>
<td></td>
<td>(21.45)</td>
<td>(43.55)</td>
<td>(86.88)</td>
<td>(169.97)</td>
<td>(332.52)</td>
</tr>
<tr>
<td>1</td>
<td>345.0212</td>
<td>341.9532</td>
<td>341.2102</td>
<td>341.0276</td>
<td>340.95221</td>
</tr>
<tr>
<td></td>
<td>0.0119</td>
<td>0.0029</td>
<td>0.0007</td>
<td>0.0002</td>
<td>0.000045</td>
</tr>
<tr>
<td></td>
<td>(36.75)</td>
<td>(76.12)</td>
<td>(2.1.45)</td>
<td>(296.42)</td>
<td>579.52</td>
</tr>
<tr>
<td>2</td>
<td>338.7260</td>
<td>340.4577</td>
<td>340.8440</td>
<td>340.93665</td>
<td>340.95943</td>
</tr>
<tr>
<td></td>
<td>0.0066</td>
<td>0.0015</td>
<td>0.0004</td>
<td>0.000089</td>
<td>0.000022</td>
</tr>
<tr>
<td></td>
<td>(53.30)</td>
<td>(107.37)</td>
<td>(216.02)</td>
<td>(419.87)</td>
<td>826.27</td>
</tr>
<tr>
<td></td>
<td>0.0048</td>
<td>0.0014</td>
<td>0.00035</td>
<td>0.000088</td>
<td>0.000022</td>
</tr>
<tr>
<td></td>
<td>(68.88)</td>
<td>(140.07)</td>
<td>(266.25)</td>
<td>(544.93)</td>
<td>1071.0</td>
</tr>
</tbody>
</table>

In this case the mesh varied from 400 to 6400 while the number of correction iterations
varied from 0 to 3. Comparing these results with the $\lambda_1$ case, we see that for a given mesh, the
quality of the approximation is much less. However, the conclusion that it is worthwhile doing at
least one corrector iteration is still valid. This latter point is illustrated in Figure 6.3, which
shows a plot of the log of relative error against the log of the time.
SCALAR AIRY EQUATION
LOWEST EIGENVALUE

Figure 6.1

LEGEND
	× = No Corr Iter
	○ = One Corr Iter
	▼ = Two Corr Iter
	□ = Three Corr Iter
SCALAR AIRY EQUATION
LOWEST EIGENVALUE

LEGEND
× = No Corr Iter
◇ = One Corr Iter
△ = Two Corr Iter
☒ = Three Corr Iter

Figure 6.2
SCALAR AIRY EQUATION
FOURTH EIGENVALUE

LEGEND

× = No Corr Iter
◇ = One Corr Iter
▽ = Two Corr Iter
⊗ = Three Corr Iter

Figure 6.3
EXAMPLE II. This system example serves to illustrate that the procedure does handle genuine systems. Also, we compute the 9-th eigenvalue to show that "large" eigenvalues can be computed. This example further illustrates that eigenvalues of multiplicity greater than one can be computed routinely. The example is

\[-(PY)' = \lambda WY \text{ on } [0,1] \text{ with } Y(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = Y(1),\]

\[P = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad Q = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad W = \begin{bmatrix} 5/2 & 1 \\ 1 & 2 \end{bmatrix}.\]

The actual eigenvalues for this problem can be computed by hand. They are:

\[\lambda_n = n^2\pi^2/4, \quad n = 1, 2, 3, \ldots,\]

with the even indices producing eigenvalues of multiplicity 2. Thus note that \(\lambda_1 = \pi^2/4, \lambda_2 = \pi^2, \lambda_3 = \pi^2, \lambda_4 = 9\pi^2/4,\) etc. In particular, for comparison with the computed eigenvalues we take

\[\lambda_1 = 2.467401\]

\[\lambda_4 = 22.206610\]

\[\lambda_9 = 6^2\pi^2/4 = 88.82644\]

<table>
<thead>
<tr>
<th>Number of correction iterations</th>
<th>100</th>
<th>200</th>
<th>400</th>
<th>800</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.464847</td>
<td>2.466122</td>
<td>2.466760</td>
<td>2.467082</td>
</tr>
<tr>
<td></td>
<td>.001064</td>
<td>.000533</td>
<td>.000267</td>
<td>.000133</td>
</tr>
<tr>
<td></td>
<td>(75)</td>
<td>(150)</td>
<td>(299)</td>
<td>(612)</td>
</tr>
<tr>
<td>1</td>
<td>2.467689</td>
<td>2.467475</td>
<td>2.467421</td>
<td>2.467407</td>
</tr>
<tr>
<td></td>
<td>.000120</td>
<td>.0000308</td>
<td>.0000083</td>
<td>.0000025</td>
</tr>
<tr>
<td></td>
<td>(126)</td>
<td>(254)</td>
<td>(521)</td>
<td>(1019)</td>
</tr>
<tr>
<td>2</td>
<td>2.467256</td>
<td>2.467362</td>
<td>2.467388</td>
<td>2.4673998</td>
</tr>
<tr>
<td></td>
<td>.000060</td>
<td>.0000163</td>
<td>.0000054</td>
<td>.0000004</td>
</tr>
<tr>
<td></td>
<td>(170)</td>
<td>(359)</td>
<td>(716)</td>
<td>(1433)</td>
</tr>
<tr>
<td>3</td>
<td>2.467252</td>
<td>2.467365</td>
<td>2.467390</td>
<td>2.467395</td>
</tr>
<tr>
<td></td>
<td>.0000620</td>
<td>.000015</td>
<td>.0000046</td>
<td>.00000254</td>
</tr>
<tr>
<td></td>
<td>(233)</td>
<td>(467)</td>
<td>(924)</td>
<td>(1867)</td>
</tr>
</tbody>
</table>

Table 6.3 Numerical Approximations to \(\lambda_1.\)
The actual value of \(\lambda_1 = \pi^2/4 = 2.467401\)
Table 6.4
Numerical Computations of $\lambda_4$.
The actual value of $\lambda_4$ is $\lambda_4 = 3^2 \pi^2 / 4 \approx 22.206610$

<table>
<thead>
<tr>
<th>Number of corrections iterations</th>
<th>Number of Mesh Subdivisions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
</tr>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>22.196332</td>
</tr>
<tr>
<td></td>
<td>.000467</td>
</tr>
<tr>
<td></td>
<td>(311)</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>22.223905</td>
</tr>
<tr>
<td></td>
<td>.000786</td>
</tr>
<tr>
<td></td>
<td>(523)</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>22.197941</td>
</tr>
<tr>
<td></td>
<td>.000394</td>
</tr>
<tr>
<td></td>
<td>(739)</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>22.197983</td>
</tr>
<tr>
<td></td>
<td>.000392</td>
</tr>
<tr>
<td></td>
<td>(961)</td>
</tr>
</tbody>
</table>

The computed value of $\lambda_9$ using 1600 steps and two iterations is computed $\lambda_9 = 88.809485$.
The actual value is $\lambda_9 = 6^2 \pi^2 / 4 \approx 88.82644$.

7. User Guide

The program consists of a main program together with seven subroutines plus five user subroutines. In addition, input data is specified via interactive terminal prompts. The results of the calculation are displayed on the terminal. The data which is specified at the terminal is as follows.

1. Number of equations: $m$
2. Index of the desired eigenvalues: $nmeig$
3. Lower and upper interval limits: $al, br$
4. Number of subintervals: $jmax$
5. Number of mesh refinements: $mxstbl$
   If this number is one, a calculation is performed using the number of subintervals $jmax$; however, if $mxstbl > 1$, then the mesh is doubled, the calculation repeated and doubled again repeatedly $mxstbl$ times.
6. Tolerance for the search: $\epsilon_p$
   The search for the desired eigenvalue is carried out to a relative error specified by $\epsilon_p$.

7. Number of correction iterations: $n_{iter}$

8. Specification of boundary condition type at left interval point: $ibcl$
   At present, when specifying boundary conditions from the terminal, the program is restricted to the basic three types of boundary conditions: Dirichlet, Neumann and Robin. Dirichlet is specified with $ibcl = 0$ while Neumann with $ibcl = 1$ and Robin with $ibcl = 2$. If Dirichlet or Neumann conditions are specified, then these conditions are imposed on each dependent variable in the system. However, if a Robin condition is specified, then the boundary condition can vary from one dependent variable to another. This is accomplished by specifying the user-supplied boundary matrices in subroutine bndry.

9. Specification of boundary condition type at right interval point: $ibcr$
   The remarks cited under item 8 apply to this situation.

   The user deck consists of five user-supplied subroutines.

   1. Subroutine $capp (m,x,p)$
      This routine defines the $m \times m$ diffusion matrix $p(x)$ for each $x$ in the interval $[a,b]$.

   2. Subroutine $capinv (m,x,pinv)$
      This routine defines the matrix inverse $pinv(x)$ to the matrix $p(x)$.

   3. Subroutine $capq(m,x,q)$
      This routine defines the absorption matrix $q(x)$ for each $x$ in the interval $[a,b]$.

   4. Subroutine $capw(m,x,w)$
      This routine defines the matrix $w(x)$ for each $x$ in the interval $[a,b]$.

   5. Subroutine $bndry (ibcl,ibcr)$
      This routine allows the user to specify the boundary matrices $A_1, A_2, B_1, B_2$ described in section 2. In addition, having specified these matrices, this routine forms the two matrices $\Theta(a)$ specified in (3.6) as
      $$\Theta(a) = (A_1^* - iA_2^*)(A_1 - iA_2)$$
      and the matrix $B(\lambda)$ specified in (3.7) as
      $$B(\lambda) = (B_1^* + iB_2^*)(B_1 + iB_2)\Theta(b,\lambda).$$

References


SUFFICIENT CONDITIONS FOR
WEIGHTED SUM INTERPOLATION INEQUALITIES IN $\mathbb{R}^n$

R. C. Brown
Mathematics Department
University of Alabama
Tuscaloosa, AL 35487

D. B. Hinton
Mathematics Department
University of Tennessee
Knoxville, TN 37996

Abstract

Weighted sum inequalities are established for functions defined on a domain in $\mathbb{R}^n$ which are contained in an appropriate weighted Sobolev space. The inequalities establish a bound for the integral of the $j$-th derivative in terms of the integrals of the function and the $m$-th derivative for $j < m$. The weights in the three integrals are allowed to be different as well as the indices defining the norm.

1. Introduction

We are concerned in this paper with sufficient conditions for a sum inequality of the following kind to hold:

$$\int_{\Omega} N |D^j u|^p \leq K \left\{ \epsilon^{\phi} \left[ \int_{\Omega} W |u|^q \right]^{p/q} + \epsilon^{\theta} \left[ \int_{\Omega} P |D^m u|^r \right]^{p/r} \right\}. \quad (1.1)$$

Here $\Omega$ is a possibly unbounded domain (open connected set) in $\mathbb{R}^n$ and $N$, $W$, $P$ are measurable functions such that $N(x) \geq 0$, $W(x), P(x) > 0$ and which satisfy certain additional conditions. The exponents $\phi, \theta$ are certain non-negative functions of $p, q, r, m, n,$ and $j$. The parameter $\epsilon$ is an arbitrary element of some interval $\Gamma = (0, \epsilon_0)$, $\epsilon_0 \leq \infty$. The function $u$ is in an appropriate weighted Sobolev space $D(\Omega)$, and the constant $K$ is independent of $u$ although it may depend on the other parameters. If $\Gamma = (0, \infty)$, it is easily shown by a calculus argument that (1.1) is equivalent to a product inequality

$$\int_{\Omega} N |D^j u|^p \leq K_1 \left[ \int_{\Omega} W |u|^q \right]^{\phi/(\phi+\theta)} \left[ \int_{\Omega} P |D^m u|^r \right]^{\theta/(\phi+\theta)}. \quad (1.2)$$

When $N = W = P = 1$, (1.2) is the standard Gagliardo Nirenberg interpolation inequality. The case $n = 1$ of (1.1)-(1.2) with general weights has been extensively studied by Kwong and Zettl [1980b, 1980a, 1981b, and 1981a], as well as by Brown and Hinton [1985, 1986, 1987a, and 1987b]. The papers of Kwong and Zettl also contain some historical information. Necessary and
sufficient conditions have been given by Lin [1986] for (1.2) when $n \geq 1$, $\Omega = R^n$, $N(t) = |t|^\beta$, $W(t) = |t|^\alpha$, $P(t) = |t|^\alpha$, $u \in C^m(\Omega)$ and certain restrictions are imposed on the parameters. Several sufficient conditions for product integrals for general weights were also given in our previous papers [1987a and 1987b]. In this paper we relax the dimensionality requirement $m - j > n/r$ (when $n > 1$) required in [Brown and Hinton 1987b].

We also note that inequalities (1.1)-(1.2) are extremely important in the spectral and perturbation theory of ordinary and partial differential operators. For example, when $p = q = r = 2$, if the embedding operator naturally defined by (1.1)-(1.2) is compact, then the spectrum of an associated operator is discrete. Applications related to the determination of the lower bound of the spectrum are given in [Brown and Hinton 1987a].

It is easy to see that the requirement $m - j > n/r$ of [Brown and Hinton 1987b] rules out some desirable cases, e.g., $m = 1$, $j = 0$, $p = q = r = 2$, and $n \geq 2$. Such cases naturally arise in studying the energy or Dirichlet form of the Laplace operator. Using a different approach, Lin [1986] required that $m \geq j$ and that $m - j - n/r \neq k$, where $k$ is a non-negative integer, thus excluding certain important cases.

Restrictions of this kind seem to be more closely connected to the technicalities of the proofs than to the fundamental validity of the inequalities. Our aim in the present paper therefore is to show how they can be replaced by more satisfactory alternative conditions. Basically we will derive (1.1) in the setting $n > 1$ if either $\{m \geq j + n\}$ or $\{m > j$ and $p \leq r\}$.

Towards this goal our strategy will be to prove two new versions (Lemmas 2.4 and 2.5 of Section 2) of Lemma 2.4 of [Brown and Hinton 1987b]. Each lemma will require new definitions of certain integral averages $S^r_{\ell,\ell}(N,W,g)$, $S^r_{\ell,\ell}(N,P,h)$ which differ from those originally given in [Brown and Hinton 1987b]. These are given in Definitions 2.2-2.3 below. Also we require at times a number of new technical conditions on the weights (hypotheses (H1)-(H11) below). However, when the necessary adjustments are made, many of the results of Brown and Hinton [1987b] concerning sum inequalities continue to hold in the new setting. These results are summarized in Section 3.

We employ standard notation for the spaces of smooth functions and Sobolev spaces; e.g., see Adams [1975]. $L^p(\Omega)$ denotes the space of equivalence classes of measurable functions $u$ with norm $(\int_\Omega |u|^p)^{1/p}$. The subscript "loc" means a local property. For $1 \leq \eta \leq \infty$, $\eta^\prime$ is the extended real number such that $1/\eta + 1/\eta^\prime = 1$. We use

$$|D^\alpha u(x)| = \sum_{|\alpha| = m} |D^\alpha u(x)|$$

where $\alpha = (\alpha_1, ..., \alpha_n)$ is a multi-index. $Q_n$ denotes a closed cube in $R^n$ with sides of length $L$ parallel to the coordinate axes. $C_x$, for $x \in Q_n$, denotes a spherical cone with vertex $x$, radius $L/2$, central angle independent of $x$, and $C_x \subset Q_n$. Define
\[
E = \left\{ u \in W_{\text{loc}}^{m,r} (\Omega) : \int_{\Omega} W |u|^q < \infty, \int_{\Omega} P |D^m u|^r < \infty \right\}
\]
with a norm given by
\[
\|u\|_E = \left[ \int_{\Omega} W |u|^q \right]^{1/q} + \left[ \int_{\Omega} P |D^m u|^r \right]^{1/r}
\]

Additionally let
\[
D'(\Omega) := \left\{ u \in C_c^m (\Omega) : \int_{\Omega} W |u|^q < \infty, \int_{\Omega} P |D^m u|^r < \infty, \text{ and } u \text{ has support in } \Omega \right\}.
\]

Note that the support of \( u \in D'(\Omega) \) need not be compact unless \( \Omega \) is bounded. In [Brown and Hinton 1987b] it was shown that the relative topological closure of \( D'(\Omega) \) in \( E \) with respect to the norm \( \| \|_E \) is a Banach space which we denote by \( D(\Omega) \).

In addition to the basic assumptions of sign and measurability on the weights \( N, W, P \), we shall consider the following hypotheses when \( n > 1 \). The set \( \tilde{\Omega} \) contains \( \Omega \) and is defined below.

(H1) \( m \geq j + n \).

(H2) \( m > j \) and \( p \leq r \).

(H3) \( m - j > n/r \).

(H4) \( N, W, P \in L_{\text{loc}}^1 (\tilde{\Omega}) \).

(H5) \( N \) is locally bounded on \( \tilde{\Omega} \).

(H6) \( N^{-1} \) is locally bounded on \( \tilde{\Omega} \).

(H7) \( W^{-r/q-q} \in L_{\text{loc}}^1 (\tilde{\Omega}) \) for \( q > r \); \( P^{-q/(r-q)} \in L_{\text{loc}}^1 (\tilde{\Omega}) \) for \( r > q \).

(H8) \( P^{-1} \) is locally bounded on \( \tilde{\Omega} \) for \( q \geq r \); \( W^{-1} \) is locally bounded on \( \tilde{\Omega} \) for \( q \leq r \).

(H9) \( P^{-1} \) is locally bounded on \( \tilde{\Omega} \).

(H10) \( W^{-q/q} \in L_{\text{loc}}^1 (\tilde{\Omega}) \) for \( q > 1 \); \( W^{-1} \) is locally bounded on \( \tilde{\Omega} \) for \( q = 1 \).

(H11) \( f \) is a positive continuous function on \( \mathbb{R}^n \).
2. Lemmas

Fundamental to the theory of Brown and Hinton [1987b] were the following two lemmas.

**Lemma 2.1.** There exists a constant $K$ depending only on $n$ and $m$ such that if $u \in C^m(\bar{Q}_n)$, $0 \leq j \leq m-1$, and $x \in \bar{Q}_n$, then

$$|D^ju(x)| \leq K \left\{ L^{-j} \int_{\bar{Q}_n} |u| + L^{m-j} \int_{\bar{Q}_n} |D^m u| + \int_{\bar{Q}_n} |x-y|^{m-1-j} dy \right\}. \quad (2.1)$$

**Lemma 2.2.** Assume (H3), (H4), (H9)-(H11) hold, $1 \leq p, q, r < \infty$ and $1 < \eta, \xi \leq \infty$. Then on any cube $Q_n = Q_{i,\varepsilon}$ we have for $u \in D(\Omega)$ the inequality

$$\int_{\bar{Q}_n} |D^j u|^p \leq K \left\{ \epsilon^{-\alpha} G_{i,\varepsilon}(g, \eta) S_{i,1,\varepsilon}^{\eta,\eta}(N, W, g) \left( \int_{\bar{Q}_n} |u| \right)^{p/q} \right\} \quad (2.2)$$

$$+ \epsilon^b G_{i,\varepsilon}(h, \xi) S_{i,1,\varepsilon}^{\eta,\eta}(N, P, h) \left( \int_{\bar{Q}_n} |D^m u| \right)^{p/r},$$

where $a = p(j + n/q - n/\eta) \eta'$ and $b = p(m - j - n/r + n/\xi) \xi' \eta$. Moreover, the constant $K$ depends only on $p, n$, and $m$. We regard $u$ as being defined on $R^n$ by setting $u(x) = 0$ on $R^n \backslash \Omega$.

In Lemma 2.2 $Q_{i,\varepsilon}$ is defined for $t \in \Omega$ and $f$ as in (H11) by

$$Q_{i,\varepsilon} = \bigvee_{i=1}^n \left[ t_i - \epsilon f(t)/2, t_i + \epsilon f(t)/2 \right].$$

Thus $t$ is the center of the cube $Q_{i,\varepsilon}$ and its edge is of length $L = \epsilon f(t)$. The quantities $G_{i,\varepsilon}(g, \eta)$, $G_{i,\varepsilon}(h, \xi)$, $S_{i,1,\varepsilon}^{\eta,\eta}(N, W, g)$, and $S_{i,1,\varepsilon}^{\eta,\eta}(N, P, h)$ were defined by Brown and Hinton [1987b]. For the sake of comparison with later definitions, we repeat them here.

**Definition 2.1.** For $Z$, $U$, $g$, $h$ positive measurable functions on $Q = Q_{i,\varepsilon}$, $1 \leq u$, $v < \infty$, and $1 < \eta, \xi \leq \infty$, let

$$T_{i,\varepsilon}^{u,v}(Z) := \begin{cases} \sup_{x \in Z^{-1}(x), x \in Q} \left( Z^{-1}(x) \right)^{u/v} \quad \text{if } v = 1, \\ L^{-n} \int_{Z^{-1}(x)}^u \left( Z^{-1}(x) \right)^{u/v} \quad \text{if } v > 1 \end{cases}$$

$$\tilde{T}_{i,\varepsilon}^{u,v}(Z) := \sup_{x \in Q} \left( L^{-n} \int_{\bar{Q}} |Z^{-1}(y)\theta_x(y)(m-n-j)^{u/v} dy \right)^{u/v}, \quad v > 1,$$

where $\theta_x(y) = |x-y|/L$. 

42
\[ S_{1,1,\xi}^{\eta,\mu}(U,Z,g) = f(t)^{-u(j+n/v-n/u\eta)} \left[ L^{n-\eta} \int_{\Omega} (gU)^{\eta} \right] T_{1,\xi}^{\mu,\nu}(Z) \]

\[ S_{2,1,\xi}^{\eta,\mu}(U,Z,h) = f(t)^{u(m-j-n/v-n/u\xi)} \left[ L^{n-\eta} \int_{\Omega} (hU)^{\xi} \right] \{ T_{1,\xi}^{\mu,\nu}(Z) + (1-\delta_{\eta,1}) \hat{T}_{1,\xi}^{\mu,\nu}(Z) \} , \]

where \( \delta_{\eta,1} \) is the Kronecker delta,

\[ G_{1,\xi}(g,\eta) = \begin{cases} \left[ \int_{\Omega} g^{-\eta} \right]^{1/\eta} & \text{if } 1 < \eta < \infty, \\ \sup_{x \in \Omega} g(x)^{-1} & \text{if } \eta = \infty, \end{cases} \]

with a similar definition for \( G_{1,\xi}(h,\xi) \).

The proof of Lemma 2.2 is an exercise in Hölder's inequality combined with the definitions. The significance of the assumptions (H3) \( m > j + n/r \) or (H9) \( P^{-1} \) is locally bounded is that a calculation shows that for \( r > 1 \),

\[ T_{1,\xi}^{p,r}(P), \hat{T}_{1,\xi}^{p,r}(P) \leq K_0 M^{p/r} , \]

where \( M \) is a bound on \( P^{-1} \) and \( K_0 \) depends only on the angle of the cone and on \( m,n,j,r \). Thus, effectively, the quantity \( S_{2,1,\xi}^{\eta,\mu} \) depends only on \( T_{1,\xi}^{p,r} \) and not on \( \hat{T}_{1,\xi}^{p,r} \). It is this rather artificial feature that made calculations with specific weights possible in [Brown and Hinton 1987b].

We now state without proof three lemmas that develop two distinct methods by which the assumption (H3) can be weakened. Throughout, \( K \) denotes a constant, whose value may change from line to line, whose value is independent of \( u, \Omega, N, W, P, Q_n \) but may depend on the parameters \( p,q,r,m,j,n \).

**Lemma 2.3.** Assume (H1), (H5), (H6), and \( 1 \leq p, v < \infty \). Then for \( u \in D(\Omega) \) the following inequality holds:

\[
\int_{\Omega} N \left| D^j u \right|^p \leq K \|N\|_{\infty, Q_n} \|N^{-1}\|_{\infty, Q_n}^{p/v} \left[ \int_{Q_n} N \left| u \right|^v \right]^{p/v} + L_p^{(m-j-n/v+n/p)} \left[ \int_{Q_n} N \left| D^m u \right|^v \right]^{p/v} .
\]

(2.3)

The method of proof is to first take \( N = 1 \) and use an induction argument. The general case follows from the \( N = 1 \) case by writing \( 1 = NN^{-1} \) in the integrals and using the appropriate \( \| \|_{\infty, Q_n} \) bounds.
To use Lemma 2.3 to construct an adequate replacement for Lemma 2.2 without the condition \( m > j + n/r \), we need to change the definitions of \( S_{2,1,\xi}^{p,q} \) and \( S_{2,1,\xi}^{p,r} \).

**DEFINITION 2.2.** Let \( 1 \leq p, q, r < \infty \), \( 1 < \eta, \xi \leq \infty \), \((H11)\) hold, and \( g, h \) be positive measurable locally bounded functions on \( Q_n = Q_{i,\xi} \). For \( q > r \), define

\[
S_{2,1,\xi}^{p,q} (N,W,g) := \| N \|_{\infty, Q_i} \| N^{-j} \|_{\infty, Q_i} \| g \|_{\infty, Q_i} f(t)^p (j+n/q-n/\eta)^{p(q-r)/q} \left[ L^{-n} \int N q^{(q-r)W^{(r-q)}} \right]^{\frac{p(q-r)}{q}}.
\]

For \( q < r \), define

\[
S_{2,1,\xi}^{p,q} (N,P,h) := \| N \|_{\infty, Q_i} \| N^{-j} \|_{\infty, Q_i} \| h \|_{\infty, Q_i} f(t)^p (m-j-n/r+n/\xi)^{p(r-q)/q} \left[ N P^{(r-q)q} \right]^{\frac{p(r-q)}{q}}.
\]

For \( q = r \), define \( S_{2,1,\xi}^{p,q} (N,W,g) \) as in the \( q < r \) case and \( S_{2,1,\xi}^{p,q} (N,P,h) \) as in the \( q > r \) case.

We can then state the following lemma.

**LEMMA 2.4.** Assume \((H1), (H5)-(H8), (H11)\), and Definition 2.2. Then on any cube \( Q_n = Q_{i,\xi} \), (2.2) is true for all \( u \in D(Q) \).

The proof uses Lemma 2.3. For example, take the case \( q > r \). Let \( v = r \) in Lemma 2.3, and apply Hölder’s inequality to \( \int N/u^r = \int (NW^{-1/q})(W^{(1/q)}W^r) \) with conjugate indices \( q/r \) and \( q/(q-r) \). Simplifying and using

\[
L^n = \int g g^{-1} \leq G_{i,\xi}(g,\eta)L^{n/\eta} \| g \|_{\infty, Q_i},
\]

with a similar inequality for \( G_{i,\xi}(h,\xi) \) completes the proof.

**DEFINITION 2.3.** Let \( 1 \leq p, q < \infty \), \( 1 < r \), \( \eta, \xi \leq \infty \), \((H11)\) hold, \( g, h \) be positive measurable locally bounded functions on \( Q_n = Q_{i,\xi} \), and \( 0 < \delta < 1 \). Let \( T_{i,\xi}^{p,r}(P) \), \( S_{2,1,\xi}^{p,q} (N,W,g) \) be as in Definition 2.1. Define

\[
T_{i,\xi}^{p,r}(P) := \sup_{x \in Q_i} \left[ L^{-n} \int P^{-r/\eta} (y) \Theta_x(y)^{-\delta} dy \right]^{\frac{p/r}{\delta}}.
\]

where \( \Theta_x(y) = |x-y|/L \).

\[
S_{2,1,\xi}^{p,q} (N,P,h) := \| h \|_{\infty, Q_i} f(t)^p (m-j-n/r+n/\xi)^{p(r-q)/q} \left[ L^{-n} \int N T_{i,\xi}^{p,r}(P) + [U_{i,\xi}(N)]^{p/r} \left[ L^{-n} \int N \right]^{(r-q)/q} \right]^{\frac{p/r}{q}}.
\]

where
Lemma 2.5. Assume (H2), (H4), (H10), (H11), and Definition (2.3). Then on any cube \( Q_n = Q_{t, \varepsilon} \), (2.2) holds for all \( u \in D(\Omega) \).

The proof involving the \( S_{t, \varepsilon} \) is the same as Lemma 2.4 of [Brown and Hinton 1987b]. The integral \( \int |D^m u| \) in (2.1) is also handled similar to that of [Brown and Hinton 1987b] but uses (2.4). The remaining integral in (2.1) also uses (2.4); however, the key step is an interchange of order of integration in

\[
\int_{Q_n} \left\{ \int_{Q_n} N(x) |D^m u(y)|^r |x-y|^{m-j} \varepsilon^{-n} \, dy \right\} \, dx.
\]

3. Sum Inequalities

We now use either Lemma 2.4 or Lemma 2.5 to replace Lemma 2.2. The condition \( m > j + n/r \) is thus replaced with either \( m \geq j + n \) or \( m > j \) and \( p \leq r \).

The derivation of sum inequalities follows exactly the same pattern as in [Brown and Hinton 1987b] and are stated in Theorem 3.1 below. The new lemmas are used to prove the sum inequality on a generic cube \( Q_{t, \varepsilon} \). Since \( S = \{ Q_{t, \varepsilon} : t \in \Omega \} \) covers \( \Omega \) and every point \( t \) is the center of some cube in \( S \), the Besicovitch covering theorem (cf. Brown and Hinton [1987b, Proposition 3.1]) may be used to extract finitely many families \( \Gamma_1, \ldots, \Gamma_s \) of mutually disjoint cubes to cover an arbitrary bounded set \( A \) in \( \Omega \). The covering theorem guarantees the critical fact that the number \( s \) depends only on \( n \) and not on \( A \). Therefore, the inequality is true on each \( \Gamma_i \) and therefore on \( U \Gamma_i \supseteq A \) (the case \( p \) is not greater than or equal to \( q, r \) requires some additional technical complications). Since \( A \) can be chosen so that the integrals \( \int_{\Omega \setminus A} W |u|^{q} \), \( \int_{\Omega \setminus A} P |D^m u|^{r} \) are arbitrarily small, the inequality will hold for \( u \in D(\Omega) \). To state this theorem precisely, we consider the following conditions where \( \Gamma \) denotes a set of positive real numbers, usually \( \Gamma = (0, \varepsilon_0) \), \( \varepsilon_0 \leq \infty \), and \( \Omega = U \{ Q_{t, \varepsilon} : t \in \Omega \} \).

(C1) \( \sup_{t \in \Omega, \varepsilon \in \Gamma} S_{t, \varepsilon}^{W, r, \varepsilon} (N, W, g) < \infty \).

(C2) \( \sup_{t \in \Omega, \varepsilon \in \Gamma} S_{t, \varepsilon}^{P, r, \varepsilon} (N, P, h) < \infty \).

(C3) \( \int_{\Omega} g^{-n} \, d\mu < \infty \).
\( (C4) \int h^{-\xi} < \infty \).

\( (C5) \) \((H3), (H4), (H9), (H10), \) Definition 2.1.

\( (C6) \) \((H1), (H5)-(H8), \) Definition 2.2.

