Advances in Computational Mathematics 2(1994)155–184

Numerical solution of eigenvalue problems for Hamiltonian systems

Marco Marletta

Department of Mathematics and Computer Science, University of Leicester, University Road, Leicester, LE1 7RH, UK

Received 27 October 1992; revised 15 June 1993

This paper discusses the numerical solution of eigenvalue problems for Hamiltonian systems of ordinary differential equations. Two new codes are presented which incorporate the algorithms described here; to the best of the author's knowledge, these are the first codes capable of solving numerically such general eigenvalue problems. One of these implements a new new method of solving a differential equation whose solution is a unitary matrix. Both codes are fully documented and are written in PFORT-verified FORTRAN 77, and will be available in netlib/aicm/sl11f and netlib/aicm/sl12f.

Keywords: Hamiltonian systems, eigenvalue problems, numerical algorithms, mathematical software.

AMS(MOS) subject classification: 65L15.

1. Introduction

A linear *Hamiltonian system* is a differential equation of the form

$$\begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} \begin{pmatrix} u'(x) \\ v'(x) \end{pmatrix} = \begin{pmatrix} S_{1,1} & S_{1,2} \\ S_{1,2}^{\mathrm{T}} & S_{2,2} \end{pmatrix} \begin{pmatrix} u(x) \\ v(x) \end{pmatrix}, \quad x \in [a,b]. \tag{1}$$

Here, $S_{1,1}$ and $S_{2,2}$ are $n \times n$ symmetric real matrices, $S_{1,2}$ is an $n \times n$ real matrix, and $u, v : \mathbb{R} \to \mathbb{R}^n$ are vector-valued C^1-functions. The matrices $S_{1,1}$, $S_{1,2}$ and $S_{2,2}$ are also functions of the independent variable x; we will assume that they are at least piecewise continuous.

In this paper, we will be concerned with numerical solution of eigenvalue problems for Hamiltonian systems. Such problems arise from many sources. For example, a well-known separation of variables method applied to quantum-mechanical many-body problems often gives rise to Hamiltonian systems in some form (see, e.g. Alexander and Manolopoulos [1]). Simple scalar Sturm–Liouville problems with non-separated boundary conditions also give rise to Hamiltonian eigenvalue

problems with $n = 2$ (see Greenberg [5]). A further example is the matrix–vector Sturm–Liouville eigenvalue problem whose equation is

$$-(P(x)y')' + Q(x)y = \lambda W(x)y,$$

where y is an n-vector and P, Q and W are symmetric matrices with P and W positive definite; in this case, the system (1) has $S_{1,1} = \lambda W(x) - Q(x)$, $S_{1,2} = 0$ and $S_{2,2} = P(x)^{-1}$. This problem was studied by the author in [7] and an algorithm proposed, based on approximation of the coefficients P, Q and W by piecewise constant matrices and the use of a spectral (miss distance) function generalised from the well-known Prüfer miss-distance for scalar Sturm–Liouville problems. The results in [7] have since been generalised by Greenberg, who replaced the matrix Sturm–Liouville differential operator by a matrix $2m$th order linear elliptic self-adjoint differential operator and showed that the spectral function proposed in [7] still works.

Under fairly mild smoothness restrictions on the coefficients, the Hamiltonian system (1) can be reduced to a second-order matrix–vector differential equation

$$-y'' + Q(x, \lambda)y = 0,$$

provided the matrix $S_{2,2}$ is positive definite. This generalisation of the Liouville transformation is due to Reid [12], and is the basis of one of the algorithms which we present.

We consider equation (1) in which the matrix coefficients have a possibly nonlinear dependence on a real eigenparameter λ; thus

$$S(x, \lambda) = \begin{pmatrix} S_{1,1}(x, \lambda) & S_{1,2}(x, \lambda) \\ S_{1,2}^{\mathrm{T}}(x, \lambda) & S_{2,2}(x, \lambda) \end{pmatrix}.$$

The matrix $S(x, \lambda)$ will be assumed to be at least continuously differentiable as a function of λ for each x. We will treat only separated boundary conditions since non-separated boundary conditions can be separated at the expense of doubling n. Thus we assume boundary conditions of the form

$$A_1^{\mathrm{T}}u(a) + A_2^{\mathrm{T}}v(a) = 0, \tag{2}$$

$$B_1^{\mathrm{T}}u(b) + B_2^{\mathrm{T}}v(b) = 0. \tag{3}$$

Here, A_1, A_2, B_1 and B_2 are $n \times n$ real matrices such that the $n \times 2n$ matrices (A_1, A_2) and (B_1, B_2) are of full rank n and the *conjointness conditions*

$$A_1^{\mathrm{T}}A_2 - A_2^{\mathrm{T}}A_1 = 0, \tag{4}$$

$$B_1^{\mathrm{T}}B_2 - B_2^{\mathrm{T}}B_1 = 0, \tag{5}$$

are satisfied. These conditions are imposed because they arise naturally when dealing with matrix–vector Sturm–Liouville problems as necessary conditions for the problem to be self-adjoint and hence have real eigenvalues; they are also necessary if we are to use matrix oscillation theory, as we do later in this paper. Their name arises from the fact that the columns of a $2n \times n$ solution $(U^T, V^T)^T$ of the Hamiltonian system with $U(a) = -A_2$ and $V(a) = A_1$ give a conjoined basis (in the sense of Reid [13]) for a subspace of the space of solutions of the differential equation.

The differential equation (1) together with the boundary conditions (2), (3) form the *eigenvalue problem*, the object of which is to find those values of λ (the eigenvalues) for which equation (1) has a non-trivial solution satisfying the boundary conditions (2), (3). At this stage, it is not clear that the eigenvalues are necessarily isolated.

The rest of this paper is organised as follows. In section 2, we consider the nature of the spectrum of the problem and some oscillation theory. We impose some restrictions on the coefficient $S(x, \lambda)$ which enable us to construct a miss-distance function $M(\lambda)$, which is a monotone increasing integer valued function whose only points of increase are discontinuities at the eigenvalues, the size of the discontinuity at an eigenvalue being equal to its multiplicity. In section 3 (the main section), we consider numerical algorithms for the computation of $M(\lambda)$. Section 4 gives the results of some numerical experiments using these algorithms.