\( (C7) \) \((H2), (H4), (H5), (H10), \) Definition 2.3.

**Theorem 3.1.** Assume \((H11), (C1), \) and \((C2)\) hold with \( \Gamma = (0, \epsilon_0) \) in \((C1), (C2)\). Set \( \eta = q/(q-p) \) if \( q > p \) and \( \xi = r/(r-p) \) if \( r > p \). Set \( \eta = \infty \) and \( g = 1 \) if \( p \geq q \) and \( \xi = \infty \), \( h = 1 \) if \( p \geq r \). Then the sum inequality \((1.1)\) holds for all \( u \in D(\Omega) \) in the following cases.

(i) \( p \geq \max\{q,r\} \), \((C5)\) or \((C6)\) holds (or \((C7)\) if \( p = r\)), \( \theta = p(m-j-n/r+n/p) \).

(ii) \( p < \min\{q,r\} \), \((C3)-(C4)\) hold together with one of \((C5), (C6), \) \((C7)\), \( \theta = p(m-j) \), \( \phi = p/j \).

(iii) \( q \leq p < r \), \((C3)\) holds together with one of \((C5), (C6), \) \((C7)\), \( \theta = p(m-j-n/r+n/p) \), \( \phi = p/j \).

(iv) \( r < p < q \), \((C4)\) holds together with \((C5)\) or \((C6)\), \( \theta = p(m-j) \), \( \phi = p(j+n/q-n/p) \).

**Example 3.1.** Let \( \Omega \) be arbitrary and set \( N(t) = (1 + |t|)^\beta \), \( W(t) = (1 + |t|)^\gamma \), and \( P(t) = (1 + |t|)^\alpha \). Take \( \Gamma = (0,1) \), \( f(t) = (1 + |t|)^\Delta_1 \) with \( \Delta_1 \leq 1 \), \( g(t) = (1 + |t|)^\Delta_2 \), and \( h(t) = (1 + |t|)^\Delta_3 \). Calculations similar to those of Brown and Hinton [1987b] show that \((C1)-(C4)\) are equivalent, respectively, to

\( (C1)' \beta \leq \gamma p/q + \Delta_1 p(j+n/q-n/p) \eta^- \).

\( (C2)' \beta \leq \alpha p/r - \Delta_1 p(m-j-n/r+n/p) \xi^- - \Delta_3. \)

\( (C3)' \eta \Delta_2 > n. \)

\( (C4)' \xi \Delta_3 > n. \)

Critical to the calculation is the fact that the ratio \((1 + |t|)/(1 + |s|)\) is bounded above and below by positive numbers for all \( t \in \mathbb{R}^n \) and \( s \in Q_{t,c} \).

Once we have shown that \((C1)-(C4)\) are equivalent to \((C1)'-(C4)'\), the remainder of the analysis, i.e., choosing \( \Delta_1, \Delta_2, \Delta_3 \), follows as in Example 1 of [Brown and Hinton 1987b]. For example, in case (i) of Theorem 3.1 we take \( \xi^- = \eta^- = 1 \) and \( \Delta_2 = \Delta_3 = 0; \Delta_1 \) is chosen by requiring that the right-hand sides of \((C1)'\) and \((C2)'\) be equal. It will then follow that \((1.1)\) holds when
\[ p \geq \max\{q, r\} \text{ and } m > j + n/r \text{ or } m \geq j + n \text{ or when } \{m > j \text{ and } p = r \geq q\} \]

if

\[ \Delta_1 := (\alpha/r - \gamma/q)/(m + n/q - n/r) \leq 1 \]

and

\[(\beta/p)(m - n/r + n/q) \leq (\gamma/q)(m - j - n/r + n/p) + (\alpha/r)(j + n/q - n/p). \tag{3.1} \]

Similar results may be derived for the other three cases of \( p \).

Inequality (1.1) determines a map \( J: D(\Omega) \to L^p(\Omega) \) defined by \( Ju = N^{1/p} D^\alpha u \), where \( \alpha \) is a fixed multi-index with \( |\alpha| = j \). Sufficient conditions were given in [Brown and Hinton 1987b, Theorem 4.1] that \( J \) be compact. The argument uses a Rellich-Kondrashov theorem that depends on the assumption \( m > j + n/r \). Since (1.1) can be obtained without this assumption, we have the following theorem.

**Theorem 3.2.** Assuming (C7) in Theorem 3.1, the mapping \( J \) determined in cases (i)-(iv) is compact if, additionally,

\[ m - j < n/r \text{ and } 1 \leq p \leq nr/(n - (m - j)r) \tag{3.2} \]

or

\[ m - j = n/r \text{ and } 1 \leq p < \infty \tag{3.3} \]

and the condition

\[(C1)^* \text{ For each } \epsilon \in (0, 1), \]

\[ \lim_{\|t\| \to \infty, t \in \Omega} S_{t, f, \epsilon}^\alpha \eta (N, W, g) = 0 \]

is substituted for (C1).

**Proof.** We give only the reasoning that differs from that in the proof of Theorem 4.1 of [Brown and Hinton 1987b]. The condition (3.2) or (3.3) implies [Kufner, John, and Fučík 1977, p. 293] that the identity map is a compact operator from \( W^{m,r}(B) \) to \( L^p(\Omega) \), where \( B \) is an open cube centered at the origin. In the language of the proof of Theorem 4.1 of [Brown and Hinton 1987b], this in turn shows that a subsequence of \( \{D^\alpha u_\epsilon, s_r(\Omega)\} \) converges in \( L^p(S(R)) \). Since \( N \) is locally bounded, \( \{N^{1/p} D^\alpha u_\epsilon, s_r(\Omega)\} \) also converges in \( L^p(S(R)) \).

The difference between (C1) and (C1)* often means replacing an inequality by a strict inequality, e.g., as in (3.1).
References


ASYMPTOTIC LINEARITY OF SOLUTIONS OF SECOND-ORDER SYSTEMS OF DIFFERENTIAL EQUATIONS

Shaozhu Chen
Department of Mathematics
Shandong University
Jinan, Shandong
People's Republic of China

Abstract

We consider the n-dimensional system

$$U''(t) + p(T)U = 0,$$

where $p$ is an $n \times n$ Hermitian matrix-valued function. Necessary and/or sufficient conditions for all solutions to be asymptotic, in varying degrees, to linear matrix functions $tB + A$ are obtained. Systems with linear or nonlinear perturbations are also discussed.

1. Introduction

We are concerned mainly with the self-adjoint system of differential equations

$$U''(t) + p(t)U(t) = 0, \quad t \geq a,$$

where $p$ is a locally integrable, $n \times n$ Hermitian matrix-valued function on $a \leq t < \infty$. An $n \times n$ matrix function $U(t)$ is said to be a solution of (1.1) if $U'(t)$ is absolutely continuous and (1.1) is satisfied a.e. on $[a, \infty)$. A solution $U$ of (1.1) is said to be prepared if it satisfies

$$U^*U - UU^* = 0.$$

A prepared solution $U$ is oscillatory if $U(t_n)$ is singular on some sequence $t_n$, $n = 1, 2, ..., t \to \infty$. Otherwise, $U$ is non-oscillatory. In view of the Sturm-type separation theorem of Morse [1934] the system (1.1) is oscillatory; that is, every prepared solution is oscillatory, provided (1.1) has a prepared solution that is oscillatory. The same is true for the non-oscillation of (1.1).

In the particular case $p(t) = 0$, (1.1) has a pair of fundamental solutions $U_1 = I$, the $n \times n$ unit matrix, and $U_2 = tI$, and any solution $U$ of (1.1) can be expressed by a linear matrix function $A + tB$ satisfying

$$t^{-1}U(t) \to B,$$

$$U'(t) \to B,$$

$$U(t) - tB \to A \quad \text{as} \ t \to \infty.$$

It is natural to pose the question of how small the function $p$ is required to be so that every solution of (1.1) is asymptotic, in the sense described in each of (1.3)-(1.5), to a linear matrix function. By virtue of the superposition principle, the above properties (1.3)-(1.5) are certainly
assured by the following requirements, respectively: (1.1) has a pair of prepared solutions \( U_1 \) and \( U_2 \) satisfying

\[
U_1(t) \to I, \quad t^{-1}U_2(t) \to I, \quad (1.6)
\]

\[
U_1(t) \to I, \quad U_2(t) \to I, \quad (1.7)
\]

\[
U_1(t) \to I, \quad U_2(t) - tI \to C \quad \text{(a constant matrix) as } t \to \infty. \quad (1.8)
\]

The purpose of this paper is to find the sufficient conditions, sometimes necessary conditions as well, imposed on the integral smallness of the function \( p(t) \) under which (1.1) possesses a pair of solutions satisfying each of (1.6)-(1.8).

After this introduction, we concentrate on a thorough study of system (1.1) in the following six sections and display the main results in Section 4.

Our fundamental tool is based on the Riccati techniques which have been extensively applied to studies of the oscillation theory for linear differential equations by Wintner [1951], Wong [1969] and others in the scalar case, Eliason and Mary [1978], and Mary [1987] in the matrix case. Note that the prepared solution \( U_1 \) in (1.6)-(1.8) is non-oscillatory, and so it is necessary for the existence of \( U_1 \) that system (1.1) be non-oscillatory. In Section 3, several well-known results in oscillation theory are viewed as lemmas for further discussions.

Section 8 studies the nonhomogeneous system

\[
U'' + p(t)U = q(t), \quad (1.9)
\]

where \( q \) is an \( n \times n \) matrix-valued function. Section 9 considers the terminal value problems for differential systems, namely, the existence and uniqueness of solutions that are asymptotic to a given linear matrix function. In Section 10 we extend our results to the nonlinear system

\[
U'' + p(t)U = f(t,U), \quad (1.10)
\]

where \( f \) is a matrix-valued function.

Many authors, e.g., Bellman [1963], Hartman and Wintner [1953], and Hille [1948], have made thorough studies of the asymptotic linearity of solutions for linear scalar equations. In the case \( n = 1 \), Bellman [1963, p. 114] proved that (1.1) has two solutions satisfying (1.6) if \( \int_0^\infty t |p(t)| dt < \infty \) (in the scalar case the solutions \( U_1 \) and \( U_2 \) in (1.6)-(1.8) are not required to be prepared); if \( p(t) \) remains of constant sign for all large \( t \), then this condition is also necessary, and in this case a necessary and sufficient condition \( \int_0^\infty t^2 |p(t)| dt < \infty \) that (1.1) have a pair of solutions satisfying (1.8) was obtained by Hille [1948]; Hartman and Wintner [1953] established the necessary and sufficient conditions for the existence of a pair of solutions satisfying either (1.6) or (1.7).
For the asymptotic linearity of solutions of nonlinear scalar differential equations, we refer to the work of Chen [1987] and the references there.

Hallam [1970] considered the perturbed vector equation
\[ y'' = A(t)y + f(t,y), \tag{1.11} \]
where \( y \) and \( f \) are \( n \)-vector functions and \( A \) is an \( n \times n \) matrix function, and obtained sufficient conditions under which (1.11) has a unique solution that is asymptotic to a given linear vector function.


The first two theorems in Section 4 are closely related to the results of Hartman and Wintner, and the results in Sections 8-10 are generalizations of the work of Chen and Hallam, but our approach is completely different from theirs.

2. Basic Hypotheses

It will be convenient to designate by \( L[a, \infty) \) the set of all locally integrable matrix functions \( f(t) \) on \([a, \infty)\) for which the limit
\[ \int_a^\infty f(t)dt = \lim_{T \to \infty} \int_a^T f(t)dt \]
exists finitely. For simplicity in notation, we shall write \( f \in L \) if \( f \in L[b, \infty) \) for some \( b \geq a \).

The symbol \( \| \cdot \| \) denotes some suitable matrix norm, and \( tr(\cdot) \) the matrix trace. If \( f(t) \) is a non-negative definite Hermitian matrix function on \([b, \infty)\), it is easy to see that \( f \in L \) if and only if either its norm or its trace is integrable over \([b, \infty)\). For the purpose of discussions below, we list two properties of the trace operator. If \( f \) and \( g \) are \( n \times n \) matrices, then
\[ tr(f + g) = tr(f) + tr(g), \tag{2.1} \]
\[ tr(fg) = tr(gf). \tag{2.2} \]

The basic hypotheses on \( p(t) \) under which the work will be carried out are

\((H1)\) \( p \in L \),

\((H2)\) \( P \in L, P(t) = \int_t^\infty p(s)ds \) and the equation \( Z' = PZ \) has a solution \( Y(t) \) such that \( Y(t) \to l \) as \( t \to \infty \).
There are a number of conditions assuring (H2), for example,

(H2)$_1$ \[ \int_a^\infty |P(t)| \, dt < \infty; \]

(H2)$_2$ \[ P \in L \text{ and for any } t \geq a, \, P(t) \int_a^t P(s) \, ds = \left[ \int_a^t P(s) \, ds \right] P(t); \]

(H2)$_3$ \[ P \in L \text{ and } \int_a^\infty |Q(s)P(s)| \, ds < \infty, \text{ where } Q(t) = \int_a^t P(s) \, ds; \]

(H2)$_4$ \[ P, QP \in L \text{ and } \int_a^\infty |R(s)P(s)| \, ds < \infty, \text{ where } R(t) = \int_a^t (QP)(s) \, ds. \]

For the first three conditions listed, we refer to Reid [1971, p. 415]. For the last one, a standard fixed-point argument asserts the existence of a unique solution of the integral equation

\[ Z(t) = (I + Q + R)^{-1}(t) - (I + Q + R)^{-1}(t) \int_a^\infty (RZP)(s) \, ds, \]

which proves (H2).

3. Riccati Techniques and Preliminary Lemmas

We shall introduce some well-known oscillation results for the system (1.1) under the following assumption on $p$:

\[ \lim \inf_{t \to \infty} (t-a)^{-1} \int_a^t p(r) \, dr > -\infty, \tag{3.1} \]

where $\pi_1, \pi_2, \ldots, \pi_n$ are a set of unit vectors spanning $\mathbb{C}^n$. For details of proof, the interested reader is referred to Eliason and Mary [1978] and Mary [1987] and the references cited in them.

**Lemma 1.** If (3.1) is valid, then the system (1.1) is non-oscillatory if and only if there exists an $n \times n$ Hermitian matrix function $W$ defined on $[b, \infty)$, for some $b \geq a$, satisfying

\[ W'(t) + W^2(t) + p(t) = 0 \quad \text{a.e. on } [b, \infty) \tag{3.2} \]

and $W^2 \in L([b, \infty))$.

Note that (H1) evidently implies (3.1) and that if (1.1) is non-oscillatory and $U$ is a non-oscillatory prepared solution, then $W = UU^{-1}$ is Hermitian and satisfies (3.2). Therefore, under the hypothesis (H1), $W$ will satisfy the Riccati integral equation.
\[ W(t) = P(t) + \int_{t}^{\infty} W^2(s) ds, \quad t \geq b. \]  

(3.3)

In what follows, for simplicity, we shall assume (H1) instead of (3.1). Let the matrix function \( X(t;s) \) be the solution of the initial value problem

\[ \frac{d}{ds} X(t;s) = P(t)X(t;s), \quad X(s;s) = I. \]  

(3.4)

Set \( V = \int_{t}^{\infty} W^2(s) ds \) and

\[ \tilde{P}(t) = \int_{t}^{\infty} X^*(s;t)P^2(s)X(s;t) ds, \]  

(3.5)

whenever the integral converges.

Observing that \( X(t;s) \) also satisfies

\[ \frac{d}{ds} X(t;s) = -X(t;s)P(s), \quad X(t;t) = I, \]  

(3.6)

we have the next result.

**Lemma 2.** Let (H1) hold. The system (1.1) is non-oscillatory if and only if \( \tilde{P}(t) \) exists finitely and there exists a Hermitian matrix function \( \tilde{V}(t) \) on \([b, \infty), for some \( b \geq a \), satisfying

\[ \tilde{V}(t) = \tilde{P}(t) + \int_{t}^{\infty} X^*(s;t)\tilde{V}^2(s)X(s;t) ds, \quad t \geq b. \]  

(3.7)

Finally we set

\[ \tilde{P}(t) = \int_{t}^{\infty} X^*(s;t)\tilde{P}^2(s)X(s;t) ds, \]  

(3.8)

whenever the integral exists.

**Lemma 3.** Let (H1) hold. If the integral \( \tilde{P}(t) \) exists finitely on \([b, \infty), for some \( b \geq a \), and

\[ \tilde{P}^2(t) \leq \frac{1}{16} \tilde{P}^2(t) \quad \text{on} \quad [b, \infty), \]  

(3.9)

then (1.1) is non-oscillatory.

**4. Main Results**

We are now ready to give our main results for the self-adjoint system (1.1) and leave their proofs until later.
We first formulate a characterization of the existence of a pair of prepared solutions of (1.1) with the asymptotic property (1.6).

**THEOREM 1.** Suppose that (H1) and (H2) hold. The system (1.1) has a pair of prepared solutions $U_1$ and $U_2$ satisfying (1.6) if and only if

(A1) $\int_a^\infty tP^2(t)dt$ exists finitely.

A straightforward calculation yields

$$tQ^2(t) - aQ^2(a) = Q^2(t) - Q^2(a) + 2\int_a^t sP^2(s)ds + \int_a^t (Qp + pQ)(s)ds.$$  \hspace{1cm} (4.1)

If the limit of the last integral in (4.1) as $t \to \infty$ exists finitely, then $tr \left[ \int_a^t sP^2(s)ds \right]$ is bounded above, which implies (A1), since otherwise $t[trQ^2(t)]'$ would remain positive for all large $t$, and hence $trQ(t)$ would not tend to zero as $t \to \infty$. Thus, we get the following result involving a non-absolute integrability condition.

**COROLLARY.** Let (H1)-(H2) hold. If

(A1)$_1 = \int_a^\infty tQ(t)p(t)dt$ exists $\forall t$ finitely,

then system (1.1) has a pair of solutions satisfying (1.6).

Next, we characterize the existence of two solutions with asymptotic behavior given by (1.7).

**THEOREM 2.** Let (H1)-(H2) hold. Then the system (1.1) has a pair of solutions $U_1$ and $U_2$ satisfying (1.7) if and only if (A1) and

(A2) $\lim_{t \to \infty} tP(t) = 0$.

Hartman and Wintner [1953] established Theorems 1 and 2 in the case $n = 1$ using a different method. It should also be mentioned that (A1) does not imply (H2)$_1$ and that the conditions (A1) and (A2) are, of course, independent. But (A2) can be replaced by the weaker condition that $\lim_{t \to \infty} tP(t)$ exists finitely since the latter is, under (H2), equivalent to (A2).

To ensure that every solution of (1.1) satisfies (1.8), we need more assumptions:

(A3) $\int_a^\infty t^2P^2(t)dt$ exists finitely,
(A4) \( Q \in L, Q(t) = \int_t^\infty p(s)ds, \)

and

(A5) \( \lim_{t \to \infty} tQ(t) \) exists finitely.

**Theorem 3.** If (H1) and (A2)-(A5) hold, then the system (1.1) has a pair of solutions \( U_1 \) and \( U_2 \) satisfying (1.8).

In the above theorem we have not assumed (H2) since (A3) obviously implies (H2). A result analogous to the corollary of Theorem 1 can be obtained by observing a similar identity

\[
i^2 Q^2(t) - a^2 Q^2(a) = 2tQ^2(t) - 2\int_a^t Q^2(s)ds \\
+ 2\int_a^t s^2 P^2(s)ds + \int_a^t s^2(Qp + pQ)(s)ds - 2aQ^2(a)
\]

(4.2)

and the fact that (A5) guarantees the integrability of \( Q^2 \) over \([a, \infty)\).

**Corollary.** Theorem 3 remains true if (A3) is replaced by

(A3)_1 \( \int_a^\infty t^2 Q(t)p(t)dt \) exists finitely.

To indicate the scope of our results, we first compare them with Theorem 1 of Hallam [1970]. To do this, let \( f = 0 \) and \( A(t) = p(t) \) in (1.11), dropping the assumption that \( A(t) \) is Hermitian, and view (1.1) as the associated matrix system of (1.11). Then Hallam's theorem becomes two statements of the following form.

The system (1.1) has a pair of solutions satisfying (1.6) if

\[
\int_0^\infty |P(t)| dt < \infty
\]

(4.3)

and

\[
\int t |Q(t)p(t)| dt < \infty.
\]

(4.4)

The system (1.1) has a pair of solutions satisfying (1.8) if

\[
\int_0^\infty |P(t)| dt < \infty
\]

(4.5)

and
\[ \int_0^\infty |Q(t)p(t)| \, dt < \infty. \] (4.6)

For simplicity of comparison we suppose further that \( n = 1 \); then the corollaries of Theorems 1 and 3 show that Hallam's theorem is true even if the integrals in (4.3), (4.4), and (4.6) converge conditionally, not absolutely, and (4.5) is replaced by the convergence of \( \int_0^\infty p(t) \, dt \) and \( \int_0^\infty Q(t) \, dt \), which is usually much weaker than (4.5).

We conclude this section with some examples.

**Example 1.** \( n = 1, p(t) = t^{-\alpha} \log^\gamma t \cos t^\beta, \alpha, \beta, \gamma > 0 \). Integration by parts gives

\[
P(t) = -\beta^{-1} t^{1-\alpha-\beta} \log^\gamma t \left[ \sin t^\beta + O(t^{-\beta}) \right],
\]
\[
Q(t) = -\beta^{-2} t^{2-2\alpha-2\beta} \log^\gamma t \left[ \cos t^\beta + O(t^{-\beta}) \right] \quad \text{as } t \to \infty.
\]

By expressing all the conditions in Theorems 1-3 in terms of ranges of parameters \( \alpha, \beta, \gamma \), we have first that (1.1) possesses two solutions with property (1.7) if and only if

\[
\alpha + \beta > 2, \text{ or } \alpha + \beta = 2 \text{ but } \gamma > 1/2.
\] (4.7)

Indeed (H1)-(H2) and (A1)-(A2) are valid in the range (4.7) while \( P \) may not be absolutely integrable, e.g., if \( \alpha + \beta = 2 \) and \( 1/2 < \gamma \leq 1 \).

We also have that (1.1) has two solutions with property (1.8) if

\[
\alpha + \beta > 5/2, \text{ or } \alpha + \beta = 5/2 \text{ but } \gamma > 1/2
\] (4.8)

and

\[
\alpha + 2\beta > 3, \text{ or } \alpha + 2\beta = 3 \text{ but } \gamma > 0.
\] (4.9)

Actually, (A3) holds in (4.8) and (A4)-(A5) hold in (4.9), while (4.5) holds in the range \( \alpha + \beta > 3 \) and \( \alpha + \beta = 3 \) and \( \gamma > 1 \), which is smaller than either (4.8) or (4.9).

We finally note that the range in which (A4) holds is larger than (4.9). We can therefore find the exact area in which (A3)-(A4) hold but (A5) fails.

**Example 2.** \( n = 1, p(t) = t^{-3} \log^{-1} t \left( 2 + 3 \log^{-1} t + 2 \log^{-2} t \right) \). In this case, (A3) and (A5) hold but (A4) does not. So, Theorem 3 fails according to Hille's result mentioned in Section 1.

It is conjectured that in Theorem 3, under the assumptions (H1) and (A2), the conditions (A3)-(A5) are also necessary.
5. Proof of Theorem 1

Our approach to proving the sufficiency part is founded mainly on a construction of a solution that tends to the unit matrix. We first ascertain the non-oscillation of the system (1.1). To this end, we shall need the following lemma.

**Lemma 4.** If $P^2 \in \mathbb{L}$, then the following statements are equivalent.

(i) (A1);

(ii) $H \in \mathbb{L}, H(t) = \int_t^\infty P^2(s)ds$;

(iii) $\bar{P} \in \mathbb{L}$.

In fact, for any $t \geq b \geq a$ with $b$ suitably large, we have

$$\int_b^t H(s)ds = tH(t) - bH(b) + \int_b^t sP^2(s)ds + \int_b^t \int_t^\infty P^2(s)ds.$$

(5.1)

Since $H$ and $P^2$ are non-negative Hermitian matrices, we have

$$\int_b^t sP^2(s)ds \leq bH(b) + \int_b^t H(s)ds \leq \int_b^t sP^2(s)ds.$$

(5.2)

Note that Hermitian matrix inequalities are taken in the positive, non-negative definite sense and that the second inequality in (5.2) makes sense when (A1) holds. Then the equivalence between (i) and (ii) follows immediately.

Let $Y(t)$ be the solution of $Z' = PZ$, satisfying $Y(t) \to I$ as $t \to \infty$. Then $X(s;t) = Y(s)Y^{-1}(t)$ and there is an $M > 0$ such that

$$|X(s;t)| \leq M, \quad |Y(t)| \leq M, \quad |Y^{-1}(t)| \leq M \quad \text{for} \quad t, s \geq a.$$

(5.3)

We observe that if $f(t)$ is Hermitian and $f(t) \geq 0$, then $f \in \mathbb{L} \iff Y^*fY \in \mathbb{L}$ since

$$|f(t)| = |(Y^{-1}Y^*fYY^{-1})(t)| \leq M^2 |(Y^*fY)(t)| \leq M^4 |f(t)|, \quad t \geq a.$$

(5.4)

On the other hand, since $P^2(t) \geq 0, \bar{P}(t) \geq 0$ and

$$\int_t^\infty Y^*(s)\bar{P}(s)Y(s)ds = \int_t^\infty Y^*(r)P^2(r)Y(r)drds$$

$$= \int_t^\infty (s-t)(Y^*P^2Y)(s)ds,$$

we have

$$\bar{P} \in \mathbb{L} \iff Y^*\bar{P}Y \in \mathbb{L} \iff tY^*P^2Y \in \mathbb{L} \iff tP^2 \in \mathbb{L} \iff (A1).$$

(5.6)
This proves Lemma 4.

From (A1) and Lemma 4, we have \( \overline{P} \in \mathcal{L} \), and hence
\[
\overline{P}^2 \in \mathcal{L}.
\] (5.7)

In view of (5.7), applying a contraction theorem to the Riccati integral equation (3.7), we can find a Hermitian solution \( V(t) \) of (3.7) on \([b, \infty)\) for some \( b \geq a \). Therefore, (1.1) is nonoscillatory. Moreover, in later discussions concerning the asymptotic behavior of solutions of (1.1), a crucial role will be played by the following estimate for \( V(t) \):
\[
\overline{P}(t) \leq V(t) \leq 2\overline{P}(t), \quad t \geq b.
\] (5.8)

To prove (5.8), let \( V_0 = 2\overline{P} \) and
\[
V_k(t) = \overline{P}(t) + \int_t^\infty X^*(s;t)V_{k-1}^2(s)X(s;t)ds, \quad k = 1, 2, \ldots.
\] (5.9)

An inductive procedure shows that the Hermitian matrix function sequence \( V_k(t) \) produced by (5.9) on \([b, \infty)\) is well defined, non-increasing, and bounded between \( \overline{P}(t) \) and \( 2\overline{P}(t) \). Thus, the limit \( V(t) \) of \( V_k(t) \) as \( k \to \infty \), which is clearly the unique solution of (3.7), exists and is itself bounded between \( \overline{P}(t) \) and \( 2\overline{P}(t) \).

We are now able to find the pair of desired solutions. Let \( W = P + V \), and note that \( V \) is absolutely integrable on \([b, \infty)\) because of (5.8). By (H2) and the absolute integrability of \( V \), the equation \( U' = WU \) has a solution \( U_1 \) satisfying
\[
\lim_{t \to \infty} U_1(t) = I.
\] (5.10)

A simple calculation shows that \( U_1 \) is a prepared solution of (1.1).

Finally, without loss of generality, we may assume \( U_1(t) \) is non-singular on \([b, \infty)\). Then one choice of the other linearly independent prepared solution of (1.1) is, e.g.,
\[
U_2(t) = U_1(t) \left( \int_b^t (U_1^*U_1)^{-1}(s)ds \right).
\] (5.11)

It may be established by an application of L'Hopital's rule that \( U_2(t)/t \to I \) as \( t \to \infty \). The proof of sufficiency is complete.

To prove the necessity, we observe first that \( U_1 \) is a non-oscillatory prepared solution and hence that (1.1) is non-oscillatory. By Lemmas 1 and 2, there are two Hermitian functions \( W(t) \) satisfying (3.3) and \( V(t) \) satisfying (3.7) on \([b, \infty)\) for some \( b \geq a \), and \( W = U_1^*U_1^{-1} \), \( V = W - P = \int_t^\infty W^2(s)ds \). Set \( U_1(t) = I - F(t) \). Choose \( b \) so large that \( |F(t)| < 1/2 \) for \( t \geq b \). Then,
\[ W(t) = -F'(t) - F'(t)F(t) - F'(t)F^2(t) - F'(t)F^3(t) - \cdots . \]  

(5.12)

Since \((F^k)' = F'F^{k-1} + FF'F^{k-2} + \cdots + F^{k-1}F'\), a repeated application of (2.1)-(2.2) to (5.12) produces

\[ \text{tr}(W(t)) = -\sum_{k=1}^{\infty} \frac{1}{k} (\text{tr}F^k(t))'; \]

(5.13)

hence,

\[ \int_t^\infty \text{tr}(W(s)) ds = \sum_{k=1}^{\infty} \frac{1}{k} \text{tr}(F^k(t)). \]

(5.14)

Since (H2) and \(V \geq 0\), (5.14) implies \(V \in L\). Furthermore, (5.8) implies \(P \in L\). Finally, (A1) follows from Lemma 4. Theorem 1 is proved.

6. Proof of Theorem 2

Without loss of generality, we always assume that \(U_2(t)\) is defined by (5.11). By means of Theorem 1, it suffices to prove that under the assumption (A1), \(U_2(t) \to I\) as \(t \to \infty\) if and only if (A2) holds. Moreover, since (A1) ensures \(U_1(t) \to I\) and \(U_2(t) = t(I + o(I))\) (\(o(I)\) is a matrix function tending to zero), and since

\[ U_2'(t) = U_1'(t)(U_1^*U_1)^{-1}(s)ds + U_1(t)(U_1^*U_1)^{-1}(t) \]

\[ = W(t)U_2(t) + U_1^{*^{-1}}(t) \]

\[ = tW(t)(I + o(I)) + U_1^{*^{-1}}(t), \]

(6.1)

where \(W = U_1^*U_1^{-1} = P + V\), it suffices to prove that \(tW(t) \to 0\) as \(t \to \infty\) if and only if (A2) holds. In the last analysis, since \(tW(t) = tP(t) + tV(t)\), it remains to prove that \(tV(t) \to 0\) under (A1). But this is true according to the following estimates:

\[ |tV(t)| \leq 2t\int_t^{\infty} |X^*(s;t)|P^2(s)X(s;t)| ds \]

\[ \leq 2M^2t\int_t^{\infty} |P^2(s)| ds \leq 2M^2\int_t^{\infty} |P^2(s)| ds \to 0 \quad \text{as} \quad t \to \infty, \]

(6.2)

obtained by an application of (5.8), (5.3), and (A1).
7. Proof of Theorem 3

For the proof of Theorem 3 we need two more lemmas.

**LEMMA 5.** If \( P^2 \in L \), then the following statements are equivalent:

(i) \((A3)\);

(ii) \( tH \in L, H(t) = \int tP^2(s)ds \);

(iii) \( tP \in L \).

The proof of Lemma 5 follows the steps of Lemma 4 except that we make use of the identity

\[
\int_b^t sH(s)ds = -\frac{1}{2} b^2 H(b) + \frac{1}{2} \int_b^t s^2 P^2(s)ds + \frac{1}{2} \int_t^\infty t^2 P^2(s)ds
\]  

(7.1)

instead of (5.1) to derive

\[
\frac{1}{2} \int_b^t s^2 P^2(s)ds \leq \frac{1}{2} b^2 H(b) + \int_b^t sH(s)ds \leq \frac{1}{2} \int_b^\infty s^2 P^2(s)ds
\]  

(7.2)

instead of (5.2), where \( b \) is a suitably large number.

It should be mentioned that if \((A1)\) holds, then the relation (iii) in Lemma 5 is, by virtue of (5.8), equivalent to

\[
\int_t^\infty ||V(r)||dr \int ds < \infty,
\]  

(7.3)

i.e.,

\[
\int_t^\infty ||V(s)||ds < \infty,
\]  

(7.4)

where \( V \) is the Hermitian solution of (3.7).

**LEMMA 6.** If \((A3)\) holds, then

\[
\int_a^\infty ||P(t)||dt < \infty \text{ for } \alpha \in [0,1/2].
\]  

(7.5)

Lemma 6 can be demonstrated by a use of Schwarz's inequality.

Having established these two lemmas, we now turn to the proof of the theorem. Since \((A3)\) implies \((A1)\), Theorem 1 provides a prepared solution \( U_1 \) of (1.1) on \([b,\infty)\) with \( U_1(t) \to I \) as
\( t \to \infty \) and

\[ U_1 = WU_1, \]  

(7.6)

where \( W = P + V, \) \( V \) is the solution of (3.7) with (5.8). Here \( W \) is absolutely integrable. Moreover, a combination of (7.4) and (7.5) leads to

\[
\int_{t}^{\infty} |W(s)| \, ds \leq \int_{t}^{\infty} |P(s)| \, ds + \int_{t}^{\infty} |V(s)| \, ds
\]

\[
\leq t^{-\alpha} \int_{t}^{\infty} |P(s)| \, ds + t^{-1} \int_{t}^{\infty} |V(s)| \, ds
\]

\[ = o(t^{-\alpha}) \text{ as } t \to \infty \text{ for } \alpha \in (0,1/2). \]  

(7.7)

We again define \( U_2 \) by (5.11). The requirements for \( U_2 \) in Theorem 3 turn out to arise from a detailed investigation of the solution \( U_1 \). To do this, we integrate (7.6) from \( t \) to \( \infty \) and get

\[ U_1(t) = I - \int_{t}^{\infty} (WU_1)(s) \, ds, \quad t \geq b. \]  

(7.8)

Two iterations of (7.8) yield

\[ U_1(t) = I - \int_{t}^{\infty} W(s) \, ds + \int_{t}^{\infty} W(s) \, \int_{s}^{\infty} W(r) \, dr - \int_{t}^{\infty} W(s) \, \int_{s}^{\infty} W(r) \, W(z) \, U_1(z) \, dz \, dr \, ds
\]

\[ = I - F(t), \]  

(7.9)

where

\[ F(t) = Q(t) + \int_{t}^{\infty} (PQ)(s) \, ds + G(t), \]  

(7.10)

\[ G(t) = \int_{t}^{\infty} V(s) \, ds - \int_{t}^{\infty} P(s) \, \int_{s}^{\infty} V(r) \, dr \, ds - \int_{t}^{\infty} V(s) \, \int_{s}^{\infty} W(r) \, dr \, ds + G_1(t), \]  

(7.11)

\[ G_1(t) = \int_{t}^{\infty} W(s) \, \int_{s}^{\infty} W(r) \, W(z) \, U_1(z) \, dz \, dr \, ds. \]  

(7.12)

We then prove that

\[ \int_{b}^{\infty} |G(t)| \, dt < \infty \text{ and } tG(t) \to 0 \text{ as } t \to \infty. \]  

(7.13)

In view of (7.3), (7.4), and
\begin{align}
|G(t)| & \leq \left[ 1 + \int_t^\infty |P(s)| ds + \int_t^\infty |W(s)| ds \right] \int_t^\infty |V(s)| ds + |G_1(t)| , \tag{7.14}
\end{align}

all that is needed in (7.13) is that $|G_1| \in L$ and $t |G_1(t)| \to 0$. But these requirements will be met by a selection of the value of $\alpha$, say, $\alpha = 2/5$, in the following estimate:

\begin{align}
|G_1(t)| & \leq M \int_t^\infty |W(s)| \int_s^\infty |W(r)| \int_r^\infty |W(z)| d\zeta dr ds \\
& = \frac{M}{3!} \left[ \int_t^\infty |W(s)| ds \right]^3 = o(t^{-3\alpha}) \text{ for } \alpha \in (0,1/2) . \tag{7.15}
\end{align}

On the other hand, because of Lemma 6 and the boundedness of $tQ(t)$, we have

\begin{align}
\int_t^\infty |PQ_1(s)| ds \leq t^{-1-\alpha} \int_t^\infty |s^\alpha |P(s)| ds |sQ(s)| ds = o(t^{-1-\alpha})
\end{align}

as $t \to \infty$ for $0 < \alpha < 1/2$, \tag{7.16}

and hence (A4)-(A5) imply that

$F \in L$ and that $\lim_{t \to \infty} tF(t) = C$ exists finitely. \tag{7.17}

Also (7.17) implies the absolute integrability of $F^2$.

Finally, we write

$$
(U_1^* U_1)^{-1} = I + F + F^* + R ,
$$

where

$$
R = U_1^{-1} (F^2 + FF^* + F^* 2) U_1^* - U_1^{-1} (F^2 F^* + FF^* 2) U_1^* \tag{7.17}
$$

is absolutely integrable because of (7.17). If we rewrite $U_2$ as

$$
U_2(t) = U_1(t) \int_b^t (U_1^* U_1)^{-1}(s) ds
$$

and hence (7.17), we can get the desired asymptotic property of $U_2$, \tag{7.17}

$$
\lim_{t \to \infty} (U_2(t) - t) = \int_b^\infty (F + F^* + R) ds - C - bI ,
$$

which completes the proof.
8. Nonhomogeneous Systems

We are now in a position to apply our main results to the non-homogeneous system

\[ U'' + p(t)U = q(t), \quad t \geq a, \quad (8.1) \]

where \( p \) is defined as before and

(H3) \( q(t) \) is an \( n \times n \) matrix-valued function on \( [a, \infty) \) and

\[
\int_a^\infty |q(t)| \, dt < \infty.
\]

A means to this end is the constant-variation formula.

**Theorem 4.** Let (H1)-(H3) hold. Then (A1) is equivalent to the following: for any given constant matrix \( A \) there is a unique solution \( U(t, A) \) of (8.1) such that

\[
\lim_{t \to \infty} U(t, A) = A, \quad \lim_{t \to \infty} U'(t, A) = 0, \quad (8.2)
\]

and for any solution \( U(t) \) of (8.1)

\[
\lim_{t \to \infty} t^{-1}U(t) \quad \text{exists finitely.} \quad (8.3)
\]

**Proof.** If (A1) holds, then Theorem 1 gives two prepared solutions \( U_1 \) and \( U_2 \) of (1.1) on \( [b, \infty) \) satisfying (1.6). Under the assumption (H3), the function

\[
U_s(t) = U_1(t) \int_U (U_1^{-1}U_2U_1^*q)(s) \, ds - U_2(t) \int_U (U_1^*q)(s) \, ds \quad (8.4)
\]

is not only well defined on \( [b, \infty) \) but a particular solution of (8.1) as well. It is not difficult to verify by a calculation that \( U_s(t) \to 0 \) and \( U'(s) \to 0 \) as \( t \to \infty \). We also note that

\[
U'(t) = W(t)U_1(t) \to 0 \quad \text{as} \quad t \to \infty.
\]

Then \( U_1A + U_s \) is the unique solution satisfying (8.2). Equation (8.3) is obviously true since the general solution of (8.1) can be written as

\[ U = U_1A + U_2B + U_s. \]

Conversely, pick up two solutions \( U(t,I) \) and \( U(t,2I) \) of (8.1), and let

\[ U_1(t) = U(t,2I) - U(t,I). \]

Consequently, \( U_1 \) is a solution of (1.1), and \( U_1(t) \to I, U'(t) \to 0 \) as \( t \to \infty \) with the aid of (8.2). Since \( U_1^* - U_1 - U_1^*U_1^* \) is a constant matrix which must be zero, \( U_1 \) is a prepared solution. Then we get (A1) from Theorem 1.

**Theorem 5.** The assumptions (A1)-(A2) are necessary and sufficient in order that for any given \( n \times n \) matrix \( A \), Equation (8.1) has a unique solution \( U(t, A) \) satisfying (8.2), and for any solution \( U \) of (8.1)
\lim_{t \to \infty} U'(t) \text{ exists finitely.} \quad \text{(8.5)}

Theorem 5 follows immediately from Theorems 2 and 4.

**Theorem 6.** If (H1), (H3), and (A2)-(A5) are valid, then for any given \( n \times n \) matrices \( A \) and \( B \), there exists a unique solution \( U \) of (8.1) satisfying

\[
U'(t) \to B \quad \text{and} \quad U(t) - tB \to A \quad \text{as} \quad t \to \infty.
\]

**Proof.** From Theorem 3, (1.1) has two solutions \( U_1 \) and \( U_2 \) so that \( U_1(t) \to I \) and \( (U_2(t) - tI) \to C \) as \( t \to \infty \). Then the desired solution is

\[
U(t) = U_1(t)(A - CB) + U_2(t)B + U_3(t).
\]

**9. Terminal Value Problems**

For any given constant matrices \( A \) and \( B \), a solution \( U(t) \) of (8.1) satisfying either

\[
U'(t) \to 0, \quad U(t) \to A, \quad \text{or} \quad U'(t) \to B, \quad U(t) - tB \to A \quad \text{as} \quad t \to \infty,
\]

will be called a solution of the terminal value problem (TVP) (8.1), (9.1) or (8.1), (9.2), respectively. Now Theorems 4 and 6 in Section 8 can be interpreted in terms of terminal value problems.

**Theorem 4'.** Let (H1)-(H3) hold. For any constant \( n \times n \) matrix \( A \) there exists a unique solution of the TVP (8.1), (9.1) if and only if (A1) holds.

**Theorem 6'.** Let (H1), (H3), and (A2)-(A5) hold. Then for any given \( n \times n \) matrices \( A \) and \( B \) there exists a unique solution of the TVP (8.1), (9.2).

**10. Nonlinear Perturbation**

In the equation

\[
U'' + p(t)U = f(t,U), \quad t \geq a,
\]

we define \( p \) as before and let \( f \) be a locally integrable, \( n \times n \) matrix function on \([a, \infty) \times \mathbb{C}^n \), satisfying the following condition:
There are two non-negative scalar functions $q_1(t)$ and $q_2(t)$ satisfying
\[ \int_a^\infty t^i q_i(t) dt < \infty, \quad i = 1, 2, \]
such that $|f(t, 0)| \leq q_1(t)$ and for any $n \times n$ matrices $U$ and $V$
\[ |f(t, U) - f(t, V)| \leq q_2(t) |U - V|. \]

We shall state two theorems in terms of terminal value problems and simply sketch the proofs.

**THEOREM 7.** If $(H1)$-$H2)$, $(H4)$, and $(A1)$ are valid, then for any given matrix $A$, the TVP (10.1), (9.1) has a unique solution.

**Proof.** Under the given assumptions, Theorem 1 provides two prepared solutions $U_1$ and $U_2$ of (1.1) satisfying (1.6). Since any solution $U$ of (10.1) can be represented by
\[ U(t) = U_1(t)A + U_2(t)B - U_1(t)\int_b^t (U_1^{-1} U_2 U_1^* f(\cdot, U))(s) ds + U_2(t)\int_b^t (U_1^* f(\cdot, U))(s) ds, \]
where the constant matrices $A$ and $B$ may depend on $U$ and $b$ is taken so large that $U_1$ is non-singular on $[b, \infty)$, we can then prove that $|U(t)|$ is bounded by a linear scalar function, say, $Kt$, for some $K > 0$. With the aid of $(H4)$, a solution of TVP (10.1), (9.1) is therefore equivalently a continuous solution of the integral equation
\[ U(t) = U_1(t)A + U_1(t)\int_b^t (U_1^{-1} U_2 U_1^* f(\cdot, U))(s) ds \]
\[ - U_2(t)\int_b^t (U_1^* f(\cdot, U))(s) ds, \quad t \geq b. \quad (10.2) \]

Again by means of $(H4)$, we may employ an argument based on the contraction mapping theorem to establish the existence and uniqueness of a solution of (10.3). We leave the details to the reader.

**THEOREM 8.** If $(H1)$, $(H4)$, and $(A2)$-$A5)$ are valid, then for any given matrices $A$ and $B$, the TVP (10.1), (9.2) has a unique solution.

Apart from some manipulations, the proof is similar to that of Theorem 7 except that we consider the integral equation
\[ U(t) = U_1(t)(A - CB) + U_2(t)B + U_1(t) \left( U_1^{-1}U_2(U^*f(\cdot, U))(s) \right) ds - U_2(t) \left( U_1^*f(\cdot, U) \right)(s) ds, \quad t \geq b, \tag{10.4} \]

in place of (10.3). We again omit the details.

It should be remarked at this point that the cited references of Kusano and Trench investigate terminal value problems (in our terminology) for various linear differential equations with nonlinear perturbations. Some of their integral smallness conditions on the perturbations permit conditional convergence in contrast with our absolute convergence conditions in (H3) and (H4). However, it seems reasonable that a "better" perturbation should be compensated for by a "worse" coefficient in the linear part of the equation, while a "better" coefficient may admit a "worse" perturbation. Therefore, it appears possible to relax the conditions on \( q \) and \( f \) only in Theorems 6 and 8.

**Acknowledgment**

I thank Professor F. V. Atkinson for encouraging me to write this paper during my visit to the University of Toronto.

**References**


OSCILLATION THEORY FOR SYSTEMS OF SECOND-ORDER DIFFERENTIAL EQUATIONS*

L. H. Erbe and Lakshmareddy Ganta†
Department of Mathematics
University of Alberta
Edmonton, Canada T6G 2G1

Abstract

We obtain some oscillation criteria for the second order differential system

\[ Y'' + Q(t)Y = 0, \quad t \in [a, +\infty), \]

where \( Q, Y \) are \( n \times n \) real matrix-valued functions defined on \( [a, +\infty) \) with \( Q = Q(t) \) symmetric.

1. Introduction

Consider the second order differential system

\[ Y'' + Q(t)Y = 0, \quad t \in [a, +\infty), \quad (1.1) \]

where \( Y(t), Q(t) \) are \( n \times n \) real continuous matrix functions and \( Q(t) \) is symmetric for all \( t \). We recall that a solution \( Y(t) \) of (1.1) is said to be nontrivial if \( \det Y(t) \neq 0 \) for at least one \( t \in [a, +\infty) \) and a nontrivial solution is said to be prepared in case

\[ Y^*(t)Y(t) - Y^*(t)Y(t) = 0, \quad t \in [a, +\infty), \quad (1.2) \]

where \( * \) denotes transpose. Equation (1.1) is said to be oscillatory on \( [a, +\infty) \) in case the determinant of every nontrivial prepared solution vanishes on \( [b, +\infty) \) for each \( b > a \). Many results [Allegretto and Erbe 1973, Etgen and Lewis 1979, Hartman 1979, Hinton and Lewis 1980, Tomastik 1971, and Walters 1980] have been obtained showing that (1.1) is oscillatory if a corresponding scalar equation that results when one applies a positive linear functional to \( Q(t) \) is oscillatory. More recently, criteria have been obtained involving eigenvalues of \( Q(t) \) (or of its integral) that yield oscillation conditions for (1.1). In particular, if for a symmetric matrix \( A \), we use the notation \( \lambda_1(A) \geq \cdots \geq \lambda_n(A) \) to denote the usual ordering of the eigenvalues, then

\[ \lim_{t \to \infty} \lambda_1 \left( \int_a^t Q(s)ds \right) = +\infty \quad (1.3) \]

implies that (1.1) is oscillatory. This result was obtained by Kaper and Kwong [1986] and (with a different proof) by Byers, Harris, and Kwong [1986]. In [Butler, Erbe, and Mingarelli 1987] and [Erbe 1986] additional criteria for oscillation of (1.1) were given which extend well-known scalar criteria. Continuing the results obtained by Butler, Erbe, and Mingarelli [1987] and Erbe [1986],

*Research supported by NSERC, Canada.
† On study leave from Osmania University, Hyderabad-500007, India.
we obtain several additional criteria for oscillation of (1.1) which treat cases that were not covered in previous work.

For completeness, we recall some notation and definitions. For a symmetric matrix \( A \), \( \sigma_k(A) \), \( \rho_k(A) \) will denote the sum of the \( k \) largest (smallest) eigenvalues:

\[ \sigma_k(A) = \sum_{i=1}^{k} \lambda_i(A), \quad \rho_k(A) = \sum_{i=1}^{k} \lambda_{n-k+1}(A). \]

For a real-valued function \( f : [\alpha, +\infty) \to \mathbb{R} \) and (extended) real numbers \( l, m \) with \( -\infty \leq l, m \leq +\infty \), we say \( \lim \text{ appr inf}_{t \to \infty} f(t) = l \) in case \( \mu(t : f(t) \leq l_1) < +\infty \) for all \( l_1 < l \) and \( \mu(t : f(t) \leq l_2) = +\infty \) for all \( l_2 > l \) (where \( \mu(\cdot) \) denotes Lebesgue measure). Similarly, \( \lim \text{ appr sup}_{t \to \infty} f(t) = m \) in case \( \mu(t : f(t) \geq m_1) = +\infty \) for all \( m_1 < m \) and \( \mu(t : f(t) \geq m_2) < +\infty \) for all \( m_2 > m \). Finally, \( \lim \text{ appr } f(t) = \gamma \) in case \( \lim \text{ appr inf}_{t \to \infty} f(t) = \lim \text{ appr sup}_{t \to \infty} f(t) = \gamma \).

2. Main Results

We recall that the matrix equation (1.1) is nonoscillatory on \([a, +\infty)\) iff the Riccati integral equation

\[ Z(t) = Z(t_0) + \int_{t_0}^{t} \mu(s) ds + \int_{t_0}^{t} Z^2(s) ds \] (2.1)

has a continuous symmetric solution on \([t_0, \infty)\) for some \( t_0 \geq a \). In particular, if \( Y(t) \) is a non-trivial prepared solution of (1.1) with \( \det Y(t) \neq 0 \) for \( t \geq t_0 \), then \( Z = -Y^{-1} \) is a symmetric solution of (2.1) on \([t_0, \infty)\). As in [Butler, Erbe, and Mingarelli 1987] we define the (extended real-valued) function \( L \) defined on the class of symmetric real continuous matrices defined on \([a, \infty)\) by

\[ L(Q) = \lim \inf_{T \to \infty} \frac{1}{T} \int_{a}^{T} \mu(s) \text{ tr}(Q(s)) ds dr. \] (2.2)

It was shown by Butler, Erbe, and Mingarelli [1987] that if (1.1) is nonoscillatory on \([a, +\infty)\), then \( L(Q) > -\infty \) iff \( \lim_{T \to \infty} \int_{a}^{T} Z^2(s) ds \) exists for any solution \( Z = -Y^{-1} \) of (2.1) (where \( Y = Y(t) \) is a prepared solution of (1.1)). This is a generalization of a fundamental lemma of Hartman [1969] for the scalar equation \( y'' + q(t)y = 0 \). In what follows, \( \eta_k(t_0), \mu_k(t_0), \Delta_k(t_0) \) will denote the expressions

\[ \eta_k(t_0) = \lim \sup_{T \to \infty} \frac{1}{T} \int_{t_0}^{T} \sigma_k \left[ \int_{t_0}^{s} Q(s) ds \right] dr, \]
\[ \mu_k(t_0) = \lim \inf_{T \to \infty} \frac{1}{T} \int_{t_0}^{T} \sigma_k \left[ \int_{t_0}^{s} Q(s) ds \right] dr, \]
\[ \Delta_k(t_0) = \eta_k(t_0) - \mu_k(t_0). \] (2.3)

We may now state our main result.
THEOREM 2.1. Assume \( L(Q) > -\infty \) and that there exists \( \delta > 0 \) and \( k, 1 \leq k \leq n \) such that
\[
\Delta_k(t_0) \geq \delta \quad \text{for all large } t_0.
\] (2.4)

Then (1.1) is oscillatory.

THEOREM 2.2. Assume \( L(Q) > -\infty \) and that there exists \( \delta > 0 \) and \( k, 1 \leq k \leq n \) such that
\[
\lim_{t \to \infty} \sup_{s \in J} \sigma_k \left[ \int_{t_0}^{t} Q(s) ds \right] - \lim_{t \to \infty} \inf_{s \in J} \sigma_k \left[ \int_{t_0}^{t} Q(s) ds \right] \geq \delta
\] (2.5)
for all large \( t_0 \). Then (1.1) is oscillatory.

REMARK. Theorems 2.1 and 2.2 may be thought of as generalizations of well-known oscillation criteria for the scalar equation \( y'' + q(t)y = 0 \) (cf. Butler, Erbe, and Mingarelli [1987]). Theorem 2.1 for the case \( k = 1 \) is Theorem 2.4 of [Erbe 1986], whereas Theorem 2.2 further extends Theorem 2.1 C,D of [Erbe 1986] (which dealt with the case \( k = 1 \)).

THEOREM 2.3. Assume \( L(Q) = -\infty \). Then (1.1) is oscillatory if
\[
\lim_{t \to \infty} \sup_{s \in J} \rho_k \left[ \int_{t_0}^{t} Q(s) ds \right] > -\infty
\] (2.6)
for some \( 1 \leq k \leq n \).

We remark that examples may be easily given (with, e.g., \( Q(t) = \text{diag}(q_1(t) \cdots q_n(t)) \)) to illustrate that any one of the conditions (2.4)-(2.6) may hold for some \( k = k_0, 1 < k_0 \leq n \), but may fail for all \( 1 \leq k < k_0 \). Thus, Theorems 2.1-2.3 each give a sequence of \( n \) tests for oscillation.

Proof of Theorem 2.1. The proof is by contradiction. We assume that \( Z = -Y \gamma^{-1} \) is a solution of (2.1) on \([t_0, \infty), t_0 \geq a\). We suppose first that both terms appearing on the right side of (2.3) are finite. We define
\[
Q_1(t;t_0) \equiv \int_{t_0}^{t} Q(s) ds, \quad A(t;t_0) \equiv Z(t_0) + \int_{t_0}^{t} Z^2(s) ds.
\] (2.7)

By Lemma 5.1 of Butler, Erbe, and Mingarelli [1987], \( \lim A(t;t_0) \) exists and hence \( \lim_{t \to \infty} \text{tr} A(t;t_0) \) exists. Let \( \varepsilon > 0 \) with \( \varepsilon < \frac{\delta}{3} \). We may suppose that \( t_0 \) is sufficiently large so that
\[
\int_{s}^{t} \text{tr} Z^2(s) ds < \varepsilon, \quad t_0 \leq s < t < \infty \quad \text{and also} \sigma_k(Z^2(t_0)) < \frac{\varepsilon^2}{k}.
\]
For symmetric matrices \( A, B \) we note that \( \rho_k(A) + \sigma_k(B) \leq \sigma_k(A + B) \leq \sigma_k(A) + \sigma_k(B) \) and \( |\sigma_k(A)|^2 \leq k \sigma_k(A^2) \), as is easily established. 

By (2.7) we can rewrite (2.1) as
\[ Z(t) = A(t; t_0) + Q_1(t; t_0) \]  
(2.8)

and hence

\[ \sigma_k(Z(t)) \leq \sigma_k(A(t; t_0)) + \sigma_k(Q_1(t; t_0)) \]  
(2.9)

and

\[ \sigma_k(Z(t)) \geq \rho_k(A(t; t_0)) + \sigma_k(Q_1(t; t_0)) . \]  
(2.10)

Further, \[ \sigma_k(A(t; t_0)) \leq \sigma_k(Z(t_0)) + \sigma_k(\int_{t_0}^t Z^2(s)ds) \]. Now since \( |\sigma_k(Z(t_0))|^2 \leq k \sigma_k(Z^2(t_0)) < \epsilon^2 \)
and \( \sigma_k(\int_{t_0}^t Z^2(s)ds) \leq t \int_{t_0}^t Z^2(s)ds < \epsilon \), we have from (2.9)

\[ \sigma_k(Z(t)) < 2\epsilon + \sigma_k(Q_1(t; t_0)), \quad t \geq t_0 . \]  
(2.11)

Also, \( \rho_k(A(t; t_0)) \geq \rho_k(Z(t_0)) + \rho_k(\int_{t_0}^t Z^2(s)ds) > -\epsilon \), so that from (2.10) we have

\[ \sigma_k(Z(t)) > -\epsilon + \sigma_k(Q_1(t; t_0)), \quad t \geq t_0 . \]  
(2.12)

We let \( T_n \to \infty \) and \( \tau_n \to \infty \) so that

\[ \lim_{T_n \to \infty} \frac{1}{T_n} \int_{t_0}^{T_n} \sigma_k(Q_1(t; t_0))dt = \eta_k(t_0) \]  
(2.13)

and

\[ \lim_{\tau_n \to \infty} \frac{1}{\tau_n} \int_{t_0}^{\tau_n} \sigma_k(Q_1(t; t_0))dt = \mu_k(t_0) . \]  
(2.14)

From (2.11) and (2.12) we obtain

\[ \frac{1}{\tau_n} \int_{t_0}^{\tau_n} \sigma_k(Z(t))dt < \frac{2\epsilon(\tau_n-t_0)}{\tau_n} + \frac{1}{\tau_n} \int_{t_0}^{\tau_n} \sigma_k(Q_1(t; t_0))dt \]  
(2.15)

and

\[ \frac{1}{T_n} \int_{t_0}^{T_n} \sigma_k(Z(t))dt > \frac{-\epsilon(T_n-t_0)}{T_n} + \frac{1}{T_n} \int_{t_0}^{T_n} \sigma_k(Q_1(t; t_0))dt . \]  
(2.16)

Now by the Schwartz inequality

\[ \left| \frac{1}{T} \int_{t_0}^{T} \sigma_k(Z(t))dt \right|^2 \leq \frac{1}{T} \int_{t_0}^{T} (\sigma_k(Z(t))^2)dt \]

\[ \leq \frac{k}{T} \int_{t_0}^{T} \sigma_k(Z^2(t))dt \]

\[ \leq \frac{k}{T} \int_{t_0}^{T} \sigma_k(Z^2(t))dt \to 0 \quad \text{as} \quad T \to \infty . \]

Hence, taking limits in (2.15) and (2.16) as \( \tau_n \to \infty \), \( T_n \to \infty \), we get

\[ -\epsilon + \eta_k(t_0) \leq 0 \leq 2\epsilon + \mu_k(t_0) \]  
so that \( \Delta_k(t_0) = \eta_k(t_0) - \eta_k(t_0) \leq 3\epsilon \), contradicting (2.4). This
proves the theorem in the case that both $\eta_k(t_0)$ and $\mu_k(t_0)$ are finite. We suppose next that $\eta_k(t_0) = +\infty$. Since $\sigma_k(A) \leq k\lambda_1(A)$, it follows that $\limsup_{T \to \infty} \frac{1}{T} \int_0^T \lambda_1(t; t_0) dt = +\infty$ and so by Theorem 2.1A of Butler, Erbe, and Mingarelli [1987], the result follows. Finally, we suppose that $\eta_k(t_0) < +\infty$ and $\mu_k(t_0) = -\infty$. Then in (2.14) we have $\lim_{t_n \to \infty} \frac{1}{t_n} \int_{t_n}^{t_0} \sigma_k(Q_1(t; t_0)) dt = -\infty$, so that when taking limits in (2.15) and (2.16) we get $-\varepsilon + \eta_k(t_0) \leq 0 \leq 2\varepsilon + \mu_k(t_0) = -\infty$, a contradiction. (Note here that $\varepsilon$ and $\delta$ are arbitrary.) This completes the proof.

**Proof of Theorem 2.2.** As in the proof of Theorem 2.1, we suppose that $Z = -Y\gamma^{-1}$ is a solution of (2.1), and we obtain (2.8). We define $l(t_0), m(t_0)$ by

$$m(t_0) = \limsup_{t \to \infty} \sigma_k(Q_1(t; t_0))$$

$$l(t_0) = \liminf_{t \to \infty} \sigma_k(Q_1(t; t_0))$$

so that $m(t_0) - l(t_0) \geq \delta > 0$ for all large $t_0$. Given $\varepsilon < \frac{\delta}{3}$, we may suppose, as in the proof of Theorem 2.1, that $t_0$ is sufficiently large so that

$$\int_{t_0}^{t_0} \sigma_k(Z^2(s)) ds < \varepsilon, \quad |\sigma_k(Z(t_0))| < \varepsilon, \quad \text{and} \quad |p_k(Z(t_0))| < \varepsilon.
$$

Thus, from (2.8) we obtain

$$\sigma_k(Z(t)) = \sigma_k(A(t; t_0)) + \sigma_k(Q_1(t; t_0))$$

$$\leq \sigma_k(Z(t_0)) + \sigma_k \left[ \int_{t_0}^{t} Z^2(s) ds \right] + \sigma_k(Q_1(t; t_0)) \tag{2.18}$$

$$< \varepsilon + \int_{t_0}^{t} \sigma_k(Z^2(s)) ds + \sigma_k(Q_1(t; t_0))$$

$$< 2\varepsilon + \sigma_k(Q_1(t; t_0))$$

and

$$\sigma_k(Z(t)) \geq \rho_k(Z(t_0)) + \sigma_k \left[ \int_{t_0}^{t} Z^2(s) ds + Q_1(T; t_0) \right]$$

$$\geq \rho_k(Z(t_0)) + \rho_k \left[ \int_{t_0}^{t} Z^2(s) ds \right] + \sigma_k(Q_1(t; t_0))$$

$$> -\varepsilon + \sigma_k(Q_1(t; t_0)) .$$

Since $\int_{t_0}^{t} \sigma_k(Z^2(s)) ds < \infty$, it follows that $\sigma_k(Z(t))$ cannot be $\leq -\eta < 0$ on a set of infinite measure and cannot be $\geq \eta > 0$ on a set of infinite measure (since for a symmetric matrix $A$, we have
\begin{align*}
\rho_k(A)l^2 &= |\sigma_k(-A)|^2 \leq k\sigma_k(A^2^2). \text{ Therefore,} \\
\lim \text{ appr } \sigma_k(Z(t)) &= 0 = \lim \text{ appr inf } \sigma_k(Z(t)) \\
&= \lim \text{ appr sup } \sigma_k(Z(t)).
\end{align*}

Therefore, from (2.18) we obtain

\begin{align*}
0 &= \lim \text{ appr inf } (\sigma_k(A(t;t_0)) + \sigma_k(Q_1(t; t_0)) \\
&\leq 2\varepsilon + \lim \text{ appr inf } \sigma_k(Q_1(t; t_0)) \\
&= 2\varepsilon + l(t_0),
\end{align*}

(2.20)

and from (2.19) we have

\begin{align*}
0 \geq &-\varepsilon + \lim \text{ appr sup } \sigma_k(Q_1(t; t_0)) \\
&= -\varepsilon + m(t_0),
\end{align*}

so (2.20), (2.21) yield the contradiction \( m(t_0) - l(t_0) < \delta \). This completes the proof of Theorem 2.2.