2. Spectral and oscillation theory

The most well-known eigenvalue problem is the two-point regular Sturm–Liouville problem, for which the eigenvalues form a countably infinite sequence. The set of eigenvalues has the form

$$\lambda_0 < \lambda_1 < \lambda_2 < \dots ,$$

each eigenvalue being simple. For the problems which we consider here, as we have stated them so far, there is almost nothing that can be said about the spectrum. It could even be the whole real line. By introducing monotonicity properties, we will now restrict our attention to problems with a countable spectrum. We start by introducing the Θ matrices of Atkinson [2]. Let $(U_L^T, V_L^T)^T$ denote the $2n \times n$ matrix solution of (1), defined by the initial condition

$$U_L(a) = -A_2, \quad V_L(a) = A_1, \tag{6}$$

and let $(U_R^T, V_R^T)^T$ be the $2n \times n$ matrix solution of (1) subject to the initial condition

$$U_R(b) = -B_2, \quad V_R(b) = B_1. \tag{7}$$

The Atkinson Θ matrices are

$$\Theta_L(x) = (V_L(x) + \mathrm{i}\, U_L(x))\,(V_L(x) - \mathrm{i}\, U_L(x))^{-1}, \tag{8}$$

$$\Theta_R(x) = (V_R(x) + \mathrm{i}\, U_R(x))\,(V_R(x) - \mathrm{i}\, U_R(x))^{-1}. \tag{9}$$

These matrices are unitary. It is not difficult to show that the conditions $\Theta_L^*\Theta_L = I$ and $\Theta_R^*\Theta_R = I$ are, respectively, equivalent to the equations

$$U_L^{\mathrm{T}}(x)V_L(x) - V_L^{\mathrm{T}}(x)U_L(x) = 0, \tag{10}$$

$$U_R^{\mathrm{T}}(x)V_R(x) - V_L^{\mathrm{T}}(x)U_L(x) = 0. \tag{11}$$

To prove the first of these, observe that by the initial condition (6), equation (10) holds when $x = a$. To prove that it holds for other values of x, one simply differentiates the left-hand side of (10) and uses the differential equation (1) to show that the result is zero. The procedure also establishes (11).

Because the Θ matrices are unitary, their eigenvalues all lie on the unit circle. We shall denote the eigenvalues of $\Theta_L(x, \lambda)$ by $(\exp(\mathrm{i}\phi_j^L(x, \lambda)))_{j=1}^n$ and the eigenvalues of $\Theta_R(x, \lambda)$ by $(\exp(\mathrm{i}\phi_j^R(x, \lambda)))_{j=1}^n$, where we have indicated the dependence on x and λ. The so-called *phase angles* $\phi_j^L(x, \lambda)$ and $\phi_j^R(x, \lambda)$ may be chosen to be continuous functions of x and λ and are uniquely determined (up to ordering) by the normalisation conditions

$$0 \le \phi_j^L(a) < 2\pi, \quad 0 < \phi_j^R(b) \le 2\pi, \tag{12}$$

at the boundaries. Note that the boundary values are independent of λ because the matrices A_1, A_2, B_1 and B_2 in (2) and (3) are assumed independent of λ.

The determinant of a matrix is the product of its eigenvalues; hence

$$\det(\Theta_L) = \prod_{j=1}^n \exp(\mathrm{i}\phi_j^L) = \exp\left(\mathrm{i}\sum_{j=1}^n \phi_j^L\right).$$

Let us therefore define

$$\arg\det\Theta_L(x, \lambda) = \sum_{j=1}^n \phi_j^L(x, \lambda), \tag{13}$$

$$\arg\det\Theta_R(x, \lambda) = \sum_{j=1}^n \phi_j^R(x, \lambda). \tag{14}$$

Given any fixed point $c \in [a, b]$, the matrix $\Theta_R^*(c, \lambda)\Theta_L(c, \lambda)$ is unitary. Its eigenvalues we denote by $(\exp(\mathrm{i}\omega_j(c, \lambda)))_{j=1}^n$, only this time we normalise the phase angles by

$$0 \le \omega_j(c, \lambda) < 2\pi, \quad j = 1, \ldots, n.$$

Because of this normalisation, the phase angles are not generally continuous functions of λ. Now define the function $M(\lambda)$ by

$$M(\lambda) = \frac{1}{2\pi}\left\{\arg\det\Theta_L(c, \lambda) - \arg\det\Theta_R(c, \lambda) - \sum_{j=1}^n \omega_j(c, \lambda)\right\}. \tag{15}$$

This function is of fundamental importance because its discontinuities are precisely the eigenvalues of our problem, and because it gives a natural indexing of the eigenvalues. It was first introduced in [7] in the less general context of matrix–vector Sturm–Liouville problems. Before considering its properties (see theorem 1 below), we require some lemmata.

LEMMA 1

Let $c \in [a, b]$ be fixed. Then λ is an eigenvalue of multiplicity d for the problem (1),(2),(3) if and only if the matrix

$$\Delta := U_R^{\mathrm{T}}(c)V_L(c) - V_R^{\mathrm{T}}(c)U_L(c) \tag{16}$$

has rank $n-d$.

Proof

Suppose that the matrix Δ has rank $n-d$, where $d \geq 0$, and let $r \geq 0$ be the multiplicity of λ as an eigenvalue of the Hamiltonian problem; our lemma will be proved if we can show that $r = d$. Let A and B be the matrices

$$A = \begin{pmatrix} -V_R^{\mathrm{T}}(c) & U_R^{\mathrm{T}}(c) \\ -V_L^{\mathrm{T}}(c) & U_L^{\mathrm{T}}(c) \end{pmatrix}, \quad B = \begin{pmatrix} U_L(c) & -U_R(c) \\ V_L(c) & -V_R(c) \end{pmatrix}.$$