**Proof of Theorem 2.3.** Suppose again that \( Z = -Y'Y^{-1} \) is a solution of (2.1). Since \( L(Q) = -\infty \), it follows that \( \int_{t_0}^t \text{ tr } Z^2(s)ds \to +\infty \) and hence \( \sigma_k(\int_{t_0}^t Z^2(s)ds) \to +\infty \) as \( t \to +\infty \) [Butler, Erbe, and Mingarelli 1987, Lemma 5.1]. For \( A = A^* \geq 0 \), we note that \( \text{ tr } (A) \leq (n - k + 1)\sigma_k(A) \leq n\sigma_k(A) \), and so from (2.1) we have

\begin{align*}
\frac{1}{n} \text{ tr } \left[ \int_{t_0}^t Z^2(s)ds \right] &= \frac{1}{n} \text{ tr } (Z(t) - Q_1(t; t_0) - Z(t_0)) \\
&\leq \sigma_k(Z(t) - Q_1(t; t_0) - Z(t_0)) \\
&\leq \sigma_k(Z(t)) - \rho_k(Q_1(t; t_0)) - \rho_k(Z(t_0)).
\end{align*}

Hence, we obtain (since \( \text{ tr } Z^2(t) \geq \sigma_k(Z^2(t)) \))

\begin{align*}
\rho_k(Z(t_0)) + \frac{1}{n} \int_{t_0}^t \sigma_k(Z^2(s))ds &\leq \sigma_k(Z(t)) - \rho_k(Q_1(t; t_0)), \quad t \geq t_0.
\end{align*}

(2.22)

If \( M = \lim \text{ appr sup } \rho_k(Q_1(t; t_0)) (= M(T_0)) \), then for any \( \varepsilon > 0 \), it follows that
\[ \mu(t; \rho_k(\mathcal{Q}_1(t; t_0)) \geq M - \varepsilon) = +\infty \]

and hence by (2.22)

\[ \mu \left( t : \rho_k(Z(t_0)) + \frac{1}{n} \int_{t_0}^t \sigma_k(Z^2(s)) ds \leq \sigma_k(Z(t)) - M + \varepsilon \right) = +\infty. \]

If the set \( E \) is defined by

\[ E = \left\{ t \in [t_0 + 1, \infty) : \frac{1}{2n} \int_{t_0}^t \sigma_k(Z^2(s)) ds \leq \sigma_k(Z(t)) \right\}, \]

then \( \mu(E) = +\infty \). Next, defining \( P(t) \) by

\[ P(t) = \int_{t_0}^t \sigma_k(Z^2(s)) ds, \]

then

\[ P'(t) = \sigma_k(Z^2(t)) \geq \sigma_k(Z^2(t)) \geq \frac{1}{k} (\sigma_k(Z(t))^2, \]

and for \( t \in E \) we have

\[ \frac{1}{P(t_0 + 1)} \geq \int_E \frac{P'(t)}{P^2(t)} \ dt \geq \frac{1}{4n^2k \mu(E) = +\infty}, \]

a contradiction. This completes the proof.

References


A USER'S GUIDE TO SUBROUTINE SPDNSF*

Charles T. Fulton
Department of Applied Mathematics
Florida Institute of Technology
Melbourne, FL 32901

Steven A. Pruess
Mathematics Department
Colorado School of Mines
Golden, CO 80401

Abstract

Subroutine SPDNSF (for Spectral Density Function) computes an approximation to the singular spectral function for Sturm-Liouville problems having continuous spectrum over $[0, \infty)$. The approximation is calculated by computing a large number of eigenvalues and norms of eigenfunctions (thousands) over $[a, b]$ for large $b$ and computing the associated step spectral function at a user-selected set of $\lambda$-points. The convergence of the step spectral function to the singular spectral function is known to be uniform on compact subsets of the continuous spectrum, a result that is supported by the numerical evidence. The code is restricted to the equation $-u'' + qu = \lambda u$ and written for a type of Sturm-Liouville problem involving linear dependence on $\lambda$ in the boundary condition at the left endpoint. The code makes use of asymptotic formulas for the large eigenvalues and a shooting algorithm of S. Pruess for the smaller eigenvalues.

1. Introduction

The code SPDNSF (for Spectral Density Function) is designed to compute the spectral function $\rho(t)$ associated with Sturm-Liouville problems that have continuous or partly continuous spectrum. In general, closed-form formulas for spectral functions of singular Sturm-Liouville problems do not exist, and for the handful of special cases for which such formulas are available, they have been obtained with the inversion formula for $\rho(t)$ in terms of the Titchmarsh-Weyl $m$-coefficient, a process that usually involves complicated manipulations of special functions for complex values of the eigenparameter as its imaginary part tends to zero. Hand calculations of this type do not lend themselves to numerical calculations because of the complexity of the limiting processes involved.

---

*This work was supported in part by the Applied Mathematical Sciences subprogram of the Office of Energy Research, U.S. Department of Energy, under Contract W-31-109-Eng-38, by NSF Grant #ECS-8515067 by the Office of Advanced Scientific Computing for supercomputing time at the National Center for Supercomputing Applications in Urbana, Illinois, by NSF Grant #DMS-8813113 (Florida Institute of Technology), and by NSF Grant #DMS-8800839 (Colorado School of Mines) by the Computational Mathematics Program.
An alternative possibility which does lead to efficient numerical algorithms is applicable when the left endpoint is regular and the right endpoint is of limit-point type. In this case, the step-spectral function associated with the interval \((a, b)\) is known to converge to the singular spectral function as \(b\) tends to the limit-point endpoint. The present code is based on computing a large number of eigenvalues and norms of the associated eigenfunctions so that the step-spectral function can be computed to a specified accuracy over a reasonably large interval \((a, b)\). Its efficiency is based on using asymptotic formulas both for the larger eigenvalues and for the norms of the eigenfunctions of the larger eigenvalues; a shooting algorithm is used to compute the smaller eigenvalues and their eigenfunctions. To obtain starting values for the shooting algorithm, the oscillation properties of the eigenfunctions are exploited in order to isolate the eigenvalues.

At present, the code is designed for problems that are regular at the left endpoint and have a boundary condition there involving a linear dependence on the eigenvalue parameter. This was chosen as the first test problem because of the simple asymptotic behavior of \(\rho(t)\) as \(t \to \infty\), namely, \(\rho(t)\) has bounded variation over \((-\infty, \infty)\) with the limit as \(t \to \infty\) being known (for both the finite and infinite interval cases) in terms of the boundary condition constants. Similar codes for the case of a standard regular boundary condition at the left endpoint, and for certain cases of limit-circle boundary conditions that allow for singular potentials, are currently under development.

2. The Basic Problem

We consider the Sturm-Liouville problem

\[-u'' + q(x)u = \lambda u \quad a \leq x \leq b\]  
\[a_1 u(a) - a_2 u'(a) = \lambda (a_1 u(a) - a_2 u'(a))\]  
\[b_1 u(b) + b_2 u'(b) = 0;\]  

we assume that

\[\alpha := a_1 a_2 - a_2 a_1 > 0,\]  

either \(b_1 = 0\) or \(b_2 = 0\) but not both,

that \(q(x)\) is continuous on \([a, \infty)\), and that \(q(x) \to 0\) as \(x \to \infty\). The basic eigenfunction expansion theory for this problem has been given in [Fulton 1977, 1980]; here we summarize a few of the basic results from these two papers. If we let \(\phi = \phi(x, \lambda)\) be the solution of (2.1) defined by the \(\lambda\)-dependent initial conditions
\[ \phi(a, \lambda) = a_2 - a_1^2 \lambda \]  
where \( \lambda \in \mathbb{C} \) is the set of complex numbers.

and if we put
\[ \Phi(x, \lambda) = \frac{\phi(x, \lambda)}{\alpha} \text{ in } H \]

where \( H = L_2(a, b) \times \mathbb{C} \) (with inner product
\[ (F, G) := \int_a^b F_1 G_1 dx + \frac{1}{\alpha} F_2 G_2 \]

for all
\[ F = \begin{bmatrix} F_1(x) \\ F_2 \end{bmatrix}, \quad G = \begin{bmatrix} G_1(x) \\ G_2 \end{bmatrix}, \]

then the eigenfunction expansion associated with (2.1)-(2.3) has the form
\[ F = \int_a^b T(F) \Phi(x, \lambda) dp_b(\lambda) \]

for all \( F \) in the Hilbert space \( H \), where
\[ T(F) := \int_a^b F_1(x) \phi(x, \lambda) dx + \frac{1}{\alpha} F_2 \alpha, \]

and \( p_b(t) \) is the step-spectral function over \([a, b]\). Here, the convergence is in the sense of the norm on \( H \). The formula for the step-spectral function is
\[ (2.5) \]

where \( \{\lambda_{nb}\} \) are the eigenvalues of (2.1)-(2.3) and \( \{r_{nb}\} \) are the reciprocals of the squares of the norms of the corresponding eigenfunctions, that is,
\[ r_{nb} = \frac{1}{\| \Phi(x, \lambda_{nb}) \|^2} \]

The most important property of \( \rho_b(t) \), and the reason the above problem was selected as a test case for the numerical code, is the known asymptotics as \( t \to \infty \), namely,
\[ \lim_{t \to \infty} \rho_b(t) = \frac{1}{\alpha}, \text{ for all } b > a. \]

Moreover, under limit point assumptions at \( \infty \), if \( \rho(t) \) is the singular spectral function associated
with (2.1) and (2.2) over \([a, \infty)\), then

\[
\lim_{t \to \infty} p(t) = \frac{1}{\alpha}.
\]  

(2.8)

It is known [Fulton and Pruess 1989] that

\[
\lim_{b \to \infty} p_b(t) = p(t) \text{ uniformly for } t \in I,
\]  

(2.9)

where \(I\) is any closed interval (finite or infinite) entirely contained in the continuous spectrum. This result provides the mathematical basis for the code SPDNSF which computes approximations to \(p(t)\) by computing \(p_b(t)\) for sufficiently large \(b\).

Numerically, the code SPDNSF computes a large number of the \(\lambda_{nb}\)'s and \(r_{nb}\)'s in (2.5) and gives output for \(p_b(t)\) at certain specified points \(t\). The parameter \(b\) is a user input, and so the length of the interval is easily increased without changing the output points where \(p_b(t)\) is computed.

Known asymptotic formulas [Fulton 1982] for \(\lambda_{nb}\) and \(r_{nb}\), for large \(n\), are used whenever possible. The formulas differ depending on the nature of the boundary conditions. For example, if \(b_2 = 0\) (Dirichlet boundary condition at \(b\)) and \(a_2 = 0\),

\[
\lambda_{nb} = \lambda_{nb}^* := \left[ \frac{n\pi}{b-a} \right]^2 + \frac{2}{b-a} \left[ \frac{a_2}{a_1} + \frac{1}{2} \int_a^b q \, dt \right].
\]  

(2.10)

Similarly, the asymptotic formula for \(r_{nb}\) in this case is

\[
r_{nb} = r_{nb}^* := \frac{2(b-a)}{(a_1 n\pi)^2}.
\]  

(2.11)

3. How to Use the Code

The central routine is subroutine SPDNSF; all other routines are called by it. The user need only become familiar with the parameter list for SPDNSF and write a calling program to drive it. In addition, the user must supply a function subprogram for the potential function \(q(x)\) whose form is discussed at the end of this section.

The call to SPDNSF has the general form

\[
\text{CALL SPDNSF}(\text{CONS, ABSERR, RELERR, NSAVE, NUMT, T, NMAX,}
\]

\[
\text{IPRINT, STORE, ISTORE, RHO, OUTEV, OUTRN, IFLAG})
\]

All real variables must be declared DOUBLE PRECISION in the present version of the code (automatic compiler options exist for converting to a single-precision version on a CRAY X-MP). The parameters have the following significance:
Input:

CONS
is an 8-vector that holds the data (in order) $a_1, a_1', a_2, a_2', b_1, b_2, a, b$.

ABSERR, RELERR
are scalar tolerance parameters specifying the absolute and relative error accuracy requested. In particular, the code attempts to control errors in such a way that

$$|\text{error estimate}| \leq \max(\text{ABSERR}, |\text{answer estimate}| \times \text{RELERR})$$

in those places where the tolerances are used. Typically, $\text{Abserr} \leq \text{Relerr} \times 10^{-2}$ and $\text{Relerr}$ in the range $10^{-3}$ to $10^{-6}$ are appropriate choices for the double-precision arithmetic.

NSAVE
is a positive integer equal to the number of $\lambda_{mb}, r_{mb}$ pairs that are returned to the calling program. Typically, one wants to compute a large number (1000 or more) of eigenvalues and norms but only wishes to see a few (NSAVE $\approx$ 20) actually printed out.

NUMT
is a positive integer equal to the number of discrete points $t$ where $\rho_b(t)$ is to be computed. $T$ is a vector dimensioned at least NUMT; this contains the values of $t$ where $\rho_b(t)$ is to be computed. Usually it is wise to have many points close to zero and then let them spread out toward $\infty$. This is because (for the case $q(x) \to 0$ as $x \to \infty$) the continuous spectrum is in $[0, \infty)$ and the spectral functions $\rho_b(t)$ for large $b$ experience the most rapid growth near $t = 0$. Of course, if it is known that there are a few negative eigenvalues, then one might wish to include some negative values of $t$ in $T$. It is assumed that the points are ordered so that

$$T(1) < T(2) < ... < T(\text{NUMT})$$

but the code does not check this.

NMAX
is an integer used to allocate storage. Roughly speaking, it is the maximum number of shooting points expected. If this is too small, then there will be insufficient storage in STORE and ISTORE, and the code returns the values IFLAG(1, )$\to$-3 or -4. NMAX is unfortunately difficult to anticipate in advance; it depends critically on the interval length $b-a$ and on the tightness of the tolerance parameters since the total number of mesh points NX in the STORE vector consumes the bulk of the space, and this is determined by the local error control which selects the mesh for each shot. Experience shows that the smallest
allowable NMAX generally increases linearly with $b-a$ and increases as ABSERR and RELERR are decreased. It is also somewhat potential dependent, particularly with potentials that are difficult to integrate. Generally, our experience has been that for $b-a$ ranging from 100 to 5000, NMAX will range from 100 to about 1000.

**IPRINT**

is an integer indicating whether or not intermediate printing is desired. If IPRINT=0, then no printing is done; otherwise, information is printed concerning initialization, and the switchover points for the asymptotic formulas.

**STORE**

is a vector of auxiliary storage. Its length must be at least $27 \times NMAX + 15$. If RCOM in function $Q$ is a vector of length $M$, then an additional $M - 1$ locations are needed in STORE.

**ISTORE**

is an integer vector of auxiliary storage. Its length must be at least $3 \times NMAX + 11$. If ICOM in function $Q$ is a vector of length $M$, then an additional $M - 1$ locations are needed in ISTORE.

**Output**

**RHO**

is a vector dimensioned at least NUMT in the calling program; it contains the computed values of the spectral density function $\rho_b(t)$ at the NUMT points $t$ in $T$.

**OUTEV**

is a vector dimensioned at least NSAVE in the calling program; it contains the values of the first NSAVE computed eigenvalues $\lambda_{nb}$. In particular,

$$\text{OUTEV}(I+1) = \lambda_{nb} \text{ for } 0 \leq I \leq \text{NSAVE}-1.$$  

**OUTRN**

is a vector dimensioned at least NSAVE in the calling program; it contains the values of the first NSAVE computed quantities $r_{nb}$. In particular,

$$\text{OUTRN}(I+1) = r_{nb} \text{ for } 0 \leq I \leq \text{NSAVE}-1.$$  

**MAXEV**

is an integer giving the total number of eigenvalues actually needed (to get to $\rho_b(t_{\text{NUMT}})$). This output is computed prior to the beginning of the computation of eigenvalues and reported to the screen; it is also passed back to the driver through Istore (10). When all eigenvalues up to $n(\text{NUMT})$ have been computed, MAXEV is checked against the actual number of eigenvalues computed, and a warning message is printed if it does not agree.
IFLAG

is an integer array dimensioned 3 by NSAVE; it contains diagnostic information about the
NSAVE output values in vectors OUTEV and OUTRN. In detail, if IFLAG(1,I+1) =

positive integer

then $\lambda_{lb}$ and $r_{lb}$ have been computed to within the requested tolerance; this is the normal
return. The value of the integer is the number of subintervals used in the crudest of three
meshes used in the most recent shot.

-1 then the matrix used to compute $r_{lb}$ (for a negative $\lambda$) was exactly singular.

-2 then the step size used by the shooting algorithm is too small. Use cruder tolerances.

-3 then there is insufficient storage in auxiliary vectors STORE and ISTORE. To overcome
this problem, increase NMAX and the declared dimensions of STORE and ISTORE in the
driver.

-4 then there is insufficient storage in auxiliary vectors STORE and ISTORE in the initialization
phase. Increase NMAX as in the case IFLAG(1, I+1) = -3.

-5 then the maximum allowable number of iterations for the inverse iteration to compute $r_{lb}$
was exceeded for this $I$. The output value may be unreliable.

-6 then the sign of $F(b, \lambda)$ following a call to $F$ is not in agreement with the number of zeros
counted in $(a,b)$. The user currently has no options to overcome this difficulty. Testing of
the code resulted in the decision to make calls to the integrator with the tolerances ABSERR
and RELERR internally set to $10^{-3}$ times the user inputs for these quantities, with the
consequence that this error return is almost never activated.

-7 then the tolerances to be used by $F$ are too close to the unit roundoff to justify making the
shot. Use cruder tolerances.

-8 then the brackets for the eigenvalue $\lambda_{lb}$ are not correct according to the zero counts obtained
for the upper or lower bound. This error return causes premature termination of the spectral
function computation and occurs rarely. If it occurs for $I \geq NUBND$, then the asymptotic
brackets are employed too soon, and NUBND should be increased by overriding the
NSWTCH output. Otherwise the brackets generated by START are suspect.

-9 then the input is illegal. Input parameters must satisfy $\alpha = a_1 a_2 - a_2 a_1 > 0$, either $b_1 = 0$
or $b_2 = 0$ but not both, NUMT > 0, NMAX > 0, RELERR > ABSERR > 0, and RELERR >
$10^4 \times$ (unit roundoff).

IFLAG(2,I+1)
gives the total number of $Q$ evaluations needed for the calculation of $\lambda_{lb}$. This is a fairly
good indicator of the amount of effort expended.
IFLAG(3,I+1)

gives the number of integration shots needed to compute $\lambda_b$; this is essentially the number of iterations needed by the zero finder.

The user must supply a function subprogram for the potential $q(x)$, called $Q$, which has the following form:

```fortran
DOUBLE PRECISION FUNCTION Q(X,ICOM,RCOM)
DOUBLE PRECISION X,RCOM
Q=
RETURN
END
```

Here $X$ is the independent variable, while $ICOM$ and $RCOM$ are used to pass other parameters (integer and real, respectively) if needed. For most problems $ICOM$ and $RCOM$ can be ignored. If used, they must be declared in $Q$ in a manner (scalar or vector or array) consistent with their declaration in the driver. One example would be if the user wanted to know the total number of calls to $Q$ in a single usage of SPDNSF. To do this, initialize $ICOM=0$ in the driver, and insert $ICOM=ICOM+1$ in the function routine $Q$. One use of $RCOM$ would be if $q$ depended on another parameter, say $r$, as well as $x$. Then $r$ could be defined in the driver and passed to $Q$ through $RCOM$. Note that if either of these parameters is declared as vectors (or arrays), then the lengths of $STORE$ and $ISTORE$ must be modified accordingly.

Several auxiliary routines from other sources have been used: (1) ZERO is a root finder taken from [Allen, Pruess, and Shampine 1989]; (2) ADAPT, an adaptive quadrature routine also from [Allen, Pruess, and Shampine 1989]. The latter two routines could be replaced by equivalent ones, if desired. ZERO is called by SPDNSF only, while ADAPT is called by SPDNSF and NSWTCH.

A sample driver can be found in the Appendix.

4. Details of Subprograms

4.1. SPDNSF

SPDNSF is called by the user's driver program and calls ADAPT, ASYMEV, ASYMR, BISECT, CMPRN, F, GETMAX, NSWTCH, START, and ZERO. Its main functions are initialization, verifying input, calling routines for the calculation of brackets for the shooting algorithm, calling the shooting algorithm, determining when to switch to the asymptotic formulas, calling the appropriate routines, and calculating the spectral function $P_b(t)$ at the desired points.

The initialization phase involves transferring variables from CONS into the storage area $STORE$, calculating the integral of the potential function $q(x)$ (using the adaptive quadrature
routine ADAPT), and setting some local variables needed for the calculation of $\rho_b(t)$. Illegal input will cause IFLAG(1,1) to be set to -9. If RELERR $\leq 10^9 \times$ (unit roundoff), then RELERR is considered too small to drive the integrator (since the integrator uses $10^{-3} \times$ RELERR in place of RELERR) and calls to $F$ will cause IFLAG(1,) to be set to -7.

The subroutine NSWTCH is called to determine the value NUBND for which the asymptotic eigenvalues can be used to generate brackets for the zero finder. To find brackets for the eigenvalues $\lambda_{ib}$, $0 \leq i < NUBND$, subroutine START is called. The output from this routine is the vector UBND(*) satisfying

$$UBND(n) < \lambda_n \leq UBND(n+1), \quad 0 \leq n < NUBND.$$  

The subroutine BISECT or ZERO is called to refine the bracket for each eigenvalue until the requested tolerance is met. In brief, an approximation $u(x)$ to the eigenfunction is initialized as in (2.4), then the differential equation (2.1) is integrated to $x = b$. Define

$$F(b,\lambda) := b_1 u(b) + b_2 u'(b). \quad (4.1)$$

The routines iterate on $\lambda$ until the upper and lower bounds, ER and EL, for an eigenvalue satisfy the tolerance test

$$ER-EL \leq \max(\text{ABSERR}, \text{RELERR} \| \lambda_n \|).$$

When $|\lambda_n - \lambda_{n+1}| \leq \max(\text{ABSERR}, \text{RELERR} \| \lambda_n \|)$ for two successive $n$, then all subsequent eigenvalues are computed by the appropriate asymptotic formula. As a byproduct of the shooting algorithm, $r_{nb}$ is produced reliably for $\lambda_{nb} \geq 0$. For $\lambda_{nb} < 0$ the routine CMPRN is called to produce $r_{nb}$. When $|r_{nb} - r_{nb}^*| \leq \max(10^{-3} \times \text{ABSERR}, 10^{-3} \times \text{RELERR} \| r_{nb}^* \|)$ for two successive $n$, then all subsequent $r_{nb}$ are computed by the appropriate asymptotic formula (4.1). The rationale for using a tighter tolerance test for the switchover to the asymptotic formula for $r_{nb}$ is that the amount of error which $r_{nb}$ can tolerate should be of about the same order of magnitude as the global error in the computed $r_{nb}$. In the absence of global error estimation for $r_{nb}$ the above tolerance test makes use of the same tolerances used by the integrator to control local errors in $r_{nb}$.

For $t \leq 0$, approximations to the singular $\rho(t)$ are obtained by computing $\rho_b(t)$ exactly as given by (2.5). For $t > 0$ the spectrum (under the assumption $q \to 0$ as $x \to \infty$) is continuous, and so for fixed $b$ equation (2.5) does not represent the best attainable approximation to the singular $\rho(t)$. This is particularly true when $t$ is close to zero where both $\rho_b$ and $\rho$ experience the steepest rise. A minor refinement is therefore employed for $t > 0$: four eigenvalues and three successive $\rho_b$ values from (2.5) are employed to generate a parabola through the midpoints of three successive steps of $\rho_b$ in the vicinity of $t$, and the value on the parabola is returned as the approximation to $\rho(t)$.
4.2. F

The routine F is called by SPDNSF, START, and the zero finders BISECT and ZERO; it
calls Q, STEP, and EXTRAP. It evaluates the function \( F(b, \lambda) \) given by (4.1) which requires the
integration of the differential equation (2.1). It is well known that typical numerical integrators
are at best inefficient when integrating (2.1) subject to initial conditions. One reason for this is
the eigenfunctions are increasingly oscillatory as \( \lambda \) gets bigger, so a method based on local poly-
nomial approximation requires small step sizes. Instead, F uses a method studied by Pruess
[1973] in which \( q(x) \) is replaced by a step function. This
has several
advantages: first, the resulting problem can be integrated exactly. Second, the solutions (for \( \lambda > 0 \)) involve piecewise sines
and cosines which are better able to approximate oscillatory solutions. For \( \lambda < 0 \) the eigenfunc-
tion calculation is unstable (for the initial value problem) so an alternative algorithm is necessary.
This is described in the section for CMPRN.

The mesh for each shot is chosen adaptively to control the local error in \( \int_a^b y^2(x) dx \). In
detail, for a given \( \lambda \) assume the differential equation has been integrated successfully up to a point
\( x \) with sufficiently accurate estimates \( Y_k = y(x_k) \), \( Y'_k = y'(x_k) \) and \( R_k = \int_a^{x_k} y^2(x) dx \). Then \( q(x) \) is
replaced by \( q(x + h/2) \), and STEP is called to produce \( Y_{k+1} = y(x_k+h) \), \( Y'_{k+1} = y'(x_k+h) \), and
\( R_{k+1} = \int_a^{x_{k+1}} y^2(x) dx \). This procedure is repeated twice using the half step \( h/2 \) and then three times
using the step \( h/3 \). Richardson's \( h^2 \)-extrapolation is then applied to the three estimates for \( Y, Y' \),
and \( R \) using the routine EXTRAP. The results have local error \( O(h^3) \), and a local error estimate is
also available. If the latter is sufficiently small, then the answers are accepted (the local error is
added in), \( h \) is adjusted, and the code proceeds to the next step. If the error is too big, then \( h \) is
reduced, and a complete restart is made at \( x_k \). Finally, this routine also calculates an index \( k \) such that

\[
\lambda_{k-1} < \lambda \leq \lambda_k.
\]

This index is used to determine brackets by the routine START. The formula for \( k \) is

\[
k = \begin{cases} 
NZERO + 1 & a_2 \neq 0, a_2/a_1^2 \leq \lambda \\
NZERO & \text{otherwise}
\end{cases}
\]

where \( NZERO \) is the number of zeros of \( u(x) \) for this \( \lambda \). The derivation of this formula relating
the eigenvalue index to the zero count is contained in [Fulton and Pruess 1989].

4.3. EXTRAP

EXTRAP is called by F and CMPRN; it performs Richardson's \( h^2 \)-extrapolation on three
quantities \( V_1, V_2, V_3 \) corresponding to meshes \( h, h/2, h/3 \), respectively.
4.4. STEP

STEP is called by F and CMPRN; it advances the solution of the approximate differential equation \(-u'' + Qu = \lambda u, Q\) a constant, one step of length \(H\). Given values \(U(x), U'(x)\), and a current \(\lambda\), let

\[
\tau := \lambda - Q, \quad w := \sqrt{|\tau|}. \quad (4.2)
\]

Then

\[
U(x+H) = U(x)G_1(H) + U'(x)G_2(H)
\]

\[
U'(x+H) = (U(x+H)G_1(H) - U(x))/G_2(H)
\]

\[
\int_a^{x+H} U^2(x)dx = \int_a^x U^2(x)dx + (U'(x))^2 [H - G_1(H)G_2(H)] / [2(\lambda - Q)] \quad (4.3)
\]

\[
+ U^2(x)[H + G_1(H)G_2(H)]/2 + U(x)U'(x)G_2(H),
\]

where

\[
G_1(H) = \begin{cases} \cos wH & \tau > 0 \\ \cosh wH & \tau < 0 \end{cases}
\]

\[
G_2(H) = \begin{cases} \sin(wH)/w & \tau > 0 \\ \sinh(wH)/w & \tau < 0 \end{cases}. \quad (4.4)
\]

For small \(|\tau|\), MacLaurin expansions are used for \(G_1\) and \(G_2\). The number of roots of \(U\) can also be calculated in this routine by setting MODE > 1.

4.5. CMPRN

CMPRN is called by SPDNSF and calls INVIT, EXTRAP, and STEP. It computes \(\int_a^b y^2(x)dx\) for eigenfunctions corresponding to negative eigenvalues. In this case \(\tau\) in (4.2) is usually negative, so hyperbolic functions occur in (4.4). The initial value problem (2.1), (2.4) is unstable in this case; consequently, the shooting algorithm is suspect. However, since BISECT and ZERO always maintain a bracket, the eigenvalue calculation is acceptable as long as overflow is avoided. The eigenfunction calculation, though, must be done by a stable algorithm, e.g., one based on the boundary value problem (2.1)-(2.3).

Given a mesh \(\{a = x_1 < x_2 < ... < x_{N+1} = b\}\), chosen by the shooting algorithm, let

\[
Q := q((x_n + x_{n+1})/2), \quad \tau_n := \lambda - Q, \quad w_n = \sqrt{|\tau_n|}
\]

on \([x_n, x_{n+1}]\) in analogy with (4.2). Then an alternative representation for \(u(x)\), in terms of \(u(x_n)\)
and \( u(x_{n+1}) \) instead of \( u(x_n) \) and \( u'(x_n) \), yields
\[
-u(x_{n-1})G_2(h_{n-1}) + [G_1(h_{n-1})/G_2(h_{n-1})] + G_1(h_n)/G_2(h_n)u(x_n) - u(x_{n+1})/G_2(h_n) = 0
\]
for \( 1 < n < N \) with \( G_1 \) and \( G_2 \) as in (4.4). The two boundary conditions (2.2) and (2.3) provide two more homogeneous equations. If \( \lambda \) is indeed an eigenvalue, then \( \{u(x_n)\}_{n=1}^{N+1} \) is a null vector for this system which can be computed by inverse iteration. The latter calculation is performed in INVIT. Three such calculations are actually made: the second for a bisected mesh, the third for a trisected mesh. Then Richardson's \( h^2 \)-extrapolation is used to produce a final estimate for \( \int_a^b u(x)^2 \, dx \) corresponding to the current \( \lambda \).