Note that

$$AB = \begin{pmatrix} \Delta & 0 \\ 0 & \Delta^{\mathrm{T}} \end{pmatrix}.$$

Because the matrix Δ has rank $n-d$, from this expression the matrix AB is clearly of rank $2n-2d$. The matrices A and B are of the same rank since B may be transformed into A by transpositions and column interchanges. The sum of the rank deficiencies of A and B must be at least the rank deficiency of the matrix AB, which is $2d$; so the matrices A and B must have rank deficiencies at least d. Thus, in particular, there exist d linearly independent vectors

$$z_j = \begin{pmatrix} \gamma_j \\ \beta_j \end{pmatrix}, \quad j = 1, \ldots, d,$$

such that $Bz_j = 0$ for $j = 1, \ldots, d$. Now define functions $u_1, \ldots, u_d$ and $v_1, \ldots, v_d$ by

$$u_j(x) = \begin{cases} U_L(x)\gamma_j & a \leq x \leq c, \\ U_R(x)\beta_j & c < x \leq b; \end{cases}$$

$$v_j(x) = \begin{cases} V_L(x)\gamma_j & a \leq x \leq c, \\ V_R(x)\beta_j & c < x \leq b. \end{cases}$$

Because $Bz_j = 0$ for $j = 1, \ldots, d$, the u_j and v_j are continuous at $x = c$; therefore, the pairs $(u_j^T, v_j^T)^T$ satisfy the differential equation throughout the whole interval $[a, b]$. They are therefore eigenfunctions of the Hamiltonian problem, and we wish to show that they are linearly independent. Suppose that there exist constants $c_1, \ldots, c_d$ such that

$$\sum_{j=1}^{d} c_d u_j(x) \equiv 0, \quad \sum_{j=1}^{d} c_d v_j(x) \equiv 0, \qquad j = 1, \ldots, d.$$

Then, in particular, we must have

$$(U_L^T(x), V_L^T(x))^T \sum_{j=1}^{d} c_j \gamma_j = 0, \quad (U_R^T(x), V_R^T(x))^T \sum_{j=1}^{d} c_j \beta_j = 0.$$

It is known that the matrices $(U_L^T, V_L^T)^T$ and $(U_R^T, V_R^T)^T$ are of full rank. Therefore, we must have

$$\sum_{j=1}^{d} c_j \gamma_j \equiv 0, \quad \sum_{j=1}^{d} c_j \beta_j \equiv 0,$$

in other words,

$$\sum_{j=1}^{d} c_j z_j = 0.$$

Since the z_j are linearly independent, this implies that the c_j are all zero, and so the $(u_j^T, v_j^T)^T$ are linearly independent. Thus, we have constructed d eigenfunctions, so $r \geq d$.

To prove the opposite inequality, suppose that $(u_1^T, v_1^T)^T, \ldots, (u_r^T, v_r^T)^T$ are linearly independent eigenfunctions. Since the eigenfunctions satisfy the boundary conditions at both ends of the interval, there exist linearly independent vectors $\gamma_1, \ldots, \gamma_r$ and $\beta_1, \ldots, \beta_r$ such that

$$u_j(x) = \begin{cases} U_L(x)\gamma_j & a \leq x \leq c, \\ U_R(x)\beta_j & c < x \leq b; \end{cases}$$

$$v_j(x) = \begin{cases} V_L(x)\gamma_j & a \leq x \leq c, \\ V_R(x)\beta_j & c < x \leq b. \end{cases}$$

Continuity of u_j and v_j at $x = c$ then gives

$$U_L(c)\gamma_j = U_R(c)\beta_j,$$

$$V_L(c)\gamma_j = V_R(c)\beta_j.$$

Premultiplying the first equation by $V_R^T(c)$ and the second equation by $U_R^T(c)$ and subtracting the first from the second, we obtain

$$(U_R^T(c)V_L(c) - V_R^T(c)U_L(c))\gamma_j = (U_R^T(c)V_R(c) - V_R^T(c)U_R(c))\beta_j.$$

The right-hand side is zero because of the conjointness condition satisfied by U_R and V_R, and the coefficient on the left-hand side is clearly the matrix Δ. Thus,

$$\Delta\gamma_j = 0 \quad \text{for } j = 1, \ldots, r.$$

The rank deficiency of Δ is therefore at least r, so $d \geq r$. We have already proved $r \geq d$ above, and so $r = d$ as required. □

The lemma above gives a condition for λ to be an eigenvalue, in terms of the matrices U_L, U_R, V_L and V_R. The next lemma gives us a condition which uses the Θ matrices Θ_L and Θ_R.

LEMMA 2

Let $c \in [a, b]$ be fixed. Then λ is an eigenvalue of multiplicity d for the problem (1),(2),(3) if and only if the unitary matrix $\Theta_R^*(c)\Theta_L(c)$ has 1 as an eigenvalue of multiplicity d.

Proof

1 is an eigenvalue of multiplicity d of the matrix $\Theta_R^*(c)\Theta_L(c)$ if and only if there exist precisely d linearly independent vectors $v_1, \ldots, v_d$ such that

$$\Theta_R^*(c)\Theta_L(c)v_j = v_j, \qquad j = 1, \ldots, d.$$

Substituting the definitions of Θ_R and Θ_L yields

$$(V_R^T + iU_R^T)^{-1}(V_R^T - iU_R^T(V_L + iU_L)(V_L - iU_L)^{-1}v_j = v_j, \quad j = 1, \ldots, d. \quad (17)$$

If we now define linearly independent vectors z_j by $z_j = (V_L - iU_L)^{-1}v_j$, then we can rewrite (17) as

$$(V_R^T - iU_R^T)(V_L + iU_L)z_j = (V_R^T + iU_R^T)(V_L - iU_L)z_j, \quad j = 1, \ldots, d.$$

This may be rearranged as

$$2(V_R^T U_L - U_R^T V_L)z_j = 0, \qquad j = 1, \ldots, d.$$

Clearly, this is true if and only if the matrix Δ in (16) has rank $n - d$, and hence if and only if λ is an eigenvalue of multiplicity d. □

We have not yet imposed any conditions to preclude the possibility that every λ is an eigenvalue. The next lemma gives a condition under which any eigenvalue is isolated.

LEMMA 3

If for every non-trivial solution $(u^T, v^T)^T$ of (1) and any $a \leq \alpha < \beta \leq b$ we have

$$\int_\alpha^\beta (u^T, v^T) \frac{\partial S}{\partial \lambda}(t, \lambda) \begin{pmatrix} u \\ v \end{pmatrix} dt > 0, \tag{18}$$

then the phase-angles $\phi_j^L(x, \lambda)$ are strictly increasing functions of λ for each $x > a$, the phase-angles $\phi_j^R(x, \lambda)$ are strictly decreasing functions of λ for each $x < b$, and the phase-angles $\omega_j(c, \lambda)$ are increasing functions of λ except where the corresponding eigenvalues of $\Theta_R^* \Theta_L(c, \lambda)$ pass through 1 on the unit circle: when this happens, they pass through 1 moving in the positive direction so that the corresponding $\omega_j(c, \lambda)$ undergo a discontinuous jump from $2\pi-$ to $0+$.

Proof

The proof is straightforward and follows the method developed by Atkinson [2, p. 307 and appendix V]. □

We come now to the result which makes the function $M(\lambda)$ defined by (15) important.

THEOREM 1

Suppose that the hypothesis of lemma 3 is satisfied. Then the integer-valued function $M(\lambda)$ is an increasing function whose only points of increase are discontinuities at the eigenvalues of the problem (1),(2),(3). As λ increases through an eigenvalue of multiplicity d, $M(\lambda)$ jumps by $+d$.

Proof

It is easy to see that $M(\lambda)$ is integer-valued, because

$$\arg \det \Theta_L - \arg \det \Theta_R \equiv \arg \det \Theta_R^* \Theta_L \pmod{2\pi}.$$

It is also evident that $M(\lambda)$ is an increasing function: this is an immediate consequence of lemma 3. The points of increase of M are precisely those λ values where one (or more) of the ω_j is discontinuous; this happens when one (or more) of the corresponding eigenvalues of the matrix $\Theta_R^* \Theta_L(c, \lambda)$ is unity, which by lemma 1

happens precisely when λ is an eigenvalue. At an eigenvalue of multiplicity d, precisely d of the $\omega_j(c, \lambda)$ are discontinuous with

$$\omega_j(c, \lambda+) - \omega_j(c, \lambda-) = -2\pi.$$

Thus, from (15) it is clear that $M(\lambda)$ has the jump property

$$M(\lambda+) - M(\lambda-) = d$$

when λ is an eigenvalue of multiplicity d. □

The last result still allows for the possibility of a spectrum unbounded below and gives a natural indexing for the eigenvalues of a problem with such a spectrum. Specifically, if we consider matrix–vector Sturm–Liouville problems, then it is known (see [8]) that $M(\lambda) = -n$ for all $\lambda < \lambda_0$, and therefore the kth eigenvalue can be found by using bisection to find the point of discontinuity where the function $M(\lambda) + n - k - 1/2$ changes sign. The same algorithm clearly carries through to the case where the spectrum is unbounded below, provided we allow negative integers k in the indexing of the eigenvalues. Nevertheless, it is useful to have a result such as the following, which guarantees a spectrum which is bounded below.

THEOREM 2

Suppose that the coefficient matrix $S(x, \lambda)$ appearing in (1) is such that

- for every non-trivial solution $(u^T, v^T)^T$ of (1) and any $a \le \alpha < \beta \le b$,

$$\int_\alpha^\beta (u^T, v^T) \frac{\partial S}{\partial \lambda}(t, \lambda) \begin{pmatrix} u \\ v \end{pmatrix} dt > 0; \tag{19}$$

- for any $x \in [a, b)$ and any real λ, the sub-matrix $S_{2,2}(x, \lambda)$ of $S(x, \lambda)$ is positive semi-definite.

Then the function $M(\lambda)$ satisfies $M(\lambda) \ge -n$ for all λ; the spectrum of the problem (1),(2),(3) is bounded below; and there can be no infinite sequence of eigenvalues decreasing towards a finite limit.

Proof

Let λ be fixed. In the case $S_{2,2}(x, \lambda) > 0$, the equation (1) is equivalent to the second-order equation

$$-(Pu')' + [(Ru)' - R^T u'] = -Qu,$$

where $P = S_{2,2}^{-1}$, $R = S_{2,2}^{-1} S_{1,2}^T$ and $Q = S_{1,2} S_{2,2}^{-1} S_{1,2}^T - S_{1,1}$. It follows from Greenberg [5] that $M(\lambda) \ge -n$ for all λ, from which all the claims of the theorem follow.

We now generalize to the case where $S_{2,2}$ is only positive *semi*-definite. We compare the problem with equation (1) with that in which the coefficient $S_{2,2}(x, \lambda)$ is replaced by $S_{2,2}(x, \lambda) + \varepsilon I$, where $\varepsilon > 0$. The corresponding M-function for this problem we denote by $M_\varepsilon(\lambda)$; the boundary conditions remain unchanged. Because $S_{2,2}(x, \lambda) + \varepsilon I$ is positive definite, it follows from the first part of the proof that

$$M_\varepsilon(\lambda) \geq -n.$$

The Θ matrices for the modified problem we denote by $\Theta_L(x, \lambda, \varepsilon)$ and $\Theta_R(x, \lambda, \varepsilon)$; by standard theory for solutions of initial value problems, these matrices are continuous functions of ε. It follows that $\arg\det\Theta_L(x, \lambda, \varepsilon)$ are continuus functions of ε. We introduce the notation $\omega_j(c, \lambda, \varepsilon)$ for the phase-angles of $\Theta_R^*\Theta_L(c, \lambda, \varepsilon)$, normalised according to

$$0 \leq \omega_j(c, \lambda, \varepsilon) < 2\pi, \tag{20}$$

and ordered so that

$$\lim_{\varepsilon \to 0+} \exp(\mathrm{i}\omega_j(c, \lambda, \varepsilon)) = \exp(\mathrm{i}\omega_j(c, \lambda)). \tag{21}$$

We want to examine the motion of the eigenvalues of the Θ matrices as functions of ε. Using methods similar to those developed in Atkinson [2, p. 307], it may be shown that

$$\frac{\partial\Theta_L}{\partial\varepsilon}(x, \lambda, \varepsilon) = \mathrm{i}\Theta_L(x, \lambda, \varepsilon)\Omega(x, \lambda, \varepsilon),$$

where Ω is the positive definite matrix given by

$$\Omega(x, \lambda, \varepsilon) = (V_L - \mathrm{i}U_L)^{-\mathrm{T}}(x, \lambda, \varepsilon)\int_a^x V_L^{\mathrm{T}}V_L(t, \lambda, \varepsilon)\mathrm{d}t \cdot (V_L + \mathrm{i}U_L)^{-1}(x, \lambda, \varepsilon).$$