4.6. INVIT

INVIT is called by CMPRN; it uses inverse iteration to solve the symmetric tridiagonal system given by (4.5) and the two boundary conditions. If this system is denoted by \( Du = 0 \), then with \( u^{(0)} := (0, 1, \ldots, 1, 0) \) solve \( Du^{(k+1)} = u^{(k)}/\sigma_k \) for \( u^{(k+1)} \). Here \( \sigma_k \) is a scale factor chosen so that an initial condition like (2.4) is satisfied. Since \( D \) is tridiagonal, the solution by Gauss elimination is cheap; however, to ensure stability, partial pivoting must be allowed. The iteration continues until \( \int_a^b [u^{(k+1)}]^2 - (u^{(k)})^2 \, dx \) is sufficiently small; the maximum number of iterations allowed is a function of the tolerance and the interval width \( b-a \).

4.7. ASYMEV

ASYMEV is called by SPDNSF, GETMAX, and START; it implements the appropriate asymptotic formula for \( \lambda_{ab}^a \), for example, (2.10).

4.8. ASYMR

ASYMR is called by SPDNSF; it implements the appropriate asymptotic formula for \( r_{ab}^a \), for example, (2.11).

4.9. START

START is called by SPDNSF and calls F, RAND, and ASYMEV. It provides the initial brackets for the shooting algorithm. First, values ULOW and UHIGH are computed satisfying
\[
ULOW \leq \lambda_0 \quad \text{and} \quad \lambda_{NUBND-1} \leq UHIGH.
\]
The search for these values uses the asymptotic formulas for the eigenvalues. For each index \( k \) between 1 and NUBND, a value for UBND(k) is sought such that
\[ \lambda_{k-1} < \text{UBND}(k) < \lambda_k. \]

A binary search is performed on [ULOW, UHIGH] with the zero count from \( F \) yielding the location of the eigenvalues according to the oscillation theory associated with the \( \lambda \)-dependent boundary condition (2.2).

### 4.10. NSWTCH

NSWTCH is called by SPDNSF and calls Q and ADAPT. It uses the asymptotic formulas for \( \lambda \) to determine when these asymptotic eigenvalues separate the actual eigenvalues. If the expansion (2.10) is carried out further, there results (for the case \( b_2 = 0 \) and \( a_2 = 0 \))

\[
\lambda_n + \left( \frac{n\pi}{b-a} \right)^2 + \frac{2c_1}{b-a} + \frac{1}{(n\pi)^2} (c_1^2 + 2c_3) + O(n^{-4}).
\] (4.6)

where

\[
c_1 = \frac{1}{2} \int_a^b q(t) dt + a_2/a_1^*
\]

\[
c_3 = c_1^2 + (b-a)\left( c_1(c_1^2/3 - R/a_1^*) + S/a_1^* \right)
\]

\[
R = a_1 - a_1^*(q(a) + q(b))/4 + \frac{1}{2} \int_a^b q(t) dt \left[ a_2 + a_1^* \int_a^b q(t) dt/4 \right]
\]

\[
S = a_2(q(b) - q(a))/4 - a_1^*(q'(b) - q'(a))/8 + a_1^* \int_a^b q^2(t) dt/8
\]

\[
+ \int_a^b q(t) dt \left[ -a_1^*/2 + a_1^*(q(b) + q(a))/8 + \int_a^b q(t) dt \left[ -a_2/8 - a_1^* \int_a^b q(t) dt/48 \right] \right].
\]

To ensure that \( \lambda_n \approx \lambda_n^* \), we choose \( n = \text{NUBD} \) as the smallest integer for which

\[
\frac{|c_1^2 - 2c_3|}{n^2 \pi^2} \leq \sigma \frac{(2n + 1)\pi^2}{(b-a)^2}.
\]

This reduces to
\[ n > \left( \frac{b-a}{\pi} \right)^2 \frac{|c_1^2 + 2c_3|}{2\sigma} \right)^{1/3} \]

We have used \( \sigma = 1/3 \) to produce NUBND from this formula.

4.11. Q2

Q2 is called by ADAPT (from NSWTCH) and is the function subprogram for \( q^2(x) \).

4.12. GETMAX

GETMAX is called by SPDNSF and calls ASYMEV; it calculates MAXEV, the total number of eigenvalues needed for the spectral function computation, by relying on the asymptotic formula for \( \lambda_{nb} \).

4.13. RAND

RAND is called by INVIT and START; it provides uniformly distributed random numbers.

4.14. BISECT

BISECT is called by SPDNSF and calls F. It implements the standard bisection method to provide a zero of F.

4.15. STORAGE ALLOCATION in STORE and ISTORE

The locations in the auxiliary vectors STORE and ISTORE are apportioned as follows:

<table>
<thead>
<tr>
<th>STORE(1)</th>
<th>STORE(15)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a_1</td>
</tr>
<tr>
<td>2</td>
<td>a_1</td>
</tr>
<tr>
<td>3</td>
<td>a_2</td>
</tr>
<tr>
<td>4</td>
<td>a_2</td>
</tr>
<tr>
<td>5</td>
<td>b_1</td>
</tr>
<tr>
<td>6</td>
<td>b_2</td>
</tr>
<tr>
<td>7</td>
<td>a</td>
</tr>
<tr>
<td>8</td>
<td>b</td>
</tr>
<tr>
<td>9</td>
<td>ABSERR</td>
</tr>
<tr>
<td>10</td>
<td>RELERR</td>
</tr>
<tr>
<td>11</td>
<td>( r_{nb} )</td>
</tr>
<tr>
<td>12</td>
<td>Estimate of unit roundoff</td>
</tr>
<tr>
<td>13</td>
<td>Used to assess stability</td>
</tr>
<tr>
<td>14</td>
<td>ULOW</td>
</tr>
<tr>
<td>15</td>
<td>(-2\times NMAX+14)</td>
</tr>
</tbody>
</table>
5. Computational Examples and Timings

Subroutine SPDNSF has been tested on 14 potential functions over a range of 8 different choices of the boundary condition parameters in (2.2) and both Dirichlet and Neumann boundary conditions in (2.3). Much of the testing has been done on the CRAY X-MP at the National Center for Supercomputing Applications at the University of Illinois, Urbana-Champaign. Runs made over large intervals using tight tolerances frequently revealed shortcomings in the code, the choice of tolerance tests for the integrator, and the tolerance tests for the switchover point to the asymptotic $r_{nb}$ being of particular significance. The problems encountered have resulted in many significant improvements to the code. In particular, in order to generate a more appropriate mesh over large intervals, the basic tolerance test for controlling local errors in $r_{nb}$ was refined; to avoid the problem of the sign of $F(b,\lambda)$ often not being in agreement with the count of zeros in the solution, it was decided to run the integrator at a significantly tighter tolerance than that used by the zero finder (BISECT and ZERO); and to retain accuracy in $\rho(t)$ for large $t$, the tolerance test
for switching to the asymptotic $r_{nb}$ was changed to have the same tightness as the calls to the integrator $F$.

Currently, subroutine SPDNSF is capable of achieving 4-5 decimal places of accuracy in the singular spectral function $\rho(t)$, uniformly over very large $t$-intervals, say $0 \leq t \leq 50,000$. However, effective use of the code at its current stage of development requires the user to develop some experience in making good choices of the interval length, $b-a$, corresponding to the choices of ABSERR and RELERR, if he expects to maximize the accuracy in the approximation to the singular $\rho(t)$. Generally, one can expect to achieve maximum accuracy by using $b-a$ in the range 100 to 5000 and $\text{ABSERR} \leq \text{RELERR} \times 10^{-2}$, where RELERR is in the range $10^{-3}$ to $10^{-6}$, assuming double-precision arithmetic. As a rule, accuracy will increase with increasing $b$ (a fixed) and will increase with decreasing ABSERR and RELERR. Also, one should request output for $\rho(t)$ at many closely spaced points $t$ for $t \text{ near zero}$ (where $\rho$ experiences a sharp rise) and let the $t$ points spread out significantly for large $t$. The choice of $b$ fixes the number of eigenvalues in $(-\infty, t_{(\text{NUMT})})$, and the choices of ABSERR and RELERR determine the accuracy with which the eigenvalues and $r_{nb}$'s are computed. For fixed $b$, it is usually the case, but not necessarily always the case, that decreasing ABSERR and RELERR will produce more accuracy in the $\rho(t)$-values being computed. For if $b$ is small (usually $b-a \leq 100$), the accuracy in $\rho(t)$ is limited by the use of (2.5) which has "large" jumps at each eigenvalue close to zero, so that highly accurate values for $\lambda_{nb}$ and $r_{nb}$ do not help to recover from the accuracy lost by the poor approximation to $\rho(t)$ afforded by $\rho_{p}(t)$; of course, the use of a parabola through 3 points of $\rho_{p}(t)$ improves the latter approximation somewhat, especially for small $t$. Similarly, for fixed ABSERR and RELERR, increasing $b$ may or may not result in higher accuracy in the $\rho(t)$-values being computed. For if $b$ is too large (usually $b-a > 1000$), the number of eigenvalues and $r_{nb}$ being computed can be too large for the given tolerances; that is, ABSERR and RELERR may not be small enough to achieve a high enough accuracy in $r_{nb}$ for very large $n$, and this will result in less accuracy in $\rho(t)$ for large $t$. Accordingly, higher accuracy in $\rho(t)$ is normally achieved by decrementing ABSERR and RELERR as $b$ is increased. Exactly how this should be done is difficult to say, because there is some potential-dependence in trying to achieve an optimal $b$ for a given choice of ABSERR and RELERR. Currently, the best approach is to run the code for several increasing values of $b$ for each fixed choice of ABSERR and RELERR and compare output for $\rho(t)$; as long as the $\rho(t)$ values are converging, it is safe to increase $b$. But past a certain point, it does not help to increase $b$, and one should decrement ABSERR and RELERR before increasing $b$ again. Work is currently in progress to try to automate this type of procedure for ensuring accuracy in $\rho(t)$.

As an example, to illustrate the accuracy which SPDNSF is capable of achieving, consider the Bessel equation of order zero on $[1,b]$:

$$
-u'' - (1/4x^2)u = \lambda u, \quad 1 \leq x < \infty
$$

(5.1)
\begin{equation}
(a_1 u(1) - a_2 u'(1)) = \lambda (a'_1 u(1) - a'_2 u'(1))
\tag{5.2}
\end{equation}

\begin{equation}
b_1 u(b) + b_2 u'(b) = 0.
\tag{5.3}
\end{equation}

In this case, the singular spectral function over \([1, \infty)\) may be obtained in closed form as in \([Fulton 1980]\) and is found to be

\begin{equation}
\rho(t) = \begin{cases}
0, & t < t_0 \\
t_0, & t_0 \leq t < 0 \\
t_0 + \int_0^t f(s) \, ds, & 0 \leq t < \infty
\end{cases}
\tag{5.4}
\end{equation}

where

\begin{equation}
f(s) = \frac{2}{\pi^2 \sqrt{[(As+B) J_0(\sqrt{s}) - \sqrt{s}[Cs+D] Y_1(\sqrt{s})]^2 + [(As+B) Y_0(\sqrt{s}) - \sqrt{s}[Cs+D] J_1(\sqrt{s})]^2}},
\tag{5.5}
\end{equation}

\begin{equation}
t_0 = \frac{E K_0(t_0) - t_0 F K_1(t_0)}{[A + (Ct_0^2 + D) Y_1(t_0)] K_0(t_0) - \frac{1}{t_0} \sqrt{[t_0 A - (A/2 - 2C/2) + (B + D) Y_2] K_1(t_0) + [(Ct_0^2 + D) Y_4] K_2(t_0)}}.
\tag{5.6}
\end{equation}

\begin{align*}
A &= a_1' - a_2'/2, & B &= -a_1 + a_2/2, & C &= -a_2', & D &= a_2, \\
E &= (a_2'/2 - a_1')/\alpha, & F &= a_2'/\alpha,
\end{align*}

and \(t_0 = -s_0^2\) where \(s_0\) is a root (of which there are at most two) of

\begin{equation}
(-As^2 + B) K_0(s) - s(-Cs^2 + D) K_1(s) = 0.
\tag{5.7}
\end{equation}

The singular spectral function (5.4) was computed to high accuracy using Bessel function routines from the Fullerton Special Function library in SLATEC, and compared with output for \(\rho(t)\) from SPDNSF. Table 1 contains the output for the "exact" spectral function \(\rho(t)\) from (5.4), together with output from SPDNSF for 27 \(t\)-values ranging from 0 to 50,000. These runs of SPDNSF were performed on the CRAY X-MP at the National Center for Supercomputing Applications, Urbana, Illinois. The driver for these runs, in double precision, is given in the Appendix. The runs were actually performed in single precision on the CRAY X-MP using an automatic compiler option for converting double precision code to single precision; single precision on the CRAY X-MP is equivalent to double-precision on VAXes or SUNs. Two runs of SPDNSF are given, the first arising from \(b = 500\) and ABSERR = 0.5D-6, RELERR = 0.5D-4, and the second from \(b = 1000\) and ABSERR = 0.5D-8, RELERR = 0.5D-6. The other user inputs and some of the other outputs from SPDNSF are given in Table 2. Comparison with the exact values for \(\rho(t)\) shows that the first run with \(b = 500\) achieved about 5 decimal places of accuracy uniformly for all 27 \(t\)-values over the range 0.0 to 50,000, and that the second run with \(b = 1000\) achieved about 6 decimal places of accuracy. On the other hand, the first run required a total of 50.2 sec of
CRAY CPU time, while the second run required 12.6 min. of CRAY CPU time, roughly 15 times as much computation. Detailed timing data for these two runs are listed in Table 3 along with detailed data on the number \( NQ \) of function evaluations of the potential function \( q(x) \) for each of four phases of the code. The four phases are defined as follows for the purpose of this timing analysis:

Phase 1. This is the initialization phase which consists of finding subintervals (called brackets) each of which contains exactly one eigenvalue and whose union contains all of the eigenvalues up to the index NUBND. The index NUBND is determined by a call to NSWTCH, and the brackets are obtained by a call to START.

Phase 2. This phase is defined as the part of the code that computes the eigenvalues \( \lambda_{nb} \) and norm reciprocals \( r_{nb} \) from \( n = 1 \) up to the index where the tolerance test for the asymptotic eigenvalues determines that it is permissible to switch to the asymptotic formulas for the eigenvalues. Normally, the latter index is greater than NUBND, and the brackets for eigenvalues with index greater than NUBND are determined by appropriate use of the asymptotic formulas, together with the previously computed eigenvalue.

Phase 3. This phase is defined as the part of the code that computes \( r_{nb} \) from the switchover point to the asymptotic eigenvalues up to the index where the tolerance test for the asymptotic \( r_{nb} \) determines that it is permissible to switch to the asymptotic formulas for \( r_{nb} \). In this phase, the asymptotic eigenvalues are used, and one shot per eigenvalue is performed to compute \( r_{nb} \).

Phase 4. The last phase of the code consists of computing \( \lambda_{nb} \) and \( r_{nb} \) from the switchover point to the asymptotic \( r_{nb} \) up to \( n = \text{MAXEV} \), where MAXEV is the smallest index such that \( t(\text{NUMT}) < \lambda_{\text{MAXEV},b} \). In this phase, both \( \lambda_{nb} \) and \( r_{nb} \) are computed at a cost of 4-5 flops apiece, and no integration, hence no \( q \)-evaluations, are required.

As the data from Table 3 demonstrate, there is a high correlation between the number of \( q \)-evaluations performed in each phase of the code and the amount of time spent in each phase. It is also evident that the efficiency of the code in computing cheap approximations to the singular \( p(t) \) derives primarily from the use of the asymptotic formulas for the eigenvalues and \( r_{nb} \)'s. For the first run with \( b = 500 \), for example, 99% of the eigenvalues and 77% of the \( r_{nb} \)'s were computed by the asymptotic formulas, while the time spent in Phase 4 was only 0.45% of the total time for the run. Similarly, for the second run with \( b = 1000 \), 97% of the eigenvalues and 38.8% of the \( r_{nb} \)'s were computed by the asymptotic formulas, and the time spent in Phase 4 was only 0.03% of the total time. We can surmise from the timing data that the sharp drop from 77% to 38.8% in the number of \( r_{nb} \) obtained from the asymptotic formulas, and the consequent rise in the number of calls to the integrator required to compute the \( r_{nb} \)'s, is primarily responsible for the jump in run
time from 50.2 sec to 12.6 min. of CRAY time. Of course, the interval length is about twice as long, and the tolerances have been decreased, so the time for each shot can be expected to increase. The fact that 15 times as much computing is required to compute about twice as many eigenvalues and $r_{nb}$'s is also indicative of the fact that the amount of accuracy that can be expected in the computation of each $\lambda_{nb}$ and $r_{nb}$ is limited by the use in the integrator of a Richardson Extrapolation which uses only 3 different meshes and therefore produces $O(h^5)$- approximations locally for the eigenfunction, its derivative, and $r_{nb}$. A higher order Richardson extrapolation with more meshes contributing on each local step can be expected to enable higher accuracy in $p(t)$ to be obtained. Generally, it has been our experience that computation times need not be more than a minute or two of CRAY time to achieve 4-5 decimal places of accuracy in $p(t)$, and that attempts to obtain higher accuracy result in sharp rises in computing time.

Acknowledgment

We thank Martin J. Vukovich, David C. Chan, and Abootaleb Vafaie-Baghaki of the Florida Institute of Technology for their assistance in testing and refining the present version of SPDNSF.

References


### Table 1. Outputs

<table>
<thead>
<tr>
<th>( t_i )</th>
<th>Exact ( p(t_i) )</th>
<th>( \text{SPDNSF RUN 1, } b=500 )</th>
<th>( \text{SPDNSF Run 2, } b=1000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>0.17803784</td>
<td>0.17803771</td>
<td>0.17803786</td>
</tr>
<tr>
<td>0.100</td>
<td>0.24722927</td>
<td>0.24722257</td>
<td>0.24722765</td>
</tr>
<tr>
<td>0.200</td>
<td>0.29887464</td>
<td>0.29887055</td>
<td>0.29887352</td>
</tr>
<tr>
<td>0.300</td>
<td>0.33944026</td>
<td>0.33943745</td>
<td>0.33943939</td>
</tr>
<tr>
<td>0.400</td>
<td>0.37298827</td>
<td>0.37298632</td>
<td>0.37298758</td>
</tr>
<tr>
<td>0.500</td>
<td>0.40156621</td>
<td>0.40156354</td>
<td>0.40156565</td>
</tr>
<tr>
<td>0.600</td>
<td>0.42639590</td>
<td>0.42639378</td>
<td>0.42639542</td>
</tr>
<tr>
<td>0.700</td>
<td>0.44826824</td>
<td>0.44828103</td>
<td>0.44828246</td>
</tr>
<tr>
<td>0.800</td>
<td>0.46779342</td>
<td>0.46779183</td>
<td>0.46779303</td>
</tr>
<tr>
<td>0.900</td>
<td>0.48534144</td>
<td>0.48534270</td>
<td>0.48534381</td>
</tr>
<tr>
<td>1.000</td>
<td>0.50125157</td>
<td>0.50125034</td>
<td>0.50125128</td>
</tr>
<tr>
<td>1.100</td>
<td>0.51576251</td>
<td>0.51576134</td>
<td>0.51576225</td>
</tr>
<tr>
<td>1.200</td>
<td>0.52907334</td>
<td>0.52907229</td>
<td>0.52907310</td>
</tr>
<tr>
<td>1.500</td>
<td>0.56325761</td>
<td>0.56325663</td>
<td>0.56325743</td>
</tr>
<tr>
<td>2.000</td>
<td>0.60674116</td>
<td>0.60674048</td>
<td>0.60674103</td>
</tr>
<tr>
<td>5.000</td>
<td>0.73112698</td>
<td>0.73112340</td>
<td>0.73112691</td>
</tr>
<tr>
<td>10.000</td>
<td>0.80443829</td>
<td>0.80443737</td>
<td>0.80443823</td>
</tr>
<tr>
<td>20.000</td>
<td>0.85970589</td>
<td>0.85970542</td>
<td>0.85970581</td>
</tr>
<tr>
<td>50.000</td>
<td>0.91049203</td>
<td>0.91049164</td>
<td>0.91049195</td>
</tr>
<tr>
<td>100.000</td>
<td>0.93652345</td>
<td>0.93652306</td>
<td>0.93652337</td>
</tr>
<tr>
<td>200.000</td>
<td>0.95504978</td>
<td>0.95504939</td>
<td>0.95504970</td>
</tr>
<tr>
<td>500.000</td>
<td>0.97154610</td>
<td>0.97154570</td>
<td>0.97154601</td>
</tr>
<tr>
<td>1000.000</td>
<td>0.97987418</td>
<td>0.97987378</td>
<td>0.97987409</td>
</tr>
<tr>
<td>2000.000</td>
<td>0.98576682</td>
<td>0.98576642</td>
<td>0.98576673</td>
</tr>
<tr>
<td>5000.000</td>
<td>0.99099736</td>
<td>0.99099780</td>
<td>0.99099727</td>
</tr>
<tr>
<td>10000.000</td>
<td>0.99363398</td>
<td>0.99363477</td>
<td>0.99363389</td>
</tr>
<tr>
<td>50000.000</td>
<td>0.99715296</td>
<td>0.99715392</td>
<td>0.99715293</td>
</tr>
<tr>
<td>Inputs to SPDNSF</td>
<td>RUN 1</td>
<td>RUN 2</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>-------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>$b$</td>
<td>500.0</td>
<td>1000.0</td>
<td></td>
</tr>
<tr>
<td>$a_1, a_1', a_2, a_2'$</td>
<td>0.0, 1.0, 1.0, 0.0</td>
<td>0.0, 1.0, 1.0, 0.0</td>
<td></td>
</tr>
<tr>
<td>$b_1, b_2$</td>
<td>1.0, 0.0</td>
<td>1.0, 0.0</td>
<td></td>
</tr>
<tr>
<td>$NMAX$</td>
<td>500</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>$NUMT$</td>
<td>27</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>$ABSERR$</td>
<td>0.5D-6</td>
<td>0.5D-8</td>
<td></td>
</tr>
<tr>
<td>$RELERR$</td>
<td>0.5D-4</td>
<td>0.5D-6</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outputs from SPDNSF</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$MAXEV$</td>
<td>35,517</td>
<td>71,106</td>
</tr>
<tr>
<td>Asymptotic $\lambda_{nb}$ start at $n =$</td>
<td>350</td>
<td>1,833</td>
</tr>
<tr>
<td>Asymptotic $r_{nb}$ start at $n =$</td>
<td>8,171</td>
<td>43,496</td>
</tr>
<tr>
<td>$NUBND$</td>
<td>83</td>
<td>166</td>
</tr>
</tbody>
</table>
TABLE 3. Timing Results

<table>
<thead>
<tr>
<th></th>
<th>Run 1, ( b = 500 )</th>
<th></th>
<th>Run 2, ( b = 1000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( NQ = \text{No. } q)-Evaluations</td>
<td>Percent of Total ( NQ )</td>
<td>Time*</td>
</tr>
<tr>
<td>Phase 1</td>
<td>52,420</td>
<td>2.05%</td>
<td>0.884</td>
</tr>
<tr>
<td>Phase 2</td>
<td>756,485</td>
<td>29.60%</td>
<td>13.580</td>
</tr>
<tr>
<td>Phase 3</td>
<td>1,746,810</td>
<td>68.35%</td>
<td>35.536</td>
</tr>
<tr>
<td>Phase 4</td>
<td>0</td>
<td>0.00%</td>
<td>0.225</td>
</tr>
<tr>
<td>Totals</td>
<td>2,555,715</td>
<td>100.00%</td>
<td>50.226</td>
</tr>
</tbody>
</table>

*All times are in seconds of CPU time on the CRAY X-MP.
Appendix

Sample Driver Program for Subroutine SPDNSF
Sample Driver Program for SUBROUTINE SPDNSF. The following
main program for calling SPDNSF was used on the CRAY X-MP
to generate the Output for Run # 1 in TABLE 1.

```fortran
Integer Istore(1511), IFlag(3,500), Nsave, Numt, Nmax,
       & Iprint, Mis
Double Precision Store(13515), Cons(8), Abserr, Relerr, T(27),
       & Rho(27), Outev(500), Outrn(500), BC(6)
External q
Double Precision t1, t2, t3, t4, t5, time1, time2, time3, time4, ttime
Common /Time/ t2, t3, t4

Data NumT/27/
Data T/0.d0, 1.d0, 2.d0, 3.d0, 4.d0, 5.d0, 6.d0, 7.d0, 8.d0, 9.d0,
       & 1.d0, 1.d0, 1.d0, 1.d0, 1.d0, 1.d0, 1.d0, 1.d0, 1.d0, 1.d0,
       & 1.d0, 1.d0, 1.d0, 1.d0, 1.d0, 1.d0, 1.d0, 1.d0, 1.d0, 1.d0,
       & 1.d0, 1.d0, 1.d0, 1.d0, 1.d0, 1.d0, 1.d0, 1.d0, 1.d0, 1.d0,

Load BC with a1, a1', a2, a2', b1, b2

Data BC/0.d0, 1.d0, 1.d0, 0.d0, 1.d0, 0.d0/

DO 5 I=1,6
   CONS(I) = BC(I)
   CONTINUE
CONS(7) = 1.0
CONS(8) = 500.0
NMAX=500
ABSERR = 0.5D-6
RELLR = 0.5D-4

Use Istore(3*Nmax+11) to Select particular q(x) desired.

Nsave = 300
Mis=3*Nmax+11
Istore(Mis) = 1
Iprint = 1
Call Second(t1)
Call Spdnsf(Cons, Abserr, Relerr, Nsave, Numt, T, Nmax,
       & IPRINT, Store, Istore, Rho, Outev, Outrn, IFlag)
Call Second(t5)

write(6,*), 'Maxev=', Istore(10)
write(6,*), 'Potential Func. Number q(x) = ',Istore(Mis)
write(6,*), 'Nmax = ',Nmax
write(6,15) Cons(8), Abserr, Relerr
write(6,20)
Do 34 I=1,Nsave
   write (6,33) I-1, Outev(I), Outrn(I), (IFlag(K,I), K=1,3)
34 Continue
write (6,35)
Do 45 I=1,NumT
   write (6,40) T(I), Rho(I)
45 Continue
write (6,50) Istore(4)
Time1 = t2 - t1
Time2 = t3 - t2
Time3 = t4 - t3
Time4 = t5 - t4
Ttime = t5 - t1
write (6,55) Time1
write (6,56) Time2
write (6,57) Time3
write (6,58) Time4
write (6,59) Ttime
```

100
write(6,60) 100.0 * Time1 / Ttime
write(6,61) 100.0 * Time2 / Ttime
write(6,62) 100.0 * Time3 / Ttime
write(6,63) 100.0 * Time4 / Ttime

C format statements for outputs from spdnsf

15 Format(‘ b=’,F9.2,/,’ AbsErr =’,D10.2,’ RelErr =’,D10.2)
20 Format(‘ I’,10x,’Ev(I)’,15x,’Rsubn(I)’,6x,’ ***Flag*** ’)
33 Format(I4,F25.15,1pd30.15,I7,I9,I5)
35 Format(‘ /’,9x,’Ev’,9x,’Rho’)
40 Format(F12.3,F25.8)
50 Format(‘ /’,’ Total Number Of Q-Evaluations is’,I30)
55 Format(‘ /’,’ Time for Phase I is ’,F20.4)
56 Format(‘ /’,’ Time for Phase II is ’,F20.4)
57 Format(‘ /’,’ Time for Phase III is ’,F20.4)
58 Format(‘ /’,’ Time for Phase IV is ’,F20.4)
59 Format(‘ /’,’ Total Time for SPDNSF is ’,F20.4)
60 Format(‘ /’,’ Percent Time Spent in Phase I is ’,F10.2)
61 Format(‘ /’,’ Percent Time Spent in Phase II is ’,F10.2)
62 Format(‘ /’,’ Percent Time Spent in Phase III is ’,F10.2)
63 Format(‘ /’,’ Percent Time Spent in Phase IV is ’,F10.2)
stop
End

C

c define the potential function q(x).

c
double precision function q(x,icom,rcom)
integer icom
double precision x,rcom
  go to (10,20),icom
10 q=-.25d0/(x*x)
  return
20 q=-1./x
  return
end
Abstract

In this article we establish the best possible value of the constant $K$ in the inequality $\|f\|_{\frac{1}{2}} \leq K \|f\| \|f''\|$ for functions $f$ that are defined and twice continuously differentiable on a compact interval $[a, b] \subset \mathbb{R}$ and whose first derivative vanishes at some point in $[a, b]$.

1. Introduction

Let $[a, b] \subset \mathbb{R}$ be compact. Let $C([a, b])$ denote the space of all real-valued functions defined and continuous on $[a, b]$. We consider $C([a, b])$ as a subspace of the Hilbert space $L^2(a, b)$, using the notation $(\cdot, \cdot)_{[a, b]}$ and $\|\cdot\|_{[a, b]}$ for the $L^2$-inner product and norm, respectively.

Let $C^2([a, b])$ be the subspace of all twice continuously differentiable elements of $C([a, b])$. We introduce the following subsets of $C^2([a, b])$:

$G = \{ f \in C^2([a, b]) : f(a) = 0, f'(b) = 0 \}$, \hspace{1cm} (1.1)

$H = \{ f \in C^2([a, b]) : f(a) = 0, f(b) = 0 \}$. \hspace{1cm} (1.2)

**Theorem 1.** Every $f \in G \cup H$ with $f'' \neq 0$ satisfies the inequality

$$\|f'\|_{[a, b]} \leq \|f\|_{[a, b]} \|f''\|_{[a, b]}.$$ \hspace{1cm} (1.3)

Equality holds if and only if $f$ is a constant multiple of $\sin(n\pi(x-a))/(2(b-a))$, $n \in \mathbb{N}$.

**Proof.** The theorem follows from Schwarz's inequality upon integration by parts. $\square$

Theorem 1 is no longer true in the larger subset $F$ of $C^2([a, b])$:

$F = \{ f \in C^2([a, b]) : f'(x) = 0 \text{ for some } x \in [a, b] \}$. \hspace{1cm} (1.4)
However, Gabushin [1967] has shown a weaker result, which we state in the next theorem.

**Theorem 2.** There exists a real number $K$ $(0 < K < \infty)$, such that every $f \in F$ with $f'' \neq 0$ satisfies the inequality

$$
\| f' \|_{[a, b]}^2 \leq K \| f \|_{[a, b]} \| f'' \|_{[a, b]}.
$$

(1.5)

Our goal is to find the infimum of those values of $K$ for which the inequality (1.5) holds for all functions under the restrictions of Theorem 2. We refer to this infimum as the best constant for the inequality (1.5). We assume the existence of an extremal function, i.e., a function $f$ for which equality holds, and use the calculus of variations to determine the value of the best constant and an expression for the extremal function.