Thus, the eigenvalues of $\Theta_L(x, \lambda, \varepsilon)$ move negatively around the unit circle with decreasing ε. Similarly, the eigenvalues of $\Theta_R(x, \lambda, \varepsilon)$ move positively with decreasing ε, and the eigenvalues of $\Theta_R^*\Theta_L(c, \lambda, \varepsilon)$ move negatively with decreasing ε. From this monotonicity and equation (20), it follows that the $\omega_j(c, \lambda, \varepsilon)$ are continuous from above as functions of ε (this would not be true if in (20) we had $0 < \omega_j(c, \lambda, \varepsilon) \leq 2\pi$). Hence, we see that all the terms in the expression for $M_\varepsilon(\lambda)$ are continuous from above as functions of ε, so

$$\lim_{\varepsilon \to 0+} M_\varepsilon(\lambda) = M(\lambda).$$

Hence, $M(\lambda) \geq -n$ for all λ. If the spectrum were unbounded below or had infinitely many eigenvalues decreasing towards a finite limit, the fact that

$$M(\lambda+) - M(\lambda-) \geq 1$$

at any eigenvalue λ would imply that M was unbounded below. Since in fact $M(\lambda) \geq -n$, the spectrum is bounded below and there can be no infinite sequence of eigenvalues decreasing to a finite limit. □

3. Numerical computation of the spectral function

In order to compute the spectral function $M(\lambda)$, we need to know $\arg\det\Theta_L$ and $\arg\det\Theta_R$ in (15). There are many different methods which could be used to compute these quantities. One would be to integrate the original differential system (1) and form the approximate Θ matrices from the appropriate matrix solutions of (1). Provided the method used was sympletic (e.g. a sympletic Runge–Kutta method [14]), the resulting Θ matrix approximations would be unitary. This is because sympletic integrators preserve the property

$$V(x)^{\mathrm{T}}U(x) = U(x)^{\mathrm{T}}V(x)$$

of our matrix solutions of (1), and it is precisely this property which makes the Θ matrices unitary. However, this approach is far from ideal since solutions of (1) can grow or oscillate rapidly. We therefore describe two alternative approaches to the problem of computing these quantities.

3.1. LIE PRODUCTS AND MATRIX EXPONENTIALS

We will use the well-known differential equations satisfied by Θ_L and Θ_R in order to compute $\arg\det\Theta_L$ and $\arg\det\Theta_R$.

Let Θ denote a Θ matrix, either Θ_L or Θ_R. The Θ satisfies the differential equation (see Atkinson [2, p. 305])

$$\frac{\mathrm{d}\Theta}{\mathrm{d}x} = \mathrm{i}\Theta\Omega, \tag{22}$$

where Ω is the Hermitian matrix

$$\begin{aligned}\Omega = \tfrac{1}{2}\{&(\Theta^* - I)S_{1,1}(\Theta - I) + \mathrm{i}(\Theta^* - I)S_{1,2}(\Theta + I) - \mathrm{i}(\Theta^* + I)S_{2,1}(\Theta - I)\\ &+ (\Theta^* + I)S_{2,2}(\Theta + I)\},\end{aligned} \tag{23}$$

and $S_{2,1} := S_{1,2}^{\mathrm{T}}$. The corresponding equation for $\arg\det\Theta$ is

$$\frac{\mathrm{d}}{\mathrm{d}x}\arg\det\Theta = \mathrm{Trace}(\Omega). \tag{24}$$

Taken together, (22) and (24) form a system of $n^2 + 1$ differential equations for the $n^2 + 1$ unknowns which are the elements of Θ and the crucial quantity $\arg\det\Theta$. In principle, one may solve this system using any reputable differential equation solver

– preferably a routine which evaluates the right-hand side of the differential equation by reverse communication, such as D02QGF in the NAG Library. In practice, such a method works quite well on easy problems where the coefficients in the Hamiltonian system (1) are not too large and the index of the eigenvalue sought is not too high. However, the method can fall down badly on problems which exhibit near-singularities (such as regularised singular problems) and yield inaccurate results. The reason for this is that the approximations to Θ generated by a standard ODE package will not generally be unitary, and so the correct qualitative behaviour of solutions of (22) is often lost near a singular point.

One way to overcome this problem is to represent Θ by an identity of the form

$$\Theta^* = \exp(\mathrm{i}H), \tag{25}$$

where H is an Hermitian matrix. For the problems which we consider here, where the boundary condition matrices A_1, A_2, B_1 and B_2 are real, the matrix H is in fact real and symmetric. The representation (25) has the additional advantage that provided the initial value $H(a)$ or $H(b)$ is appropriately chosen, the quantity $\arg\det\Theta$ is given by

$$\arg\det\Theta = -\mathrm{Trace}(H). \tag{26}$$

The difficulty now is to obtain a differential equation for H. In the case $n = 1$ (e.g. a normal scalar Sturm–Liouville problem), this is easy; for $n > 1$, we must use the theory described by Magnus in [6]; we outline this here for completeness.

Given two matrices A and B, the Lie product of A and B is

$$[A, B] = AB - BA.$$

Higher-order Lie products $\{A, B^m\}$ are defined inductively by

$$\{A, B^m\} := [\{A, B^{m-1}\}, B] = \{A, B^{m-1}\}B - B\{A, B^{m-1}\},$$

with $\{A, B^0\} := A$, and it is easy to see that

$$\{A, B^m\} = \sum_{k=0}^{m} \binom{m}{k} (-1)^k B^k A B^{m-k}.$$

If P is a power series defined by

$$P(B) = \sum_{k=0}^{\infty} p_k B^k, \tag{27}$$

then the Lie product $\{A, P(B)\}$ is defined by

$$\{A, P(B)\} = \sum_{k=0}^{\infty} p_k \{A, B^k\}.$$

Note that this is not the same as $[A, P(B)]$ in general, since $[A, B^k] \neq \{A, B^k\}$. With these definitions and results, we can state the differential equation for H, which is

$$\frac{dH}{dx} = -\left\{\Omega, \frac{iH}{I - \exp(-iH)}\right\}. \tag{28}$$

The right-hand side of this formula may be evaluated by expanding the term

$$\frac{iH}{I - \exp(-iH)}$$

in a power series in the usual way. This gives an infinite series. In order to sum the series, it is useful to have available the orthonormal eigenvectors $v_1, \ldots, v_n$ and corresponding real eigenvalues $\mu_1, \ldots, \mu_n$ of H. Let R be the orthogonal matrix whose columns are the eigenvectors of H, and let Δ be the matrix given by

$$\Delta = R^T \frac{dH}{dx} R,$$

so that $dH/dx = R\Delta R^T$. If we can compute Δ, then we can compute dH/dx. Now observe that the (p, q) term of Δ is given by $\Delta_{p,q} = v_p^T (dH/dx) v_q$, and so

$$\begin{aligned}
\Delta_{p,q} &= -v_p^T \left\{\Omega, \frac{iH}{I - \exp(-iH)}\right\} v_q \\
&= -v_p^T \left\{\Omega, \sum_{m=0}^{\infty} \beta_m H^m\right\} v_q \qquad \text{(series expansion)} \\
&= -\sum_{m=0}^{\infty} \beta_m v_p^T \{\Omega, H^m\} v_q \\
&= -\sum_{m=0}^{\infty} \beta_m v_p^T \sum_{k=0}^{m} \binom{m}{k} (-1)^k H^k \Omega H^{m-k} v_q \\
&= -\sum_{m=0}^{\infty} \beta_m \sum_{k=0}^{m} \binom{m}{k} (-1)^k \mu_p^k \mu_q^{m-k} v_p^T \Omega v_q \\
&= -\sum_{m=0}^{\infty} \beta_m (\mu_q - \mu_p)^m v_p^T \Omega v_q \quad \text{(binomial theorem)} \\
&= -\frac{i(\mu_q - \mu_p)}{1 - \exp(-i(\mu_q - \mu_p))} v_p^T \Omega_{\,q}.
\end{aligned} \tag{29}$$

Thus, we have been able to sum the infinite series. It remains only to compute the term $v_p^T \Omega v_q$. This is easily done using the expression (23), bearing in mind that

$$v_p^T \Theta^* = v_p^T \exp(iH) = e^{i\mu_p} v_p^T$$

and

$$\Theta v_q = \exp(-iH) v_q = e^{-i\mu_q} v_q.$$

From these expressions and (23), it is easy to see that

$$\begin{aligned} 2v_p^T \Omega v_q = {} & (e^{i\lambda_p} - 1)(e^{-i\lambda_q} - 1) v_p^T S_{1,1} v_q + i(e^{i\lambda_p} - 1)(e^{-i\lambda_q} + 1) v_p^T S_{1,2} v_q \\ & - i(e^{i\lambda_p} + 1)(e^{-i\lambda_q} - 1) v_p^T S_{2,1} v_q + (e^{i\lambda_p} + 1)(e^{-i\lambda_q} + 1) v_p^T S_{2,2} v_q. \end{aligned}$$

When this expression is substituted back into (29), the resulting expression simplifies considerably to yield

$$\begin{aligned} \Delta_{p,q} = {} & f(\lambda_p, \lambda_q) v_p^T S_{1,1} v_q + g(\lambda_p, \lambda_q) v_p^T S_{1,2} v_q \\ & - g(\lambda_q, \lambda_p) v_p^T S_{2,1} v_q + h(\lambda_p, \lambda_q) v_p^T S_{2,2} v_q, \end{aligned} \tag{30}$$

where f, g and h are the functions defined by

$$f(x, y) = \frac{(x - y)(\cos((x - y)/2) - \cos((x + y)/2))}{\sin((x - y)/2)},$$

$$g(x, y) = \frac{(x - y)(\sin((x + y)/2) + \sin((x - y)/2))}{\sin((x - y)/2)},$$

$$h(x, y) = \frac{(x - y)(\cos((x - y)/2) + \cos((x + y)/2))}{\sin((x - y)/2)}.$$

This formula shows some important features. Firstly, the singularity which appears when $\lambda_p = \lambda_q$ is removable. This is important since it means that we can compute the diagonal elements of Δ. Secondly, the best way to compute Δ is to find the eigenvectors of H, and hence the matrix R; then compute $R^T S_{1,1} R$, $R^T S_{1,2} R$, $R^T S_{2,1} R$ and $R^T S_{2,2} R$ (this can be done in less than $8n^3$ flops); then observe that, for example, $v_p^T S_{1,1} v_q = (R^T S_{1,1} R)_{p,q}$, and hence compute all the elements of Δ in a further $O(n^2)$ flops uisng (30). Thirdly, problems arise when for some p and q,

$$\lambda_p = \lambda_q + 2m\pi \quad (m \neq 0). \tag{31}$$

When this happens, the singularity is not removable. The matrix H genuinely fails to be differentiable. The trick which saves this situation is to change to a different

matrix H_{new} which is that $\exp(\mathrm{i}H) = \exp(\mathrm{i}H_{new})$, but whose eigenvalues λ^{new} are such that

$$\lambda_p^{new} = \lambda_q^{new}.$$

The singularity in the expression for the associated Δ matrix Δ_{new} is then removable. To see how H_{new} is formed, observe that for any diagonal matrix D, the matrix $R^{\mathrm{T}}DR$ commutes with H, and therefore

$$\exp(\mathrm{i}H - \mathrm{i}2\pi R^{\mathrm{T}}DR) = \exp(\mathrm{i}H)\exp(-\mathrm{i}2\pi R^{\mathrm{T}}DR) = \exp(\mathrm{i}H)R^{\mathrm{T}}\exp(-\mathrm{i}2\pi D)R. \quad (32)$$

The matrix H_{new} is chosen to be of the form

$$H_{new} = H - 2\pi R^{\mathrm{T}}DR,$$

where the diagonal matrix D has integer elements on the diagonal. From (32), this ensures that $\exp(\mathrm{i}H_{new}) = \exp(\mathrm{i}H) = \Theta^*$. Clearly, H_{new} has the same eigenvectors as H, and if $D = \mathrm{diag}(d_1, \ldots, d_n)$, then the eigenvalues of H_{new} are related to those of H by

$$\lambda_k^{new} = \lambda_k - 2\pi d_k.$$

Therefore, from (31) we have

$$\lambda_p^{new} = \lambda_q^{new} + 2\pi(d_p - d_q + m).$$

Provided the matrix D is chosen to ensure that $d_q - d_p = m$, the matrix H_{new} will be differentiable. Moreover,

$$\mathrm{Trace}(H) = \mathrm{Trace}(H_{new}) + 2\pi\mathrm{Trace}(D),$$

and hence

$$-\arg\det\Theta = \mathrm{Trace}(H_{new}) + 2\pi\mathrm{Trace}(D).$$

Thus, the all-important $\arg\det\Theta$ is recoverable from H_{new}.