We note that the average value of the extremal function over $(a, b)$ is zero. For, let $P$ be the projection operator that maps $L^2(a, b)$ onto the subspace of elements whose average over $(a, b)$ is zero. Then, for any $f \in L^2(a, b)$, $\| Pf \|_{[a, b]} \leq \| f \|_{[a, b]}$, with equality if and only if $f = Pf$. On the other hand, corresponding derivatives of $f$ and $Pf$ have equal norms. Now, suppose that $f$ is an extremal function. Then $\| (Pf)' \|_{[a, b]} = \| f' \|_{[a, b]} = K \| f \|_{[a, b]} \| f'' \|_{[a, b]} \geq K \| Pf \|_{[a, b]} \| (Pf)' \|_{[a, b]}$. To be consistent with the definition of $K$ as the best possible constant in the inequality (1.5), we must have equality here, so $f = Pf$, i.e., the average of $f$ over $(a, b)$ must be zero, as claimed.

We refer the reader to the classical treatise of Hardy, Littlewood, and Polya [1934, Chapter VII] for the general background of variational inequalities.

In Section 2, we rephrase the problem of finding the best constant for the inequality (1.5) as a variational problem. In Section 3, we discuss the solution of this variational problem. In Section 4, we calculate the best constant for the inequality (1.5) and give the general form of the extremal function.

### 2. Reformulation of the Problem

In this section we convert the problem of finding the best constant for the inequality (1.5) into a problem of the calculus of variations. Without loss of generality we may take $a = 0$ and $b > 0$. Thus, $b$ is a parameter of the problem. Henceforth, we shall indicate the dependence on this parameter explicitly. That is, we shall use the definition

$$
F_b = \{ f \in C^2([0, b]) : f'(x) = 0 \text{ for some } x \in [0, b] \} \tag{2.1}
$$

and look for the best constant $K_b$ for the inequality

$$
\| f' \|_{[0, b]}^2 \leq K_b \| f \|_{[0, b]} \| f'' \|_{[0, b]}, \tag{2.2}
$$

where $f \in F_b, f'' \neq 0$.

Let $Q(b, \cdot)$ be the following functional on $F_b$:  

104
\[
Q(b, f) = \frac{\| f' \|_{[0,b]} \| f'' \|_{[0,b]}}{\| f \|_{[0,b]} \| f'' \|_{[0,b]}}, \quad f \in F_b. \tag{2.3}
\]

Then
\[
K_b = \sup\{Q(b, f) : f \in F_b, f'' \neq 0\}. \tag{2.4}
\]

Next, we introduce another subset of \(C^2([0, b])\),
\[
F_b' = \{f \in C^2([0, b]) : f'(0) = 0\}. \tag{2.6}
\]

By analogy with (2.4), we define
\[
K_b' = \sup\{Q(b, f) : f \in F_b', f'' \neq 0\}. \tag{2.7}
\]

Clearly, \(F_b' \subset F_b\), so \(K_b' \leq K_b\) for each fixed \(b > 0\).

**Lemma 1.** \(K_b = K_1\) and \(K_b' = K_1\) for all \(b > 0\).

**Proof.** Choose an arbitrary element \(f \in F_1\) for which \(f'' \neq 0\). Define \(g \in F_b\) by the identity \(g(y) = f(y/b), 0 \leq y \leq b\). Then \(Q(b, g) = Q(1, f)\). Hence, \(K_b = K_1\). The identity for \(K_1\) is shown similarly. \(\Box\)

Let \(R(b, \cdot)\) be the following functional on \(F_b\):
\[
R(b, f) = \frac{2 \| f' \|_{[0,b]} \| f'' \|_{[0,b]}}{\| f \|_{[0,b]} \| f'' \|_{[0,b]}}, \quad f \in F_b. \tag{2.8}
\]

Using the arithmetic-geometric mean inequality, one verifies that \(R(b, f) \leq Q(b, f)\) for each \(b > 0\) and \(f \in F_b\). Therefore, if we define
\[
L_b = \sup\{R(b, f) : f \in F_b, f'' \neq 0\}, \tag{2.9}
\]
then \(L_b \leq K_b\). But \(K_b = K_1\), by Lemma 1, so taking the supremum over all \(b > 0\), we find that \(\sup(L_b : b > 0) \leq K_1\). Similarly, if we define
\[
L_b' = \sup\{R(b, f) : f \in F_b', f'' \neq 0\}, \tag{2.10}
\]
then \(\sup(L_b' : b > 0) \leq K_1\). Clearly, \(L_b' \leq L_b\) for each fixed \(b > 0\).

**Lemma 2.** \(K_1 = \sup\{L_b : b > 0\}\) and \(K_1' = \sup\{L_b' : b > 0\}\).

**Proof.** It suffices to prove the inequalities \(K_1 \leq \sup\{L_b : b > 0\}\) and \(K_1' \leq \sup\{L_b' : b > 0\}\). We prove only the first of these; the proof of the second is similar.

We choose an arbitrary element \(f \in F_1\) for which \(f'' \neq 0\) and fix \(b = \| f' \|_{[0,1]} \| f'' \|_{[0,1]}\). Then we define \(g \in F_b\) by the identity \(g(y) = f(y/b), 0 \leq y \leq b\). One verifies that not only \(Q(b, g) = Q(1, f)\), but also \(R(b, g) = Q(1, f)\), so \(Q(b, g) \leq L_b\). Taking the supremum over all \(g \in F_b\), we find that \(K_b \leq L_b\). But \(K_b = K_1\), by Lemma 1, so \(K_1 \leq L_b\). Then it is certainly the case that \(K_1 \leq \sup\{L_b : b > 0\}\). \(\Box\)
This result enables us to prove that $K_1$ and $K_1$ coincides.

**LEMMA 3.** $K_1 = K_1$.

**Proof.** It suffices to prove the inequality $K_1 \leq K_1$.

We choose an arbitrary element $f \in F_b$ for which $f'' \neq 0$. Let

$$c = \inf \{x \in [0, b] : f'(x) = 0\}, \quad d = \sup \{x \in [0, b] : f'(x) = 0\}.$$  \hspace{1cm} (2.11)

By continuity, $f'(c) = 0$ and $f'(d) = 0$. We have

$$\|f'\|_{[0, c]} + \|f'\|_{[c, d]} + \|f'\|_{[d, b]}. \hspace{1cm} (2.12)$$

We estimate each term separately. For example,

$$\|f'\|_{[0, c]} \leq \frac{1}{2} K_1 (\|f\|_{[0, b]} + \|f'\|_{[0, c]}), \hspace{1cm} (2.13)$$

and similarly for the other two terms, with $K_{d-c}$ and $K_{b-d}$ replacing $K_c$. According to Lemma 1, $K_c = K_{d-c} = K_{b-d} = K_1$, so we can add the three inequalities thus obtained. The result is

$$\|f'\|_{[0, b]} \leq \frac{1}{2} K_1 \left(\|f\|_{[0, b]} + \|f'\|_{[0, c]}\right), \hspace{1cm} (2.14)$$

so $R(b, f) \leq K_1$. Taking the supremum over all $f \in F_b$, we obtain the estimate $L_b \leq K_1$. Taking the supremum over all $b > 0$, we find the desired inequality $K_1 \leq K_1$. □

The combined results of Lemmas 1, 2, and 3 give us a handle on the computation of the best constant for the inequality (2.2). Let $M(b, \cdot)$ be the following functional on $F_b$:

$$M(b, f) = \|f\|_{[0, b]} + \|f''\|_{[0, b]}, \quad f \in F_b, \hspace{1cm} (2.15)$$

and let $M(b)$ be the following infimum of $M(b, \cdot)$:

$$M(b) = \inf \{M(b, f) : f \in F_b, \|f'\|_{[0, b]} = 1\}. \hspace{1cm} (2.16)$$

**THEOREM 3.** The best constant for the inequality (2.2) is independent of $b$. Its value is

$$K_1 = 2 \left[ \inf \{M(b) : b > 0\} \right]^{-1}. \hspace{1cm} (2.17)$$

**Proof.** According to Lemma 1, the best constant $K_b$ is constant and equal to $K_1$ for each $b > 0$. According to Lemma 3, $K_1 = K_1$, where $K_1 = \sup \{L_b : b > 0\}$. According to the definition (2.10), $L_b$ is the supremum of the functional $R(b, \cdot)$ over $F_b$. Because $R(b, \cdot)$ is homogeneous on $F_b$, we have

$$\sup \{R(b, f) : f \in F_b\} = 2 \left[ \inf \{M(b, f) : f \in F_b, \|f'\|_{[0, b]} = 1\} \right]^{-1}, \hspace{1cm} (2.18)$$

where $M(b, \cdot)$ is defined by (2.15). Hence,

$$\sup \{L_b : b > 0\} = 2 \left[ \inf \{M(b) : b > 0\} \right]^{-1}, \hspace{1cm} (2.19)$$

where $M(b)$ is defined in (2.16). □
The computation of \( M(b) \) is a constrained minimization problem, which we solve using the method of Lagrange multipliers; cf. Courant and Hilbert [1953, Chapter 4, Section 7]. That is, we minimize the functional \( J \),

\[
J(f) = \| f \|_{0, b}^2 + \| f'' \|_{0, b}^2 - \lambda \| f' \|_{0, b}^2,
\]

over \( F_b \), where \( \lambda \), the Lagrange multiplier, is a real parameter. We discuss the solution of this variational problem in the following section.

3. The Variational Problem

Let \( f \) be an arbitrary element of \( F_b \) for which \( f'' \neq 0 \). The first variation of the functional \( J \) at \( f \) is the bilinear form

\[
J_1[f, g] = (f, g)_{[0, b]} + (f'', g)_{[0, b]} - \lambda (f', g)_{[0, b]}, \quad g \in F_b.
\]

Upon integration by parts, we find

\[
J_1[f, g] = (f''(0) + \lambda f''(b))g(0) + f''(b)g''(0).
\]

A necessary condition for \( f \) to be an extremal for \( J \) in \( F_b \) is that the first variation of \( J \) vanish at \( f \), i.e., \( J_1[f, g] = 0 \) for all \( g \in F_b \); cf. Courant and Hilbert [1957, Chapter 4, Section 3]. Hence, any extremal must be a solution of the boundary value problem

\[
f''(x) + \lambda f''(x) + f = 0, \quad x \in (0, b);
\]

\[
f'(0) = 0, \quad f''(0) = 0; \quad f''(b) = 0, \quad f''(b) + \lambda f'(b) = 0.
\]

The differential equation (3.3a) is the Euler equation for \( J \). The essential boundary condition \( f'(0) = 0 \) is implicit in the assumption that \( f \in F_b \); the remaining conditions in (3.3b) are the natural boundary conditions. The boundary value problem (3.3) is linear in \( f \) and can be solved by the standard techniques.

**Lemma 4.** If \( f_\lambda \in F_b \) is a solution of (3.3) that satisfies the normalization condition \( \| f_\lambda \|_{[0, b]} = 1 \), then

\[
\lambda = \| f_\lambda \|_{[0, b]}^2 + \| f_\lambda'' \|_{[0, b]}^2.
\]

**Proof.** Let \( f_\lambda \in F_b \) be a solution of (3.3). Then

\[
\| f_\lambda \|_{[0, b]}^2 + \| f_\lambda'' \|_{[0, b]}^2 = \lambda \| f_\lambda' \|_{[0, b]},
\]

as one readily verifies by integration by parts. \( \Box \)

Since the expression in the right member of (3.4) is always positive, Lemma 4 implies that it suffices to consider positive values of \( \lambda \).
Nontrivial solutions of (3.3) exist if and only if \( \lambda \) satisfies the solvability condition

\[
D(b, \lambda) = \frac{1}{2} \sqrt{2 + \lambda} (\lambda - 1) (e^{b \sqrt{2 - \lambda}} - e^{-b \sqrt{2 - \lambda}}) + \frac{1}{2f} \sqrt{2 - \lambda} (\lambda + 1) (e^{ib \sqrt{2 + \lambda}} - e^{-ib \sqrt{2 + \lambda}}) = 0.
\]

(3.5)

Apart from a multiplicative constant, the solutions are given by

\[
f_\lambda(x) = \sigma^2 (e^{\sigma b} + e^{-\sigma b}) (e^{\beta x} + e^{-\beta x}) - \rho^2 (e^{\sigma b} + e^{-\sigma b}) (e^{\omega x} + e^{-\omega x}),
\]

(3.6)

where \( \rho = \frac{1}{2} (\sqrt{2 - \lambda} - i \sqrt{2 + \lambda}) \) and \( \sigma = \frac{1}{2} (\sqrt{2 - \lambda} + i \sqrt{2 + \lambda}) \).

4. The Best Constant and the Extremal Function

The equation (3.5) defines a curve in the interior of the first quadrant of the \((b, \lambda)\)-plane. If we define

\[
\lambda_0 = \inf \{ \lambda > 0 : D(b, \lambda) = 0 \text{ for some } b > 0 \},
\]

(4.1)

then the best constant for the inequality (2.2) is \( K_1 = 2/\lambda_0 \).

**Lemma 5.** \( D(b, 0) < 0 \) for all \( b > 0 \).

**Proof.** \( D(b, 0) \) is found from (3.5),

\[
D(b, 0) = -\sqrt{2} (\sinh(b \sqrt{2}) + \sin(b \sqrt{2})).
\]

(4.2)

This expression is negative for all \( b > 0 \). \( \square \)

It follows from Lemma 5 that there exists a \( \lambda^* > 0 \), such that \( D(b, \lambda) < 0 \) for all \((b, \lambda) \in (0, \infty) \times (0, \lambda^*)\). Clearly, the quantity \( \lambda_0 \), defined in (4.1), is the supremum of all such \( \lambda^* \), so

\[
\lambda_0 = \sup \{ \lambda > 0 : D(b, \lambda) < 0 \text{ for all } b > 0 \}.
\]

(4.3)

As we shall see, it suffices to consider the range \( 0 < \lambda < 2 \), where \( D \) is given by the expression

\[
D(b, \lambda) = \sqrt{2 + \lambda} (\lambda - 1) \sinh(b \sqrt{2 - \lambda}) - \sqrt{2 - \lambda} (\lambda + 1) \sin(b \sqrt{2 + \lambda}).
\]

(4.4)

**Lemma 6.** The function \( \lambda \to D(b, \lambda) \) is strictly increasing on \((0, 1)\) for all \( b > 0 \).

**Proof.** The derivative of \( D \) with respect to \( \lambda \) is

\[
\frac{\partial D}{\partial \lambda} = \frac{b}{2} \sqrt{2 + \lambda} (1 - \lambda) \cosh(b \sqrt{2 - \lambda}) + \frac{3}{2} \frac{1 + \lambda}{\sqrt{2 + \lambda}} \sinh(b \sqrt{2 - \lambda})
\]
As long as $0 < \lambda < 1$, we can estimate this expression from below,

$$
\frac{\partial D}{\partial \lambda} \geq \frac{b}{2} \sqrt{4 - \lambda^2} \left( \frac{1}{2 - \lambda} + \frac{3}{2 + \lambda} \right) \frac{1}{2 + \lambda} - \frac{3}{2 - \lambda} \sin(b \sqrt{2 + \lambda}).
$$

The lower bound is positive for all $b > 0$. □

A straightforward computation yields the inequality $D(2, 0.9) > 0$. It must therefore be the case that $\lambda_0 \leq 0.9$.

Next, we consider the dependence of $D$ upon $b$, keeping $\lambda$ fixed in the range $0 < \lambda \leq 0.9$.

**Lemma 7.** Let $\lambda$ be fixed, such that $0 < \lambda \leq 0.9$. The partial derivative $\partial D/\partial b$ vanishes for at most two positive values of $b$. If it has no zeros, then $D(b, \lambda) < 0$ for all $b > 0$. Otherwise, if $b_1$ and $b_2$ are its zeros and $b_1 < b_2$, then $D(b, \lambda) < \max\{0, D(b_2, \lambda)\}$.

**Proof.** We have $D(0, \lambda) = 0$. Furthermore,

$$
\frac{\partial D}{\partial b}(b, \lambda) = \sqrt{4 - \lambda^2} \left( (\lambda - 1) \cosh(b \sqrt{2 + \lambda}) - (\lambda + 1) \cos(b \sqrt{2 + \lambda}) \right).
$$

The first term inside the braces is negative and strictly decreasing as $b$ increases. The second term is oscillatory with an amplitude at most equal to 1.9; it is negative when $(4n - 3)(\pi/2) < b \sqrt{2 + \lambda} < (4n - 1)(\pi/2)$, $n \in \mathbb{N}$, strictly decreasing as $b \sqrt{2 + \lambda}$ increases from $(4n - 3)(\pi/2)$ to $(4n - 2)(\pi/2)$ and strictly increasing as $b \sqrt{2 + \lambda}$ increases from $(4n - 2)(\pi/2)$ to $(4n - 2)(\pi/2)$. It follows that the expression in the right member of (4.7) vanishes for at most two values of $b$ if $\pi/2 < b \sqrt{2 + \lambda} < 3\pi/2$. If $b \sqrt{2 + \lambda} > 3\pi/2$, then

$$
(\lambda - 1) \cosh(b \sqrt{2 + \lambda}) < (\lambda - 1) \cosh\left( \frac{5\pi}{4} \right) < -\frac{1}{10} \cosh\left( \frac{5\pi}{4} \right) = -2.54 \ldots,
$$

so $(\partial D/\partial b)(b, \lambda) < 0$.

These considerations imply that $\partial D/\partial b(b, \lambda)$ is (strictly) negative everywhere, except in the interval between its two zeros (or at its zero, if there is only one). The assertions of the lemma follow. □

We are now ready to prove the final lemma.

**Lemma 8.** There is a unique positive number $b_0$, such that $D(b_0, \lambda_0) = 0$.

**Proof.** For each $\lambda > 0$, we define the set
According to Lemma 7, $B_\lambda$ is an interval if $0 < \lambda \leq 0.9$. The set $B_\lambda$ is non-empty if $\lambda > \lambda_0$. Furthermore, $B_\lambda \subset B_\mu$ if $0 < \lambda \leq \mu \leq 1$. Therefore, the set $B$, defined by

$$B = \bigcap \{B_\lambda : \lambda_0 \leq \lambda \leq 0.9\},$$

is non-empty. Let $b$ be a point in $B$. Then $D(b, \lambda) \geq 0$ for all $\lambda$ satisfying $\lambda_0 < \lambda \leq 0.9$. By continuity, it follows that $D(b, \lambda_0) \geq 0$. Fixing $\lambda$ at $\lambda_0$, we conclude from Lemma 7 that the function $b \to (\partial D/\partial b)(b, \lambda_0)$ has a zero at a point $b_0 \in B$ and that $b \to D(b, \lambda_0)$ has a maximum at $b_0$. Also, $D(b_0, \lambda_0) \geq 0$.

Suppose that $D(b_0, \lambda_0) = \varepsilon$, where $\varepsilon > 0$. Since the function $\lambda \to D(b_0, \lambda)$ is continuous, there exists a $\delta > 0$, such that $D(b_0, \lambda_0 - \delta) > \varepsilon - \varepsilon = 0$. This conclusion contradicts the definition of $\lambda_0$. Therefore, $D(b_0, \lambda_0) = 0$, and $D(b, \lambda_0) < 0$ for all $b \neq b_0$. \(\square\)

A straightforward numerical computation gives

$$\lambda_0 = 0.855580 \ldots, \quad b_0 = 2.451845 \ldots.$$  \hspace{1cm} (4.11)

We summarize our results in the following theorem.

**Theorem 9.** Let $D$ be defined by

$$D(b, \lambda) = \frac{1}{2} \sqrt{2 + \lambda} (\lambda - 1)(e^{b \sqrt{2 - \lambda}} - e^{b \sqrt{2 + \lambda}}) + \frac{1}{2i} \sqrt{2 - \lambda} (\lambda + 1)(e^{i b \sqrt{2 + \lambda}} - e^{-i b \sqrt{2 + \lambda}}).$$  \hspace{1cm} (4.12)

If $\lambda_0$ is defined by

$$\lambda_0 = \sup \{\lambda > 0 : D(b, \lambda) < 0 \text{ for all } b > 0\},$$  \hspace{1cm} (4.13)

then the best constant for the inequality (2.2) is $K_1 = 2/\lambda_0$; its numerical value is $K_1 = 2.33759 \ldots$. If $b_0$ is the (unique) solution of the equation

$$D(b_0, \lambda_0) = 0,$$  \hspace{1cm} (4.14)

and $\rho_0$ is determined by

$$\rho_0 = e^{i \phi_0}, \quad \lambda_0 = -2 \cos \phi_0, \quad \pi/2 < \phi_0 < \pi,$$  \hspace{1cm} (4.15)

then the extremal function for the inequality (2.2) is a constant multiple of

$$f_0(x) = \text{Im}(\rho_0^2 \cosh(\rho_0 x) \cosh(\rho_0 x)).$$  \hspace{1cm} (4.16)

Here, $^-$ denotes complex conjugation.
References


HIGH-ORDER ASYMPTOTIC APPROXIMATIONS TO THE EIGENVALUES
OF STURM-LIOUVILLE PROBLEMS

B. J. Harris
Department of Mathematical Sciences
Northern Illinois University
DeKalb, IL 60115

Abstract

We consider the eigenvalue problem

\[(py')' + qy = \lambda wy \quad \infty < a \leq b < \infty\]

\[y(a) \sin \alpha - p(a)y'(a) \cos \alpha = 0\]

\[y(b) \cos \beta + p(b)y'(b) \sin \beta = 0\]

for \( \alpha, \beta \in [0, \pi] \). We derive estimates of the eigenvalues, \( \lambda_n \), of the form

\[\lambda_n = \phi(n) + O(n^{-k})\]

where the constant \( k \) increases with the smoothness of \( p, q \) and \( w \).

1. Introduction

We consider the problem of estimating eigenvalues, real numbers \( \lambda \), which are such that the linear differential equation

\[(p(x)y'(x))' + (\lambda w(x) - q(x))y(x) = 0 \quad \infty < a \leq x \leq b < \infty\]

has solutions that satisfy the conditions

\[y(a) \sin \alpha - p(a)y'(a) \cos \alpha = 0\] (1.2)

\[y(b) \cos \beta + p(b)y'(b) \sin \beta = 0\] (1.3)

for \( \alpha, \beta \in [0, \pi] \).

We suppose that \( p, w, \) and \( q \) are real-valued functions defined on \( [a, b] \) with

\[p(x) > 0 \quad \text{and} \quad w(x) > 0 \quad \text{for} \quad x \in [a, b].\]

Further conditions relating to the smoothness of \( p, q, w \) will be imposed below. It is well known that these conditions ensure that the eigenvalues of (1.1)-(1.3) are real and may be arrayed as a sequence

\[\lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \lambda_3 \ldots,\]

where \( \lambda_n \to \infty \) as \( n \to \infty \).
The problem we consider here is that of obtaining the asymptotic form of \( \lambda_n \) as \( n \to \infty \). That is to say, we seek a function \( \phi(n) \) such that

\[
\lambda_{n+1} = \phi(n) + O(n^{-k}) \quad \text{as} \quad n \to \infty
\]

(1.5)

for some constant \( n^* \) where \( k \) is as large as possible. Results of this nature have been obtained before, notably by Fix [1967], Fulton [1982], Atkinson [1957], and Hochstadt [1961]. Our results share with those of Atkinson and Hochstadt the feature that the constant \( k \) of (1.5) increases with the differentiability of \( p, q, w \), but have the advantage of yielding an algorithm for the computation of \( \phi(n) \) that is more readily applicable and that may be implemented in one of the computer languages using exact symbolic algebra. We give an example using the computer language REDUCE.

2. Assumptions and Auxiliary Functions

We suppose throughout that for some integer \( N \geq 1 \),

\[
p, w \in C^{N+2}[a,b]
\]

(2.1)

\[
q \in C^N[a,b]
\]

(2.2)

We define the following functions for \( x \in [a,b] \).

\[
a_{-1}(x) := i(pw)^\frac{1}{2}
\]

(2.3)

\[
a_0(x) := \frac{pa_{-1}}{a_{-1}} = \frac{(wp)^\frac{1}{2}}{4wp}
\]

(2.4)

\[
a_1(x) := \frac{-i}{2(pw)^\frac{1}{2}} (p(q-a_0) - a_0^2)
\]

(2.5)

\[
a_{v+1} := \frac{i}{2} (\frac{p}{w})^{\frac{1}{2}} \left[ a_v + \sum_{0 \leq m \leq N \& m+v = \nu} a_m a_n \right] \quad \nu = 1, \ldots, N - 1
\]

(2.6)

\[
R(x, \lambda) := \sum_{v=0}^{N} a_v(x) \lambda^{-v/2}
\]

(2.7)

\[
i \theta(x, \lambda) := \sum_{v=-1}^{N} a_v(x) \lambda^{-v/2}
\]

(2.8)

We show that for \( \lambda \) sufficiently large, \( R \) and \( \theta \) are real valued.
Further define
\[ \theta := \int_a^b p(t)^{-1} \theta(t, \lambda) dt = \sum_{\nu=1}^N - i \left( \int_a^b p^{-1} a_\nu dt \right) \lambda^{-\nu/2} \]  
(2.9)

\[ \rho := \int_a^b p(t)^{-1} R(t, \lambda) dt = \sum_{\nu=0}^{\nu \text{ even}} \left( \int_a^b p^{-1} a_\nu dt \right) \lambda^{-\nu/2} \]  
(2.10)

3. Results

**THEOREM 1.** In the Dirichlet case in which \( \alpha = \frac{\pi}{2} \), \( \beta = 0 \), the eigenvalues of (1.1)-(1.3) satisfy

\[ \theta(\lambda) = m\pi + O(\lambda^{-N/2}) \]

as \( \lambda \to \infty \) for integral \( m \).

**THEOREM 2.** In the case \( \alpha = \frac{\pi}{2} \), \( \beta \in (0,\pi) \), the eigenvalues of (1.1)-(1.3) satisfy

\[ \theta(\lambda) = m\pi - \cot^{-1} \left( \frac{\cot \beta + R(b)}{\theta(b)} \right) + O(\lambda^{-N-1/2}) \]

as \( \lambda \to \infty \) for integral \( m \).

In the cases \( \alpha \neq \frac{\pi}{2} \), we define a sequence \( \{c_n\} \) by the relations

\[ c_1 := \frac{i}{a_{-1}(a)} (\tan \alpha - a_0(a)) \]  
(3.1)

\[ c_{\nu+1} := \frac{-i}{a_{-1}(a)} \left[ a_\nu(a) - i \sum_{1 \leq m, n \leq N-1 \atop m+n=\nu \atop m, n \text{ odd}} a_m(a)c_n \right] \]  
(3.2)

for \( \nu = 2, \ldots, N-1 \) an even integer.

\[ c := \sum_{p=1 \atop p \text{ odd}}^{N-1} c_p \lambda^{-p/2} \]  
(3.3)

**THEOREM 3.** In the case \( \alpha \neq \frac{\pi}{2} \), \( \beta = 0 \), the eigenvalues of (1.1)-(1.3) satisfy

\[ \theta(\lambda) = m\pi + \cot^{-1}(c) + O(\lambda^{-N/2}) \]

as \( \lambda \to \infty \) for integral \( m \).
THEOREM 4. In the case of $\alpha \neq \frac{\pi}{2}$, $\beta \in (0, \pi)$, the eigenvalues of (1.1)-(1.3) satisfy

$$\theta(\lambda) = m\pi + \psi + O(\lambda^{-(N-1)/2})$$

as $\lambda \to \infty$ for integral $m$, where

$$\psi = \tan^{-1}\left[ \frac{\cot \beta + R(b) + c\theta(b)}{c \cot \beta - \theta(b) + cR(b)} \right].$$

Theorems 1-4 all allow the use of reversion; see Erdélyi [1956] and Olver [1974]. In particular, this procedure may be implemented in the computer language REDUCE and, in the circumstances of Theorem 1 with $p(x) = w(x) = 1$, yields the following results for some constant $\tau$:

$q(t) = x, \ [a,b] = [0,1].$

$$\lambda_{m+\tau} = (8m^5 \pi^5 + 58m^3 \pi^3 + (965/6)m\pi)/(8m^4 \pi^4 + 56m^2 \pi^2 + 147) + O(m^{-5}).$$

$q(x) = x, \ [a,b] = [0,2].$

$$\lambda_{m+\tau} = (\sqrt[m]{m^3 \pi^3} + 22m\pi) / (m^2 \pi^2 + 42) + O(m^{-5}).$$

$q(x) = x, \ [a,b] = [1,3].$

$$\lambda_{m+\tau} = (\sqrt[m]{m^3 \pi^3} + 44m\pi) / (m^2 \pi^2 + 84) + O(m^{-5}).$$

$q(x) = x^2, \ [a,b] = [0,1].$

$$\lambda_{m+\tau} = (43200m^9 \pi^9 + 201600m^7 \pi^7 + 395520m^5 \pi^5 + 413200m^3 + 685813/3m\pi) / (43200m^8 \pi^8 + 194400m^6 \pi^6 + 374040m^4 \pi^4 + 393120m^2 \pi^2 + 237931) + O(m^{-5}).$$

$q(x) = x^3, \ [a,b] = [0,2].$

$$\lambda_{m+\tau} = (\frac{1}{2}m^3 \pi^3 + 44m\pi) / (m^2 \pi^2 + 84) + O(m^{-5}).$$

4. An Auxiliary Function

Let $r(\cdot, \lambda) \in C^1[a,b]$ be a function to be chosen later, and let $y = y(x, \lambda)$ be the solution of (1.1) that satisfies
\[ y(a, \lambda) = 1 \quad p(a)y'(a, \lambda) = r(a, \lambda). \quad (4.1) \]

Let
\[ v(x, \lambda) := r(x, \lambda) - p(x)y'(x, \lambda) / y(a, \lambda). \quad (4.2) \]

We note from (4.2) that
\[ v(a, \lambda) = 0. \quad (4.3) \]

We substitute (4.2) into (1.1) and obtain the equation
\[ v' = Q + p^{-1}v^{2} - 2rp^{-1}v, \quad (4.4) \]

where
\[ Q(x, \lambda) := \lambda w - q + r^{-1} + r^{2}/p. \quad (4.5) \]

It follows from (4.4) that, so long as \( v(x, \lambda) \) exists,
\[ v(x, \lambda) = \int_{a}^{x} \exp \left[ -2 \int_{t}^{x} p^{-1}r \, d\xi \right] \left( Q(t) + p(t)^{-1}v(t)^{2} \right) dt. \quad (4.6) \]

Define
\[ V(x, \lambda) := \sup_{a \leq x} \| v(t, \lambda) \| \]
\[ A := \sup_{a \leq x \leq b} \left| \int_{a}^{x} \exp \left[ -2 \int_{t}^{x} p^{-1}r \, d\xi \right] Q(t) dt \right| \]
\[ B := \sup_{a \leq x \leq b} \int_{a}^{x} \left| \exp \left[ -2 \int_{t}^{x} p^{-1}r \, d\xi \right] \right| dt. \]

**Lemma 1.** If \( 4AB < 1 \), then \( V(b, \lambda) < 2A \) if \( \lambda \) is sufficiently large.

**Proof.** We show that
\[ V(x, \lambda) < 2A \quad \text{for all } x \in [a, b] \quad \text{if } \lambda \text{ is sufficiently large}. \quad (4.7) \]

Our approach follows that of Atkinson [1981].