In practice, this sort of change of the H matrix is carried out not only when $\lambda_p = \lambda_q + 2m\pi$ for some pair (p, q) and non-zero integer m, but also when such an equality is "almost" satisfied. There are various implementation details involved in coping with the case where there are several offending pairs (p, q) and in interpreting the "almost" so as to obtain a robust algorithm, but they are uninspiring so we do not discuss them here.

Before summarising the algorithm outlined above, we make some important remarks concerning its efficiency; we wish, in particular, to emphasise that it presents substantial advantages over the standard Magnus method, where one uses "local" H matrices, writing

$$\Theta(x) = \Theta(x_{i-1})\exp(-\mathrm{i}H_{loc}(x))$$

over each interval $[x_{i-1}, x_i]$, and where one truncates the infinite series for dH_{loc}/dx. This is undesirable for many reasons: it hides the possible points of non-differentiability in the matrix H_{loc}, resulting in regions where the stepsize becomes small for no apparent reason; it limits accuracy; and, most importantly, it is definitely not less expensive than the method we propose here. For a start, the matrices H_{loc} have to be complex Hermitian rather than real symmetric. The computation of a single Lie product then requires $4n^3$ flops, and one requires at least four such products to achieve fourth-order accuracy; the matrix Ω must be evaluated, for which an exponential of a complex matrix plus further complex matrix multiplications are required. The exact summation method which we have proposed does indeed require the diagonalisation of a real symmetric matrix, but in compensation there are no Lie products at all. The result is that it runs faster than the classical Magnus method.

In summary, then, here is the algorithm for computing $\arg\det\Theta$ currently implemented in the author's subroutine SL11F.

ALGORITHM FOR ARG DET Θ

Step 1. From initial conditions, form an initial H matrix such that $\exp(iH) = \Theta^*$. Set a real variable T to zero.

Step 2. Find the eigenvalues and eigenvectors of H. If for some pair (p, q) and integer $m \neq 0$, the equality (31) is almost satisfied, then switch to a new H as discussed above, and add to the current value of T the quantity $-2\pi\text{Trace}(D)$, where D is the matrix which shifts to the new H.

Step 3. Choose a stepsize h, and carry out a Runge–Kutta 5-4 step with error estimate, computing the required values of dH/dx as indicated above from appropriate Δ matrices. If a non-removable singularity occurs, then the step length is too long; reduce and repeat step.

Step 4. If the error estimate is too large, reduce h in the usual way for an RK 5-4 code and repeat step.

Step 5. If the end of the range of integration has not been reached, go to step 2.

Step 6. Recover $\arg\det\Theta$ as $-\text{Trace}(H) + T$.

This method is fifth-order accurate because we use a fifth-order RK method.

3.2. COEFFICIENT APPROXIMATION

A second method of finding $\arg\det\Theta_L$ and $\arg\det\Theta_R$ consists of transforming the Hamiltonian system to a second-order equation in Liouville form, to which the ideas in [8] may be applied. This is only possible when the matrix $S_{2,2}$ is positive definite. However, the computational costs may be lower than with the Magnus method described above, particularly where stiffness is present or when high index eigenvalues are sought, so it is worth having a separate code which implements this method.

As above, let Θ denote a Θ matrix, either by Θ_L or Θ_R. The coefficient approximation algorithm in the author's code involves four stages:

(1) the coefficients $S_{i,j}$ appearing in (1) are replaced by piecewise constants;

(2) on each subinterval on which the coefficients are constant, the differential equation is transformed to Liouville normal form (with, in general, non-constant coefficients);

(3) the non-constant coefficient matrices in the Liouville normal form of the ODE are replaced by constant approximations;

(4) the value of $\arg\det\Theta$ is advanced across the interval using a generalisation of the algorithm described in [8].

The first of these stages consists of choosing a mesh $(x_i)_{i=0}^N$; on each mesh interval (x_{i-1}, x_i), the coefficients $S_{p,q}$ in the differential equation are replaced by their values $\hat{S}_{p,q}$ at the centres $x_{mid} = \frac{1}{2}(x_{i-1} + x_i)$:

$$\hat{S}_{p,q} = S_{p,q}(x_{mid}), \quad p, q = 1, 2.$$

On a typical interval $[x_{i-1}, x_i]$, we then denote the corresponding approximation to Θ by Θ_1; note that if h is the greatest mesh length, then

$$\|\Theta(x_i) - \Theta_1(x_i)\| \leq Ch^2, \tag{33}$$

and hence, since the eigenvalues of a unitary matrix are well-conditioned, we may arrange that

$$|\arg\det\Theta(x_i) - \arg\det\Theta_1(x_i)| \leq Ch^2. \tag{34}$$

At the second stage, we apply Reid's transformation. Because $S_{2,2}$ is assumed to be positive definite, $\hat{S}_{2,2}$ has a positive definite square root. Thus, we may define the matrices

$$A^{(0)} = \hat{S}_{2,2}^{-1/2}\hat{S}_{1,2}^{\mathrm{T}}\hat{S}_{2,2}; \tag{35}$$

$$C^{(0)} = -\hat{S}_{2,2}^{1/2}\hat{S}_{1,1}\hat{S}_{2,2}^{1/2}; \tag{36}$$

$$A^{(1)} = \tfrac{1}{2}[A^{(0)} + (A^{(0)})^{\mathrm{T}}]; \tag{37}$$

$$A^{(2)} = \tfrac{1}{2}[A^{(0)} - (A^{(0)})^{\mathrm{T}}]. \tag{38}$$

Notice that $A^{(2)}$ is anti-symmetric. Let $M(x)$ be the orthogonal matrix given for $x_{i-1} \leq x \leq x_i$ by

$$M(x) = \exp(A^{(2)}(x - x_{mid})).$$

Suppose that Θ_1, our approximation to Θ, is given by $\Theta_1 = (V_1 + iU_1)(V_1 - iU_1)^{-1}$, where $(U_1^T, V_1^T)^T$ is a $2n$ by n solution of (1) with approximated coefficients. Define new matrices U_2 and V_2 by

$$U_2(x) = M(x)^T \hat{S}_{2,2}^{-1/2} U_1(x), \tag{39}$$

$$V_2(x) = M(x)^T [\hat{S}_{2,2}^{1/2} V_1(x) + A^{(1)} \hat{S}_{2,2}^{1/2} U_1(x)]. \tag{40}$$