From (4.6)
\[ |v(x, \lambda)| \leq A + \int_a^x \left| \exp \left[ -2 \int_t^x p^{-1} r d\xi \right] \right| |v(t)|^2 dt \]

\[ \leq A + V(x)^2 \int_a^x \left| \exp \left[ -2 \int_t^x p^{-1} r d\xi \right] \right| dt \]

\[ \leq A + V(x, \lambda)^2 B. \tag{4.8} \]

In fact, (4.8) remains valid if we replace \( x \) on the left-hand side by any lesser number. Thus,

\[ V(x, \lambda) \leq A + B V(x, \lambda)^2 \quad \text{for} \quad x \in [a, b]. \tag{4.9} \]

Suppose (4.7) were false. We know that \( V(0) = 0 < 2A \), so let \( x_0 \) denote the least value of \( x \) for which

\[ V(x_0, \lambda) = 2A. \]

Then, from (4.9),

\[ V(x_0, \lambda) = 2A + 4A^2 B = A(1 + 4AB), \]

which gives a contradiction and completes the proof.

It follows from Lemma 1 that for all \( x \in [a, b] \)

\[ |r(x, \lambda) - p(x)y'(x, \lambda) / y(x, \lambda)| < 2A, \tag{4.10} \]

if \( \lambda \) is sufficiently large. Thus,

\[ \frac{y'(x, \lambda)}{y(x, \lambda)} = p^{-1} r + \sigma \tag{4.11} \]

for some \( \sigma \) with

\[ |\sigma(x, \lambda)| \leq 2p(x)^{-1} A \tag{4.12} \]

for \( x \in [a, b] \) if \( \lambda \) is sufficiently large. It follows from (4.11) that

\[ \log(y(x, \lambda) / y(a, \lambda)) = \int_a^x p(t)^{-1} r(t, \lambda) dt + \int_a^x \sigma(t, \lambda) dt. \]

From (4.1) we may write

\[ y(x, \lambda) = (1 + \varepsilon_1(x, \lambda)) \exp \left[ \int_a^x p(t)^{-1} r(t, \lambda) dt \right], \tag{4.13} \]

\[ y'(x, \lambda) = (1 + \varepsilon_1(x, \lambda))(p(x)^{-1} r(x, \lambda) + \sigma(x, \lambda)) \exp \left[ \int_a^x p(t)^{-1} r(t, \lambda) dt \right], \tag{4.14} \]

where
1 + e_1(x, \lambda) := \exp \left[ \int_0^x \sigma(t, \lambda) \, dt \right].

5. The Choice of r

We recall that

\[ Q := \lambda w - q + r' + p^{-1} r^2. \]  \hspace{1cm} (5.1)

We now follow the approach of Harris [1984] and let

\[ r(x, \lambda) := \sum_{v=-1}^N a_v(x) \lambda^{-v/2} \]

so that

\[ r'(x, \lambda) = \sum_{v=-1}^N a'_v(x) \lambda^{-v/2} \] \hspace{1cm} (5.2)

\[ r(x, \lambda)^2 = \sum_{v=-2}^{2N} \lambda^{-v/2} \sum_{m+n=v} a_m(x) a_n(x). \] \hspace{1cm} (5.3)

We substitute (5.2) and (5.3) into (5.1) and obtain

\[ Q = \lambda w - q + a_{-1} \lambda^{1/2} + a_0 + \sum_{v=1}^N a_v \lambda^{-v/2} + p^{-1} \left[ \lambda a_{-1}^2 + 2 \lambda^{1/2} a_{-1} a_0 + 2 a_{-1} a_1 + a_0^2 + \sum_{v=1}^{2N} \lambda^{-v/2} \sum_{m+n=v} a_m a_n \right]. \] \hspace{1cm} (5.4)

We now choose

\[ a_{-1} = i(p w)^{1/2} \quad a_0 = -p a_{-1} / a_{-1} \]

\[ a_1 = \frac{-i}{2(p w)^{1/2}} (p(q-a_0^2) - a_0^2), \]

and (5.4) becomes

\[ Q = \sum_{v=1}^{N-1} \lambda^{-v/2} \left[ a_v + p^{-1} \sum_{-1 \leq m+n \leq v+1} a_m a_n \right] \]

\[ + a_N \lambda^{-N/2} + \sum_{v=N}^{2N} \lambda^{-v/2} \sum_{m+n=v} a_m a_n. \]

The choice
\[
a_{v+1} := \frac{i}{2} \left( \frac{D}{w} \right)^{\chi} \left\{ a_{\nu} + \sum_{0 \leq m \leq N} a_m a_n \right\} \quad \text{for} \quad v = 1, \ldots, N-1
\]

yields

\[
Q(x, \lambda) = a_{\nu} \lambda^{-N/2} + \sum_{\nu=N}^{2N} \lambda^{-\nu/2} \sum_{m+n=\nu} a_m a_n .
\]  \hspace{1cm} (5.5)

It may be verified that with this choice of \( a_{\nu} \), we do indeed have \( r(\cdot, \lambda) \in C^1[a,b] \).

**Lemma 2.** For \( v = -1, \ldots, N \) \( a_{\nu}(x) \) is real if \( v \) is even, and imaginary if \( v \) is odd.

**Proof.** This is the same as [Harris 1985, Lemma 6(i)].

It follows from (5.5) that

\[
| Q(x, \lambda) | < C \lambda^{-N/2} \quad \text{for} \quad x \in [a,b] .
\]  \hspace{1cm} (5.6)

Moreover, from Lemma 2, since \( \lambda > 0 \), the functions \( R \) and \( \Theta \) of (2.7), (2.8) defined by

\[
R(x, \lambda) := \sum_{\nu=0}^{N} a_{\nu}(x) \lambda^{-\nu/2} = \text{Re}(r(x, \lambda))
\]

\[
i\Theta(x, \lambda) := \sum_{\nu=-1}^{N} a_{\nu}(x) \lambda^{-\nu/2} = i \text{Im}(r(x, \lambda)).
\]

are both real valued.

We note too that

\[
\left| \exp \left\{ \int_{t_s}^{x} p(\xi)_1^{-1} r(\xi) d\xi \right\} \right| = \exp \left\{ \int_{t_s}^{x} R(\xi, \lambda) d\xi \right\} < C
\]

so that

\[
A < C \lambda^{-N/2} \quad \text{and} \quad B < C .
\]  \hspace{1cm} (5.7) \hspace{1cm} (5.8)

**6. A Fundamental Set of Solutions**

It follows from (5.7) and (5.8) that with this choice of \( r \) the hypotheses of Lemma 1 are satisfied, and the analysis of Section 4 shows that we have a complex-valued solution, \( y \), of (1.1) which satisfies (4.1) and
\[ y(x, \lambda) = (1 + O(\lambda^{-N/2})) \exp \left( \int_a^x p^{-1} R + ip^{-1} \theta dx \right). \]

We recall that since the coefficients of (1.1) are real, the real and imaginary parts of \( y \) are also solutions of (1.1). We write

\[ y = y_1 + iy_2 \quad (6.1) \]

and from (4.11)

\[ y'_1 + iy'_2 = (p^{-1} R + ip^{-1} \theta + \sigma)(y_1 + iy_2) \quad (6.2) \]

so that

\[ y'_1 = p^{-1} R y_1 - p^{-1} \theta y_2 + O(\| \sigma \| \| y \|), \]

\[ y'_2 = p^{-1} R y_2 + p^{-1} \theta y_1 + O(\| \sigma \| \| y \|). \]

Since \( |y(x, \lambda)| < C \), we have from (6.2) that two solutions of (1.1) satisfy the following for \( \lambda \) sufficiently large:

\[ y_1(x, \lambda) = \exp \left( \int_a^x p^{-1} R dt \right) \cos \left( \int_a^x p^{-1} \theta dt \right) + O(\lambda^{-N/2}) \]

\[ y'_1(x, \lambda) = p(x)^{-1} \exp \left( \int_a^x p^{-1} R dt \right) R(x) \cos \left( \int_a^x p^{-1} \theta dt \right) - \theta(x) \sin \left( \int_a^x p^{-1} \theta dt \right) + O(\lambda^{-N/2}) \]

\[ y_2(x, \lambda) = \exp \left( \int_a^x p^{-1} R dt \right) \sin \left( \int_a^x p^{-1} \theta dt \right) + O(\lambda^{-N/2}) \]

\[ y'_2(x, \lambda) = p(x)^{-1} \exp \left( \int_a^x p^{-1} R dt \right) R(x) \sin \left( \int_a^x p^{-1} \theta dt \right) + \theta(x) \cos \left( \int_a^x p^{-1} \theta dt \right) + O(\lambda^{-N/2}) \quad (6.3) \]

**Lemma 3.** If \( \lambda \) is sufficiently large, then \( y_1 \) and \( y_2 \) are linearly independent.

**Proof.** Consider

\[ py_1 y'_2 - py'_1 y_2 = R y_1 y_2 + \theta y_1^2 - R y_1 y_2 + \theta y_2^2 + O(\| \sigma \| \| y \|) \quad \text{from (6.2)} \]

\[ = \theta(x, \lambda) |y(x, \lambda)|^2 + O(\lambda^{-N/2}) \quad \text{by (6.3)}. \]

The result now follows, since by (1.4) and (2.3) \( \theta(x, \lambda) = \lambda^\gamma \).

If \( \lambda \) is sufficiently large, we thus have a fundamental set of solutions of (1.1) which we use to prove Theorems 1-4.
7. The Boundary Condition at $a$

The boundary condition at $a$ is

$$y(a)\sin \alpha - p(a)y'(a)\cos \alpha = 0.$$  \hspace{1cm} (7.1)

If $\lambda$ is sufficiently large, then all solutions of (1.1) may be written as

$$y(x) = c_1 y_1 + c_2 y_2,$$  \hspace{1cm} (7.2)

where $y_1$ and $y_2$ are the solutions of Section 6. We note from (6.3) that

$$y_1(a) = 1 + O(\lambda^{-N/2})$$  \hspace{1cm} (7.3)

$$p(a)y_1'(a) = R(a) = \sum_{\nu = 0}^{N-1} a_\nu(a)\lambda^{-\nu/2} + O(\lambda^{-N/2})$$  \hspace{1cm} (7.4)

$$y_2(a) = O(\lambda^{-N/2})$$  \hspace{1cm} (7.5)

$$p(a)y_2'(a) = \theta(a) + O(\lambda^{-N/2}) = \sum_{\nu = 1}^{N-1} (-ia_\nu(a)) + O(\lambda^{-N/2}).$$  \hspace{1cm} (7.6)

We substitute (7.3)-(7.6) into (7.1) and observe that $c_1$ and $c_2$ must satisfy

$$c_1 \sin \alpha - c_1 R(a)\cos \alpha - c_2 \theta(a)\cos \alpha = O(\lambda^{-N/2}).$$  \hspace{1cm} (7.7)

We note that $c_1$ and $c_2$ are unique up to a common factor. Consider first the case $\alpha = \frac{\pi}{2}$. We normalize with $c_2 = 1$ and observe that in this case we have

$$c_1 = O(\lambda^{-N/2}), \, \, \, c_2 = 1 + O(\lambda^{-N/2}).$$  \hspace{1cm} (7.8)

For $\alpha \neq \frac{\pi}{2}$ we normalize by setting $c_1 = 1$ and set

$$c_2 := \sum_{\nu = 0}^{N-1} c_{2, \nu} \lambda^{-\nu/2},$$  \hspace{1cm} (7.9)

where the $c_{2, \nu}$ are chosen from below. We substitute (7.9) into (7.7) and, from (2.7) and (2.8), obtain

$$\sin \alpha - \cos \alpha a_0(a) - \cos \alpha \sum_{\nu = 2}^{N-1} a_\nu(a)\lambda^{-\nu/2}$$

$$- \cos \alpha \sum_{\nu = 0}^{N-1} \lambda^{-\nu/2} \sum_{\nu \text{ even}} \sum_{m+n = \nu} (-ia_m(a))c_{2,n} = O(\lambda^{-N/2}).$$

122
Thus,
\[
\sin\alpha - a_0(a)\cos\alpha + ia_{-1}(a)c_{2,1}\cos\alpha + \cos\alpha \sum_{\nu=2}^{N-1} \lambda^{-\nu/2} \left[ a_\nu(a) + ia_{-1}(a)c_{2,\nu+1} + \sum_{1\leq m,n \leq \nu-1}^{\nu-1} a_m(a)c_{2,n} \right] = O(\lambda^{-N/2}) .
\] (7.10)

We choose
\[
c_{2,1} := \frac{1}{a_{-1}(a)} (\tan\alpha - a_0(a))
\] (7.11)
and for \(\nu = 2, \ldots, N-1\) even,
\[
c_{2,\nu+1} = \frac{-i}{a_{-1}(a)} \left[ a_\nu(a) - i \sum_{1\leq m,n \leq \nu-1}^{\nu-1} a_m(a)c_{2,n} \right] .
\] (7.12)

It may be seen from Lemma 2 that the \(c_{2,\nu}\) are real. In the sequel we suppress the subscript 2 and write \(c_p\) for \(c_{2,p}\).

8. The Boundary Condition at \(b\)

We recall that
\[
\theta(\lambda) := \int_a^b p(t)^{-1} \theta dt = -i \sum_{\nu=0}^{N} \lambda^{-\nu/2} \int_a^b p(t)^{-1} a_\nu(t) dt
\]
and
\[
\rho(\lambda) := \int_a^b R dt = \sum_{\nu=0}^{N} \lambda^{-\nu/2} \int_a^b p(t)^{-1} a_\nu(t) dt
\]
so that, from (6.3),
\[
y_1(b,\lambda) = e^p \cos\theta + O(\lambda^{-N/2})
\] (8.1)
\[
y_1'(b,\lambda) = p(b)^{-1} \left[ R(b)\cos\theta - \theta(b)\sin\theta \right] + O(\lambda^{-N/2})
\] (8.2)
\[
y_2(b,\lambda) = e^p \sin\theta + O(\lambda^{-N/2})
\] (8.3)
\[
y_2'(b,\lambda) = p(b)^{-1} \left[ R(b)\sin\theta + \theta(b)\cos\theta \right] + O(\lambda^{-N/2})
\] (8.4)
9. Proof of Theorem 1

We now have $\alpha = \frac{\pi}{2}$, so by (7.8) the solution of (1.1) that satisfies (1.2) also satisfies

$$y(x, \lambda) = y_2(x, \lambda) + O(\lambda^{-N/2}),$$

where $y_2$ satisfies (6.3). It follows from (8.3) that

$$y(b, \lambda) = e^\theta \sin \theta + O(\lambda^{-N/2}),$$

$$y'(b, \lambda) = y(b, \lambda) = p(b)^{-1} e^\theta [R(b)\sin \theta + \theta(b)\cos \theta] + O(\lambda^{-N/2}).$$

Since $\beta = 0$ in (1.3), we require that $\lambda$ is such that

$$y(b, \lambda) = 0$$

or $\sin \theta = O(\lambda^{-N/2})$ from which follows the conclusion of Theorem 1.

10. Proof of Theorem 2

We now consider the case in which $\alpha = \frac{\pi}{2}$ but $\beta \neq 0$. The results (9.1)-(9.3) still apply, and from (1.3), (9.2), (9.3) we require that $\lambda$ is such that

$$\cos \beta \sin \theta + \sin \beta [R(b)\sin \theta + \theta(b)\cos \theta] = O(\lambda^{-N/2}),$$

which we rewrite as

$$[(\cos \beta + R(b))^2 + \theta(b)^2]^{\gamma} \sin(\theta + \gamma) = O(\lambda^{-N/2}),$$

where

$$\gamma = \cot^{-1} \left( \frac{\cos \beta + R(b)}{\theta(b)} \right).$$

Since $\theta(b) = \lambda^2$, we thus require that

$$\sin(\theta + \gamma) = O(\lambda^{-(N-1)/2})$$

from which Theorem 2 follows.

11. Proof of Theorem 3

We consider now the case in which $\alpha \neq \frac{\pi}{2}$ so that by (7.9) the solution $y(x, \lambda)$ of (1.1) that satisfies (1.2) also satisfies

$$y(x, \lambda) = y_1(x, \lambda) + cy_2(x, \lambda)$$

(11.1)
where we have written \( c \) for the \( c_2 \) given by (7.9)-(7.12) and \( y_1 \) and \( y_2 \) satisfy (6.3). Since \( \beta = 0 \), we require from (1.3) that

\[
y(b, \lambda) = 0,
\]

whence, from (6.3) and (11.1),

\[
\cos \theta + c \sin \theta = O(\lambda^{-N/2}). \tag{11.2}
\]

We rewrite (11.2) as

\[
(1 + c^2)^{1/4} \sin(\theta + \gamma) = O(\lambda^{-N/2}),
\]

where \( \gamma = \cot^{-1}(c) \). Theorem 3 now follows.

12. Proof of Theorem 4

Since \( \alpha \neq \frac{\pi}{2} \), the solution \( y(x, \lambda) \) of (1.1) that satisfies (1.2) is again given by (11.1). The condition (1.3) now requires, by (6.3), (8.1)-(8.4), that \( \lambda \) be such that

\[
\cos \beta (\cos \theta + c \sin \theta) + \sin \beta (R(b)\cos \theta - \theta(b)\sin \theta)
\]

\[
+ c(R(b)\sin \theta + \theta(b)\cos \theta)) = O(\lambda^{-N/2}). \tag{12.1}
\]

We rewrite (12.1) as

\[
([\cot \beta + R(b) + c \theta(b)]^2 + [ccot \beta - \theta(b) + cR(b)]^2)^{1/2} \sin(\theta + \gamma) = O(\lambda^{-N/2}), \tag{12.2}
\]

where

\[
\gamma = \tan^{-1} \frac{\cot \beta + R(b) + c \theta(b)}{ccot \beta - \theta(b) + cR(b)}.
\]

Since \( \theta(b) = \lambda^N \), it follows that

\[
\sin(\theta + \gamma) = O(\lambda^{(N-1)^2}).
\]

The proof of Theorem 4 is now complete.

References


LANDAU'S INEQUALITY FOR THE DIFFERENCE OPERATOR

Man Kam Kwong
Mathematics and Computer Science Division
Argonne National Laboratory
9700 South Cass Avenue
Argonne, IL 60439

A. Zettl
Department of Mathematical Sciences
Northern Illinois University
DeKalb, IL 60115

Abstract

The best constants for Landau's inequality with the classical \( p \)-norms are known explicitly only when \( p = 1, 2, \) and \( \infty \). This is true for both the discrete and the continuous versions of the inequality and for both the "whole line" and "half line" cases. In each of the six known cases the best constant for the discrete version is the same as the best constant for the continuous version. Here we show that for many other values of \( p \) the discrete constants are strictly greater than the corresponding continuous ones. In addition, we show that the "three norm version" of the inequality, established by Nirenberg and Gabushin in the continuous case, is also valid in the discrete case.

1. Introduction

The inequality under investigation here is

\[
\| \Delta^k x \|_q \leq C \| x \|_p^a \| \Delta^n x \|_p^\beta .
\] (1.1)

Throughout this paper we assume that \( k \) and \( n \) are integers satisfying \( 1 \leq k < n \), \( 1 \leq q, p, r \leq \infty \). For \( x = \{x_j\}_{j \in M} \), where \( M = \mathbb{Z} = \{-\infty, -2, -1, 0, 1, 2, \ldots\} \) or \( M = \mathbb{Z}^+ = \{0, 1, 2, \ldots\} \), we have

\[
\| x \|_p^p = \sum_{j \in M} |x_j|^p , \quad 1 \leq p < \infty
\]

\[
\| x \|_\infty = \sup_{j \in M} |x_j| , \quad p = \infty .
\]

The operator \( \Delta \) is the classical difference operator

\[
\Delta x = (x_{j+1} - x_j)_{j \in M} , \quad \Delta^n = \Delta(\Delta^{n-1}) , \quad n = 2, 3, \ldots ;
\] (1.2)

\[
\alpha = (n-k-r^{-1}+q^{-1})/(n-r^{-1}+p^{-1}) , \quad \beta = 1-\alpha ,
\]

with the usual convention that \( s^{-1} = 0 \) when \( s = \infty \).
Inequality (1.1) is the discrete analog of the well-known Gabushin [1967] continuous version

\[ \|y^{(k)}\|_q \leq K \|y\|_p^\alpha \|y^{(\alpha)}\|_r^\beta . \] (1.3)

In (1.3) the norm is the classical \( L^p(J) \) norm with \( J = \mathbb{R} = (-\infty, \infty) \) or \( J = \mathbb{R}^+ = (0, \infty) \):

\[ \|y\|_p^p = \int_J |y(t)|^p dt, \quad 1 \leq p < \infty, \]

\[ \|y\|_\infty = \text{ess. sup. } |y(t)|, \quad t \in J. \]

Nirenberg [1955] and Gabushin [1967] have shown that for given \( p, q, r, 1 \leq p, q, r \leq \infty \), and given \( J = \mathbb{R} \) or \( \mathbb{R}^+ \) and \( \alpha, \beta \) given by (1.2) there exists a positive number \( K \) such that (1.3) holds for all \( y \) in \( W^n(p,r;J) = \{ y \in L^p(J); y^{(\alpha-1)} \text{ is locally absolutely continuous on } J \text{ and } y^{(\alpha)} \in L^r(J) \} \) if and only if

\[ nq^{-1} \leq (n-k)p^{-1} + kr^{-1} . \] (1.4)

Clearly if (1.3) holds with some positive constant \( K \) there is a smallest such constant which we denote by \( K = K(n, k, q, p, r; J) \) to emphasize its dependence on these quantities. The names of Landau, Hardy-Littlewood, and Kolmogorov are often attached to the following special cases of (1.3): (i) \( n = 2, k = 1, p = q = r = \infty, J = \mathbb{R}^+ \), (ii) \( n = 2, k = 1, p = q = r = 2, J = \mathbb{R}^+ \), and (iii) general \( n, k, p = q = r = \infty, J = \mathbb{R} \), respectively.

Similarly the best constant in (1.1) is denoted by \( C = C(n, k, q, p, r; M) \) with \( M = \mathbb{Z} \) or \( \mathbb{Z}^+ \).

The major purpose of this paper is twofold:

1. To show that \( C(2, 1, p, p, p; \mathbb{Z}) > K(2, 1, p, p, p; \mathbb{R}) \) for many values of \( p \) in the range \( 3 < p < \infty \).

2. There exists a finite constant \( C \) such that (1.1) is valid if and only if there exists a finite \( K \) in (1.3).

Remark. We say that inequality (1.1) holds or is valid if (and only if) for given integers \( n,k, 1 \leq k < n, \) given \( q, p, r, 1 \leq p, q, r \leq \infty \), and given \( M = \mathbb{Z} \) or \( \mathbb{Z}^+ \) there exists a positive number \( C \) such that (1.1) holds for all \( x \) in \( l^p(M) \) satisfying \( \Delta^nx \in l^r(M) \).

We also raise a number of questions about these inequalities.
2. The Second-Order Case

In this section we assume that \( n = 2, k = 1 \) and \( q = p = r \) and use the notation \( C(p, M) = C(2, 1, p, p, p ; M), K(p, J) = K(2, 1, p, p, p ; J) \). Even in these special cases the best constants are known explicitly only for \( p = 1, 2, \infty \). These results are summarized in Theorem 1.

**Theorem 1.** \( C(2, Z) = 1 = K(2, R) \), \( C(1, Z^+) = \sqrt{5/2} = K(1, R^+) \), \( C(\infty, Z^+) = 2, C(1, Z) = C(\infty, Z) = C(2, Z^+) = \sqrt{2} = K(1, R) = K(\infty, R) = K(2, R^+) \).

**Proof.** Proofs of these results can be found in the references given in Kwong and Zettl [1980].

It is interesting to note that the discrete constants are the same as the corresponding continuous constants in all six cases where they are known explicitly. Thus one might be tempted to conjecture that \( C(p, Z) = K(p, R) \) and \( C(p, Z^+) = K(p, R^+) \) for all \( p, 1 \leq p \leq \infty \). That this is not so, at least for the whole line case, is shown by

**Theorem 2.** For some values of \( p > 3 \) we have \( C(p, Z) > K(p, R) \).

**Proof.** First we establish the following lemma.

**Lemma 1.** For \( 1 \leq p \leq 2 \) we have

\[
C(p, Z) \geq 2^{2p-1}, \tag{2.1}
\]

and for \( 2 \leq p < \infty \) we have

\[
C(p, Z) \geq 2^{1-2/p}. \tag{2.2}
\]

**Proof of Lemma 1.** A sequence \( x = (x_j) \) is said to be \( P \)-periodic if \( P \) is a positive integer such that for all \( j \) in \( Z \)

\[
x_{j+p} = x_j.
\]

For such a sequence \( x \) we define its "periodic \( l^p \) norm" as

\[
\| x \|_{p,P} = \left( \sum_{j=0}^{P-1} |x_j|^p \right)^{1/p}, \quad 1 \leq p < \infty.
\]

Note that if \( x \) is \( P \)-periodic, then so is \( \Delta x \).

By Theorem 8 of Kwong and Zettl [1987] we have

\[
C(p, Z) = \sup \frac{\| \Delta x \|_{p,P}^2}{\| x \|_{p,P} \| \Delta^2 x \|_{p,P}}, \tag{2.3}
\]
where the supremum is taken over all nonzero $P$-periodic sequences $x$ in $l^\infty(Z)$ with $\Delta^2 x \neq 0$ for all $P = 1, 2, 3, \ldots$.

Applying (2.3) to the 4-periodic sequences $\ldots 0 1 0 -1 \ldots$ and $\ldots 1 1 -1 -1 \ldots$ we get (2.1) and (2.2), respectively. (Actually both (2.1) and (2.2) hold for all $p$, $1 \leq p < \infty$, but are interesting only for the ranges of $p$ indicated.)

It was shown in Franco, Kaper, Kwong, and Zettl [1983] that

$$K(p, R) \leq U(p),$$

where

$$U(p) = (q-1)^{2-p/3} \left\{ \prod_{i=1}^{n} \left( \frac{q}{q^{-1} - 1} \right) \right\}^{2(q-1)}$$

for $2 < p < \infty$, $p^{-1} + q^{-1} = 1$, $n = \lceil \log_2 q \rceil^{-1}$ and $\lceil \cdot \rceil$ is the greatest integer function.

From (2.2) and (2.3) (see also [Franco, Kaper, Kwong and Zettl 1983, p. 261]) for $p = 4$ we get

$$K(4, R) \leq U(4) = (15/7)^{3/8} \approx 1.33082962 < 2^{1/2} \approx 1.414214 \leq C(4, Z).$$

Similarly, we obtain

$$K(5, R) \leq U(5) = 4^{47/125} (11/9)^{8/25} (19/61)^{32/125}$$

$$\approx 1.33222966 < 2^{1-25} \approx 1.515717 \leq C(5, Z).$$

$$K(6, R) \leq U(6) = 5^{19/108} (19/11)^{5/18} (59/91)^{25/108}$$

$$\approx 1.39745611 < 2^{1-26} \approx 1.587401 \leq C(6, Z).$$

Remark 1. Numerical evidence strongly suggests that $C(p, Z) > K(p, R)$ for all $p$ satisfying $3 < p < \infty$. In fact this follows from the lower bound (2.2) for $C(p, Z)$ and the upper bound $U(p)$ for $K(p, R)$ for every $p$ in the range $3 < p < \infty$ for which we have made the computation including values of $p$ up to $p = 10^5$. A natural question arises.

Question 1. Is $C(p, Z) > K(p, R)$ for $1 < p < 2$ and for $2 < p < \infty$?

It seems to us that the upper bound $U(p)$ of $K(p, R)$ is "good" when $p > 3$ but not so good when $p < 3$. We expect that before question 1 is answered for $1 < p < 2$ and $2 < p < 3$ a better upper bound than $U(p)$ needs to be found.

The "half line" version of question 1 is as follows:

Question 2. Is $C(p, Z^+) > K(p, R^+)$ for $1 < p < 2$ and $2 < p < \infty$?
In answering question 2 one is hampered by the lack of "good" upper bounds for \( K(p, R^+) \) and "good" lower bounds for \( C(p, Z^+) \).

There seem to be no upper bounds comparable to \( U(p) \) known for \( K(p, R^+) \) and no lower bounds comparable to (2.1), (2.2) known for \( C(p, Z^+) \).

Can the lower bounds (2.1), (2.2) be improved or can we even do better?

**Question 3.** Is \( C(p, Z) = 2^{1-2p} \) for \( 2 < p < \infty \)?

**Question 4.** Is \( C(p, Z) = 2^{2/p-1} \) for \( 1 < p < 2 \)?

### 3. The Discrete Inequality with Three Norms

Here we establish the discrete analog (1.1) of the well-known inequality (1.3). The proof of the continuous result (1.3) does not seem to be extendable to the discrete case. Here we use the continuous result to prove the discrete case.

**Theorem 3.** Let \( k \) and \( n \) be integers with \( 1 \leq k < n \); let \( p, q, r \) satisfy \( 1 \leq p, q, r \leq \infty \); let \( \alpha, \beta \) be given by (1.2); let \( M = Z \) or \( M = Z^+ \). Then there exists a positive number \( C \) such that

\[
\| \Delta^k x \|_q \leq C \| x \|_p^\alpha \| \Delta^nx \|_r^\beta
\]  \hspace{1cm} (1.1)

for all \( x \) in \( l^p(M) \) satisfying \( \Delta^n x \in l^q(M) \) if and only if

\[
nq^{-1} \leq (n-k)p^{-1} + kr^{-1} \]  \hspace{1cm} (1.4)

**Proof.** Using standard approximation arguments it can be shown that

\[
C(n, k, q, r; Z, Z^+) \geq -K(n, k, q, r; R, R^+) \]  \hspace{1cm} (3.1)

This is done in Ditzian [1983]; see also Kaper and Spellman [1987] for the case \( q = p = r = \infty \) with \( Z \) and \( R \). Since the proof for the general case is similar, we omit the details. Thus if (1.4) fails, then \( C \geq K = \infty \). By this we mean that (1.3) and consequently (1.1) are not valid. Assume condition (1.4) is satisfied. We will show that (1.3) implies (1.1). For this it is sufficient to prove the case \( n = 2 \) since \( n > 2 \) then follows by induction. (The induction argument is not completely straightforward — see Kwong and Zettl [1980] for details.) We proceed with the “whole line” version of \( n = 2 \), i.e., \( M = Z \). The case \( M = Z^+ \) is similar and hence omitted.

To relate the discrete case to the continuous case, we use a construction due to Ditzian [1983].

Given a sequence \( x = \{x_j\}_{j \in M} \) define a function \( f = Tx \) on \( R \) by
\[ f(t) = \sum_{j \in M} x_j B_{j,3}(t), \quad t \in \mathbb{R}, \tag{3.2} \]

where \( B_{j,3} \) is the B-spline of order 3 with support in \([j, j+3]\). See [de Boor 1978, Chapter IX] for a discussion of B-splines. Then \( f' \) is the piecewise linear interpolant of \( \Delta x \) and \( f'' \) is the piecewise constant interpolant of \( \Delta^2 x \) with constant values of -1 and +1.

\[ f'(t) = \sum_{j \in M} (\Delta x)_{j-1} B_{j,2}(t) \]
\[ f''(t) = \sum_{j \in M} (\Delta^2 x)_{j-2} B_{j,1}(t). \]

Thus if \( \Delta^2 x \in l'(\mathbb{Z}) \) then \( f'' \in L'(\mathbb{R}) \) and

\[ \|f''\|_r = \|\Delta^2 x\|_r. \tag{3.3} \]

Now we show that there is a positive number \( A \), independent of \( x \) and \( f \), such that

\[ \|f\|_p \leq A \|x\|_p. \tag{3.4} \]

Note that each of \( B_{j,3}(t) \) is a translate of \( B_{0,3} \) and has support in an interval of length 3. Let \( M \) be a bound of \( C_{0,3}(t) \). Then for \( 1 \leq p < \infty \)

\[ \|f\|_p \leq \sum_{j=-\infty}^{\infty} x_j B_{j,3}(t) \|x\|_p \leq \sum_{j=-\infty}^{j+1} \int x_j B_{j,3}(t) + x_{j+1} B_{j+1,3}(t) + x_{j+2} B_{j+2,3}(t) \|x\|_p \, dt \]
\[ \leq \sum_{j=-\infty}^{\infty} M \|x_j + x_{j+1} + x_{j+2}\|_p \leq 3 \cdot M^p \cdot 2^{p-1} \sum_{j=-\infty}^{\infty} \|x_j\|_p = 3 \cdot 2^{p-1} M^p \|x\|_p. \]

The proof of the \( p = \infty \) case is similar.

Next we show that there is a positive number \( B \), independent of \( x \) and \( f \), such that

\[ \|f'\|_q \geq B \|\Delta x\|_q. \tag{3.5} \]

Since \( f' \) is the piecewise linear function joining the points \((n, (\Delta x)_n)\) we need only show, when \( 1 \leq q < \infty \), that for some positive number \( D \) we have

\[ \int_{j}^{j+1} \|f'(t)\|_q \, dt \geq D \left[ \| (\Delta x)_j \|_q + \| (\Delta x)_{j+1} \|_q \right]. \tag{3.6} \]

Let \( a = (\Delta x)_j, b = (\Delta x)_{j+1} \).

Case 1. \( a \) and \( b \) have the same sign. Suppose \( a \geq 0, b \geq 0 \). Let \( f_1 \) denote the straight line through the points \((j, a)\) and \((j+\frac{a}{b}, 0)\), and let \( f_2 \) denote the straight line through the points \((j+\frac{a}{b}, 0)\) and \((j+1, b)\). Then
A similar construction establishes the case when \( a \leq 0 \) and \( b \leq 0 \). Hence (3.5) holds with \( B = \frac{1}{2(q+1)} \) in this case.

**Case 2.** \( a \) and \( b \) have opposite sign. Suppose \( a \geq 0 \) and \( b \leq 0 \) and \( a \geq |b| \). Let \( g \) denote the straight line through the points \((j,a)\) and \((j+\frac{1}{q},0)\). Then

\[
\int_{j}^{j+1} |f'|^q \geq \int_{j}^{j+1} |f_1'|^q + \int_{j}^{j+1} |f_2'|^q = \frac{a^q}{2(q+1)} + \frac{b^q}{2(q+1)}.
\]

In the last step we used \( |b| \leq a \). Clearly (3.5) follows from these inequalities.

The other subcases are established similarly, as is the case \( q = \infty \).

Using (3.3), (3.4), and (3.5), we have that for all \( x \) in \( l^p(Z) \) such that \( x \neq 0 \) and \( 0 \leq \Delta^n x \) is in \( l^r(Z) \)

\[
\|\Delta x\|_q \|\Delta^2 x\|_r \leq B^{-1} A \cdot \|f\|_p \|f'\|_q \leq AB^{-1} K.
\]

**Remark.** The proof of Theorem 3 — i.e., inequality (1.1) — yields an upper bound for \( C \) in terms of \( K \):

\[
C(2, 1, q, p, r; Z) \leq B^{-1} A K(2, 1, q, p, r; R).
\]

Similar upper bounds for \( C \) in terms of \( K \) follow for \( n > 2 \) and all \( k, 1 \leq k < n \). However, these upper bounds for \( C \) are rough.

**References**


REMARKS ON THE FIRST EIGENSPACE FOR POLYHARMONIC OPERATORS*

Patrizia Pucci
Dipartimento di Matematica Pura ed Applicata
Università degli Studi di Modena
Italy

and

James Serrin
School of Mathematics
University of Minnesota
Minneapolis, MN 55455

Abstract

This paper considers several questions concerning the first eigenvalue and eigenspace of the polyharmonic operator with Dirichlet boundary conditions.

1. Introduction

In this paper we discuss several questions concerning the first eigenvalue and eigenspace of the polyharmonic operator with Dirichlet boundary conditions. In particular, we consider the problem

\[ (-\Delta)^K u = \lambda u \quad \text{in } B \]
\[ u = Du = \cdots = D^{K-1} u = 0 \quad \text{on } \partial B , \]

where \( B \) is a bounded domain in \( \mathbb{R}^n \) and \( K \) is a positive integer.

Recently Coffman [1987] showed that when \( B \) is a ball, then the first eigenspace of (1) is simple (with positive eigenvalue \( \lambda \)) and that the corresponding eigenfunctions \( u \) are radial and of constant sign in \( B \). Our first result is that when \( K = 2 \) or \( K = 3 \), these eigenfunctions not only are radial and of one sign but also are strictly monotone functions of the radius (measured from the center of \( B \)).

Our second result provides an estimate of the first eigenvalue \( \lambda \) of (1) in terms of the first eigenvalues for the Laplace operator. In this case, we need not restrict the discussion to the radial case nor to \( K \leq 3 \). Specifically we show that

* This paper will appear in the Atti del Seminario Matematico e Fisico dell' Università di Modena.
\[ \lambda \geq \begin{cases} 
(\lambda_1 \mu_2)^m & \text{when } K = 2m \text{ even} \\
(\lambda_1 \mu_2)^{m+1} & \text{when } K = 2m+1 \text{ odd} 
\end{cases} \]

where \( \lambda_1 \) is the first eigenvalue for the Laplace operator in \( B \) with Dirichlet data, and \( \mu_2 \) is the first non-zero eigenvalue for the Laplace operator in \( B \) with Neumann data.

In Sections 2 and 3 we consider the monotonicity of the first eigenfunction for balls, when \( K = 2 \) and when \( K = 3 \). In Section 4 we give the eigenvalue estimates noted earlier.

2. Monotonicity When \( K = 2 \)

One may define the least eigenvalue of (1) by the following variational problem, to minimize the integral

\[ I(u) = \int_B F(D^K u) \, dx, \]

where the admissible functions \( u \) are in the set

\[ \mathcal{S} = \{ v \in H_0^1(B) : \int_B v^2 \, dx = 1 \} \]

and

\[ F(D^K u) = \frac{1}{2} \left( \frac{1}{\lambda} \left( \right)^2, \text{ for } K = 2m \text{ even} \right) \]

\[ \frac{1}{2} \left( \frac{1}{\lambda} \right)^2, \text{ for } K = 2m+1 \text{ odd} \right). \]

Now suppose \( B \) is a ball, which for simplicity we assume has a unit radius and center at the origin. In this case Coffman [1987] has shown that the preceding variational problem has a unique positive solution \( u = \phi(r) \), \( r = |x| \), \( x \in B \), and that \( \phi \) satisfies (1) for some real positive \( \lambda \), the smallest eigenvalue.

Multiplying (1) by \( \phi \) and integrating over \( B \) lead at once to the relation

\[ I(\phi) = \int_B F(D^K \phi) \, dx = \frac{1}{2} \lambda \int_B \phi^2 \, dx = \frac{1}{2} \lambda, \]

so that the minimum of \( I \) on \( \Sigma \) is exactly \( \lambda/2 \).

Now suppose \( K = 2 \). Since \( \phi \) is real analytic on \([0,1]\), it is clear that the set of critical points of \( \phi \) consists of isolated points, say,

\[ t_0, t_1, \ldots, t_l \]

in decreasing order, with \( t_0 = 1 \) and \( t_l = 0 \). Because \( \phi > 0 \) on \([0,1]\), it follows that \( \phi' < 0 \) on \((t_1, 1)\). If \( t_1 = 0 \), then \( \phi \) is strictly decreasing on \([0,1]\) as asserted, and we are done.

Thus assume for contradiction that \( t_1 > 0 \) and correspondingly that \( l \geq 2 \). Or, if \( t_l \) is an inflection point of \( \phi \) from the list of values \( t_i \), we may assume without loss of generality that \( \phi \) is
strictly decreasing on \((t_1, 1)\), strictly increasing on \((t_2, t_1)\), and so forth.

Now introduce the function \(\phi: [0, 1] \rightarrow \mathbb{R}\) defined by

\[
\phi(r) = \begin{cases} 
\phi(r), & r \in [t_1, 1] \\
\phi(r) + 2\phi(t_1), & r \in [t_2, t_1) \\
\phi(r) - 2(\phi(t_2) - \phi(t_1)), & r \in [0, t_2), 
\end{cases}
\]

see Figure 1. If \(t_2 = 0\), there is no need to consider the definition of \(\phi\) in the interval \([0, t_2)\). It is easy to see that \(\phi\) is continuous, with continuous first derivatives and bounded second derivatives. Hence \(\phi\) is in \(H^2_0(B)\).

We shall show that

\[
\phi(r) \leq \phi\phi(r) \text{ for } r \in [0, 1].
\]

This is obvious if \(r \in [t_1, 1]\). In the interval \([t_2, t_1)\) we have

\[
\phi(r) - \phi(r) = 2(\phi(t_1) - \phi(r)) > 0,
\]

since \(\phi\) is strictly increasing there. Finally, in \((0, t_2)\) we have
\[
\phi(r) - \phi'(r) = 2[\phi(t_1) - \phi(t_2)],
\]
and this again is strictly positive. This proves (2), moreover with the equality sign holding only in the first interval \([t_1, 1]\). It now follows that
\[
a^2 = \int_B |\phi'|^2 \, dx > \int_B |\phi|^2 \, dx = 1.
\]
Consider the function \(\bar{\phi} = \phi/a\). Then \(\bar{\phi} \in H_0^2(B)\) and
\[
\int_B |\phi|^2 \, dx = 1.
\]
In other words \(\bar{\phi} \in \Sigma\). Moreover, \(|\Delta \bar{\phi}| = |\Delta \phi|\) a.e. in \(B\), and so
\[
I(\bar{\phi}) = \frac{1}{2} \int_B |\Delta \bar{\phi}|^2 \, dx = \frac{1}{2a^2} \int_B |\Delta \phi|^2 \, dx = \frac{1}{2a^2} \int_B |\Delta \phi|^2 \, dx = \frac{\lambda}{2a^2} < \frac{\lambda}{2},
\]
which is impossible since \(\lambda/2\) is the minimum value for \(I(\mu)\) on \(\Sigma\).

3. Monotonicity When \(K = 3\)

Next consider the case \(K = 3\). We shall again show that \(\phi\) is strictly monotone decreasing on \([0, 1]\). The proof will be a similar, but more subtle, version of the previous "reflection" argument.

The relevant idea here is to consider the zeros of the function \(A(r)\), rather than of \(\phi'(r)\), where
\[
A(r) = \begin{cases} 
0 & \text{for } r = 0, \\
\phi'(r) - \frac{\phi'(r)}{r} & \text{for } 0 < r < 1
\end{cases}
\]
(3)
Since \(A\) is real analytic in \([0, 1]\), it is evident that the zeros of \(A\) are isolated unless \(A \equiv 0\). In the latter case we have \(\phi(r) = \alpha + \beta r^2\), which is impossible because \(\phi \neq 0\) and \(\phi(1) = \phi'(1) = 0\). Therefore, the zeros of \(A\) can be written as before
\[
t_0, t_1, \ldots, t_l,
\]
in decreasing order with \(t_0 = 1, t_l = 0\).

Since \(\phi > 0\) and since \(\phi', \phi''\) have only a finite number of zeros, it is easy to see that \(A > 0\) in some interval \((t_1, 1)\) near \(r = 1\). If \(t_1 = 0\), then since \(A(r) = r(\phi'/r)'\), we find that \(\phi'/r\) is strictly increasing on \((0, 1)\) and hence, since \(\phi'(1) = 0\), that \(\phi'/r < 0\) on \((0, 1)\). In other words, \(\phi\) is strictly decreasing in \([0, 1]\), as required.
Thus assume for contradiction that $t_1 > 0$ and $l \geq 2$. Then as in Section 2 we may assume without loss of generality that $A \geq 0$ in $[t_1, 1)$, $A \leq 0$ in $[t_2, t_1)$, and so forth. Of course, in each of these intervals the zeros of $A$, if any, are isolated.

Now introduce the function $\phi: [0, 1] \to \mathbb{R}$ defined by

$$
\phi(r) = \begin{cases} 
\phi(r), & r \in [t_1, 1] \\
-\phi(r) + 2\psi_1(r), & r \in [t_2, t_1) \\
\phi(r) - 2[\psi_2(r) - \psi_1(r)], & r \in [0, t_2),
\end{cases}
$$

where

$$
\psi_j(r) = \phi(t_j) + \frac{1}{2} \phi''(t_j)(r^2 - t_j^2), \quad j = 1, 2;
$$

see Figure 2. if $t_2 = 0$, then $\phi$ is of course defined only by the first two lines of (4). It is not difficult to see that $\phi$ and $\phi''$ are continuous at $t_1$ and $t_2$. The condition that $\phi''$ be continuous at these points is just that

$$
t_k \phi''(t_k) = \phi'(t_k), \quad k = 1, 2,
$$

and this is satisfied since $A(t_k) = 0$. Thus $\phi$ is of class $C^2$ on $[0, 1]$ with $\phi'(0) = 0$, and its third derivatives are bounded. Hence $\phi \in H^3_0(B)$. 

Figure 2
We assert that
\[ \phi(r) \leq \phi(r) \text{ for } r \in [0, 1], \] (6)
with the equality sign holding only in the first interval \([t_1, 1]\). To this end, we first show that
\[ \phi - \psi_1 < 0 \text{ in } [t_2, t_1]. \] (7)
The functions \(\phi\) and \(\psi_1\) have the same value and the same first and second derivatives at \(t_1\). Moreover, \(\psi_1''(r) \equiv 0\), while (3) gives
\[ \phi''(t_1) = A''(t_1) + A(t_1)/t_1. \]
Clearly \(A(t_1) = 0\) and \(A'(t_1) \geq 0\). If \(A'(t_1) > 0\), then \(\phi''(t_1) > 0\) and in turn
\[ \phi - \psi_1 < 0 \] (8)
for \(r \in [t_2, t_1]\) sufficiently near \(t_1\). If \(A'(t_1) = 0\), a similar argument applies at the first non-zero higher derivative of \(A\) at \(t_1\), and again (8) holds for \(r\) near \(t_1\).

Now suppose (8) fails at some point in \([t_2, t_1]\). Then there exists a point \(p \in [t_2, t_1]\) such that \(\phi(p) = \psi_1(p)\), and by Rolle's theorem another point \(q \in (p, t_1)\) for which \(\phi'(q) = \psi_1'(q)\), that is,
\[ \phi'(q) = \phi'(t_1)q. \] (9)
Since \(A \leq 0\) in \([t_2, t_1]\) and since \(A(r) = r(\phi''r)'\), it is evident that
\[ \frac{\phi'(r)}{r} - \frac{\phi'(t_1)}{t_1} > 0 \text{ for } r \in [t_2, t_1]; \] (10)
ote that the equality does not hold. Taking \(r = q\) in this inequality and using (9) and (5), we get
\[ \phi''(t_1) - \phi''(t_1) > 0, \]
which is impossible. Thus (7) is proved.

We next show that
\[ \psi_2 - \psi_1 < 0 \text{ in } [0, t_2]. \] (11)
To begin with,
\[ (\psi_2 - \psi_1)(t_2) = (\phi - \psi_1)(t_2) < 0, \]
where we used (7) at the last step. Moreover, for \(r \in (0, t_2)\) we have
\[ (\psi_2 - \psi_1)'(r) = r[\phi''(t_2) - \phi''(t_1)] = r \left[ \frac{\phi'(t_2)}{t_2} - \frac{\phi'(t_1)}{t_1} \right] > 0 \]
by (10), and (11) follows at once.
We are now ready to prove (6). This is obvious if \( r \in [t_1, 1] \). We show that it holds in \([t_2, t_1]\). Indeed, by (4) we have for \( r \in [t_2, t_1]\)
\[
\tilde{\phi} = -\phi + 2\psi_1 = \phi - 2(\phi - \psi_1) > \phi,
\]
the last step taking (7) into account.

If \( r \in [0, t_2) \), then again by (4),
\[
\tilde{\phi} = \phi - 2(\psi_2 - \psi_1) > \phi
\]
in view of (11). This completes the proof of (6).

Since each of the functions \( \psi_1, \psi_2 \) in (4) is of the form \( \alpha + \beta r^2 \), with \( \alpha \) and \( \beta \) constant, it is clear that \( D\Delta \psi_1 = D\Delta \psi_2 = 0 \). Consequently, we have \( |D\Delta \tilde{\phi}| = |D\Delta \phi| \) almost everywhere in \( B \). Moreover, by (6)
\[
\int_B |\tilde{\phi}|^2 \, dx > \int_B |\phi|^2 \, dx,
\]
and accordingly we obtain a contradiction exactly as in the case \( K = 2 \).

We have actually shown in the case \( K = 3 \) that \( \phi/r \) is strictly increasing, a slightly stronger conclusion than simply that \( \phi \) is strictly monotone.

REMARKS. The preceding proofs can also be used to show that a radially symmetric nontrivial solution of (1) corresponding to the least eigenvalue \( \lambda \) is of one sign, independent of Coffman's argument.

Our demonstration showed \( \tilde{\phi} \geq \phi \) and used the fact that \( \phi > 0 \) to complete the proof. If we drop the condition \( \phi > 0 \) but instead use a slightly more sophisticated function \( \tilde{\phi} \), then one can derive by the same proof that \( \tilde{\phi} \geq |\phi|, |D\Delta \tilde{\phi}| = |D\Delta \phi| \) and \( |\Delta \tilde{\phi}| = |\Delta \phi| \) when \( K = 2 \) and even more \( |D\Delta \tilde{\phi}| = |D\Delta \phi| \) when \( K = 3 \). The required definitions for \( \tilde{\phi} \) are
\[
\tilde{\phi}(r) = \begin{cases} 
\phi(r), & r \in [t_1, 1] \\
-\phi(r) + \phi(t_{k+1}) + \phi(t_k), & r \in [t_{k+1}, t_k), \ k \ odd \\
\phi(r) - \phi(t_{k+1}) + \phi(t_k), & r \in [t_{k+1}, t_k), \ k \ even
\end{cases}
\]
when \( K = 2 \) and \( 1 \leq k \leq l-1 \), and
\[
\tilde{\phi}(r) = \begin{cases} 
\phi(r), & r \in [t_1, 1] \\
(-1)^k \phi(r) + 2 \sum_{j=1}^{k} (-1)^{j+1} \psi_j(r), & r \in [t_{k+1}, t_k)
\end{cases}
\]
when \( K = 3 \) and \( 1 \leq k \leq l-1 \), where
\[ \psi_j(r) = \Phi(t_j) + \frac{1}{2} \Phi''(t_j)(r^2 - t_j^2). \]

The case \( K = 4 \) cannot be treated by a similar reflection method, since the construction of the corresponding functions \( \psi \) and \( \Phi \), such that

\[ \Delta^2 \psi = 0, \quad \Phi \in H_0^4(B), \]

require more conditions than can be simultaneously handled.

4. Eigenvalue Estimates

In this section we estimate the first eigenvalue of problem (1) in terms of the first non-zero eigenvalues of the Laplace operator \(-\Delta\) in \( B \). Here we assume that \( B \) is any regular bounded domain of \( \mathbb{R}^n \).

By the Poincaré inequality, if \( u \in H_0^1(B) \), then

\[ \int_B |D\omega|^2 \, dx \geq \lambda_1 \int_B u^2 \, dx, \tag{12} \]

where \( \lambda_1 \) is the first eigenvalue of \(-\Delta\) in \( B \) with homogeneous Dirichlet boundary conditions.

Moreover, if \( u \in H^2(B) \) with \( \partial u / \partial \nu = 0 \) on \( \partial B \), then

\[ \int_B |\Delta u|^2 \, dx \geq \mu_2 \int_B |Du|^2 \, dx, \tag{13} \]

where \( \mu_2 \) is the first non-zero eigenvalue of the Laplace operator \(-\Delta\) with homogeneous Neumann boundary conditions. This result is due to Smoller [1983: Theorem 11.11]. For completeness, and in the interest of a useful later generalization of the result, we give the proof.

Denote by \( \{ \phi_l \}, l = 1, 2, \ldots \), the eigenfunctions of \(-\Delta\) in \( B \) with homogeneous Neumann boundary conditions, and by \( \{ \mu_l \} \) their corresponding eigenvalues, so that

\[ 0 = \mu_1 < \mu_2 \leq \cdots \leq \mu_l \leq \cdots, \quad \phi_1 = 1/\sqrt{|B|}, \]

\[- \Delta \phi_l = \mu_l \phi_l \text{ in } B, \quad \partial \phi_l / \partial \nu = 0 \text{ on } \partial B.\]

Consider any function \( u \in C^2(\bar{B}) \) with \( \partial u / \partial \nu = 0 \) on \( \partial B \). Certainly \( \Delta u \) is an element of \( L_2(B) \), so that we can consider its Fourier expansion with respect to the orthonormal bases \( \{ \phi_l \} \), namely,

\[ \Delta u = \sum_{l=1}^{\infty} \alpha_l \phi_l, \quad \alpha_l = \int_B \phi_l \, \Delta u \, dx, \quad l = 1, 2, \ldots. \tag{14} \]

By the divergence theorem
\[ \alpha_1 = \int_B \phi_1 \Delta u \, dx = \phi_1 \int_B \Delta u \, dx = \frac{1}{\sqrt{1|B|}} \int_{\partial B} \frac{\partial u}{\partial \nu} \, d\sigma = 0 , \]
so that the first term in (14) can be dropped.

Since \( \partial \phi_i / \partial \nu = \partial u / \partial \nu = 0 \) on \( \partial B \), we also have
\[ \alpha_l = \int_B \phi_l \Delta u \, dx = \int_B u \Delta \phi_l \, dx = -\mu_l \int_B u \phi_l \, dx , \quad l = 2,3, \ldots . \]

Now put
\[ \bar{u} = \frac{1}{\sqrt{1|B|}} \int_B u \, dx = \int_B u \phi_1 \, dx . \]

Then \( u \) can be represented in \( L_2(B) \) by
\[ u = \bar{u} \phi_1 - \sum_{l=2}^{\infty} \alpha_l \mu_l^{-1} \phi_l . \quad (15) \]

Since \( \partial u / \partial \nu = 0 \) on \( \partial B \), we get
\[ \int_B |Du|^2 \, dx = -\int_B \Delta u \, dx = \sum_{l=2}^{\infty} \alpha_l^2 \mu_l^{-1} \leq \mu_2^{-1} \sum_{l=2}^{\infty} \alpha_l^2 = \mu_2^{-1} \int_B |\Delta u|^2 \, dx \]
for any function \( u \in C^2(B) \) with \( \partial u / \partial \nu = 0 \) on \( \partial B \). By standard completion arguments (13) now follows at once.

For elements \( u \in H_0^2(B) \) both inequalities (12) and (13) apply, yielding
\[ \int_B |\Delta u|^2 \, dx \geq \lambda_1 \mu_2 \int_B u^2 \, dx . \quad (17) \]

Let \( u \) be a solution of (1). If \( K = 2m \) is even, we immediately derive from the boundary conditions on \( u \) and from (17) that
\[ \lambda \int_B u^2 \, dx = \int_B u \Delta^K \, u \, dx = \int_B |\Delta^m u|^2 \, dx \geq \cdots \geq (\lambda_1 \mu_2)^m \int_B u^2 \, dx , \]
which implies
\[ \lambda \geq (\lambda_1 \mu_2)^m , \]
provided \( u \) is a non-zero solution of (1).

Analogously, if \( K = 2m + 1 \) is odd, then we have
\[ \lambda \int_B u^2 \, dx = \int_B u \Delta^K \, u \, dx = \int_B |D \Delta^m u|^2 \, dx , \]
whence from (12) and (17) applied as necessary we obtain
\[
\lambda \int_B u^2 \, dx \geq \lambda_1 \int_B |\Delta u|^2 \, dx \geq \cdots \geq \lambda_1 (\lambda_1 \mu_2)^m \int_B u^2 \, dx.
\]

Again this implies that
\[
\lambda \geq \lambda_1 (\lambda_1 \mu_2)^m,
\]
if \(u\) is a non-zero solution of (1).

In conclusion, we have shown that the first eigenvalue \(\lambda\) of problem (1) is such that
\[
\lambda \geq \begin{cases} 
\lambda_1 \mu_2)^m & \text{if } K = 2m \text{ even} \\
\lambda_1 (\lambda_1 \mu_2)^m & \text{if } K = 2m+1 \text{ odd},
\end{cases}
\]
where \(\lambda_1\) is the first eigenvalue of the Laplace operator \(-\Delta\) in \(B\) with homogeneous Dirichlet data and \(\mu_2\) is the first non-zero eigenvalue of \(-\Delta\) in \(B\) with homogeneous Neumann boundary conditions.

In passing, we observe that when \(n = 1\) and \(B = (-1, 1)\), then \(\lambda_1 = \mu_2 = \pi^2/4\). Polya [1952] has proved, however, that \(\mu_2 < \lambda_1\) when the dimension \(n \geq 2\).

If the solution \(u\) of (1) is radial, as is the case when \(B\) is a ball, the estimate (18) can be significantly improved. Indeed, in the representations (14) and (15) it is not necessary to use any non-radial eigenfunctions, for these will be orthogonal to any radial function. But then (16), and hence (13), (17) and (18), hold with \(\mu_2\) replaced by the first non-zero eigenvalue associated with a radial eigenfunction, say \(\mu = \mu_{\text{rad}}\). In particular, using the theory of Bessel functions, we see that the corresponding eigenfunction must have the form, up to a constant,
\[
\phi_{\text{rad}} = r^{-(n-2)/2} J_{(n-2)/2}(\sqrt{\mu}r).
\]
Hence, when \(B\) has radius 1,
\[
\mu_{\text{rad}} = (J_{(n/2)})^2.
\]
Of course, \(\lambda_1\) is always associated with a radial eigenfunction and, in turn, \(\lambda_1 = (J_{(n-2)/2})^2\). In contrast to Polya's result, clearly \(\mu_{\text{rad}} > \lambda_1\).

It is interesting to note also that Levine and Weinberger [1986] have shown that if \(B\) is a bounded convex domain of \(\mathbb{R}^n\), then \(\mu_{l+n} \leq \lambda_l, l = 1, 2, \ldots\). Hence, in the case of a ball we immediately derive that \(\mu_{\text{rad}} \geq \mu_{2+n}\).

Note added in proof. The authors have recently seen a paper of P. L. Lions: "The concentration-compactness principle in the calculus of variations. The limit case - Part 1," Revista Matematica Iberoamericana, 1 (1985), pp. 145-201, in which a stronger result (see Sect. 1.4) is obtained than our first theorem. Our method of proof is entirely different, however, and may have independent interest.
References


Internal:

J. M. Beumer (30)
F. Y. Fradin
H. G. Kaper (30)
M. K. Kwong
G. K. Leaf
G. W. Pieper
D. P. Weber
C. L. Wilkinson

ANL Patent Department
ANL Contract File
ANL Libraries
TIS Files (3)

External:

DOE-OSTI, for distribution per UC-405 (66)
Manager, Chicago Operations Office, DOE
Mathematics and Computer Science Division Review Committee:
J. L. Bona, Pennsylvania State U.
T. L. Brown, U. of Illinois, Urbana
P. Concus, Lawrence Berkeley Laboratory
H. B. Keller, California Inst. of Technology
M. J. O'Donnell, U. of Chicago
L. R. Scott, Pennsylvania State U.
C. D. Ahlbrandt, U. of Missouri, Columbia
W. Allegretto, U. of Alberta, Canada
F. V. Atkinson, U. of Toronto, Canada
D. Austin, ER-DOE, Washington
P. Bailey, Sandia National Laboratories
Chr. Bennewitz, U. of Uppsala, Sweden
H. Benzinger, U. of Illinois, Urbana
R. C. Brown, U. of Alabama, Tuscaloosa
A. Castro, North Texas State U.
C. Y. Chan, U. of Southwestern Louisiana
S. Chen, Shandong U., People's Republic of China
L. Erbe, U. of Alberta, Canada
W. N. Everitt, U. of Birmingham, U.K.
C. T. Fulton, Florida Inst. of Technology
L. Ganta, U. of Alberta, Canada
M. Garbey, Ecole Normale Superieure de Lyon, France
D. W. Gillman, Massachusetts Institute of Technology
J. Goldstein, Tulane U.
G. Halvorsen, Inst. for Energy Technology, Norway
B. J. Harris, Northern Illinois U.
D. Hinton, U. of Tennessee, Knoxville
V. Jurdjevic, U. of Toronto, Canada
A. M. Krall, Pennsylvania State U.
G. Michael, Lawrence Livermore Laboratory
A. B. Mingarelli, U. of Ottawa, Canada
J. Neuberger, North Texas State U.
S. Pruess, Colorado School of Mines
P. Pucci, U. degli Studi di Modena, Italy
T. Read, Western Washington U.
J. Ridenhour, Utah State U.
B. Schultze, U. Gesamthochschule Essen, West Germany
G. Serrin, U. of Minnesota, Minneapolis
J. K. Shaw, Virginia Polytechnic Inst. and State U.
E. Socolovsky, U. of Pittsburgh, Pittsburgh
R. C. Thompson, Utah State U.
L. Veron, Université Tours, France
A. Zettl, Northern Illinois U., c/o Universitat Gesamthochschule Essen, West Germany