This transformation maps our matrix Θ_1 to a new unitary matrix $\Theta_2 = (V_2 + iU_2)(V_2 - iU_2)^{-1}$, where

$$U_2'(x) = V_2(x), \tag{41}$$

$$V_2'(x) = Q(x) U_2(x), \tag{42}$$

and the symmetric matrix $Q(x)$ is given by

$$Q(x) = M(x)^T [C^{(0)} + A^{(1)} A^{(2)} - A^{(2)} A^{(1)}] M(x).$$

Thus, the Hamiltonian system is reduced to the Liouville normal form

$$-U_2'' + Q(x) U_2 = 0. \tag{43}$$

Θ_2 is not an approximation to Θ_1 in general. However, there is a connection formula which enables one to recover $\arg\det\Theta_1$ if $\arg\det\Theta_2$ is known. Suppose that the eigenvalues of Θ_1 are $\exp(i\alpha_j)$, where $0 \le \alpha_j < 2\pi$ for $j = 1, \ldots, n$, and suppose that the eigenvalues of Θ_2 are $\exp(i\beta_j)$, where $0 \le \beta_j < 2\pi$ for $j = 1, \ldots, n$. Then there exists a constant integer N such that for all $x \in [x_{i-1}, x_i]$,

$$\arg\det\Theta_1(x) = \arg\det\Theta_2(x) + \sum_{j=1}^{n} \alpha_j(x) - \sum_{j=1}^{n} \beta_j(x) + 2\pi N. \tag{44}$$

The connection formula may be proved (in outline) as follows. Observe that everything in the formula is continuous as a function of x except at a point where either Θ_1 or Θ_2 has an eigenvalue equal to 1, when one of the phase-angles α_j or β_j will be discontinuous. It is easy to see that 1 is an eigenvalue of Θ_1 if and only if U_1 is singular, which from (39) occurs if and only if U_2 is singular. Thus, 1 is an eigenvalue of Θ_1 if and only if 1 is an eigenvalue of Θ_2. By a more careful analysis, one may show that the multiplicity of 1 as an eigenvalue of Θ_1 is equal to the multiplicity of 1 as an eigenvalue of Θ_2, and if an eigenvalue of either Θ matrix is equal to 1 at a given value of x, then it is moving in a positive direction around the unit circle at that value of x. The net effect of these properties is that the discontinuities in the term $\sum_{j=1}^{n} \alpha_j(x)$ are exactly cancelled by the discontinuties in the term $-\sum_{j=1}^{n} \beta_j(x)$, so the integer N in (44) must be constant.

Equation (44) would therefore provide a means of computing $\arg\det\Theta_1$ were it not for the unfortunate fact that, since Q is a function of x, $\arg\det\Theta_2$ cannot be computed exactly. To overcome this problem, we approximate Q by its value $\hat{Q}$ at the centre of the interval $[x_{i-1}, x_i]$. This results in an $O(h^2)$ approximation $\hat{\Theta}_2(x_i)$ to $\Theta_2(x_i)$. Unfortunately, it also results in the loss of the connection formula (44). We therefore define a new approximation $\hat{\Theta}_1$ to Θ_1 by applying the inverse of the Reid transformation to $\hat{\Theta}_2$. The Reid transformation and its inverse always preserve connection formulae and, since $\hat{\Theta}_2(x_i)$ is an $O(h^2)$ approximation to $\Theta_2(x_i)$, it follows that $\hat{\Theta}_1(x_i)$ is an $O(h^2)$ approximation to $\Theta_1(x_i)$ and hence also to $\Theta(x_i)$, our original Θ matrix. The connection formula between $\hat{\Theta}_2$, the matrix which we can compute exactly, is the obvious analogue of (44), namely

$$\arg\det\hat{\Theta}_1(x) = \arg\det\hat{\Theta}_2(x) + \sum_{j=1}^{n}\hat{\alpha}_j(x) - \sum_{j=1}^{n}\hat{\beta}_j(x) + 2\pi\hat{N}. \tag{45}$$

Here, the $\hat{\alpha}_j$ are the phase-angles of $\hat{\Theta}_1$ and the $\hat{\beta}_j$ are the phase-angles of $\hat{\Theta}_2$; the actual value of $\hat{\Theta}_1(x_i)$ from which these are computed (using an eigenvalue routine) is easily recovered from the computed value of $\hat{\Theta}_2$.

We will now obtain an expression for $\arg\det\hat{\Theta}_2$. Let C and S be "sine-like" and "cosine-like" n by n matrix solutions of the equation

$$-U'' + \hat{Q}U = 0, \tag{46}$$

defined by the initial conditions

$$S(x_{i-1}) = 0,\ \ S'(x_{i-1}) = I,\ \ C(x_{i-1}) = I,\ \ C'(x_{i-1}) = 0. \tag{47}$$

Observe that $C(x) = S'(x)$ for all $x_{i-1} \le x \le x_i$. Actual values of C and S may be obtained by diagonalising the equation (46); we will do this later. The diagonalisation process may also be used to show that C and S commute. Now define the auxiliary matrix $\Phi(x)$ by

$$\Phi(x) = (C(x) - \mathrm{i}S(x))^{-1}(C(x) + \mathrm{i}S(x))\hat{\Theta}_2(x_{i-1}). \tag{48}$$

$\Phi(x)$ is a product of unitary matrices and is therefore unitary. The reason for introducing Φ is that $\arg\det\Phi$ is very easy to compute and, as may be shown following [8], we have a connection formula between $\hat{\Theta}_2$ and Φ.

In fact, let the eigenvalues of $\Phi(x)$ be $\exp(\mathrm{i}\varphi_j(x))$, where $0 \le \varphi_j(x) < 2\pi$. Recall that the eigenvalues of $\hat{\Theta}_2(x)$ are $\exp(\mathrm{i}\hat{\beta}_j(x))$. Let $\arg\det\Phi(x)$ be uniquely determined by the initial condition $\arg\det\Phi(x_{i-1}) = \arg\det\hat{\Theta}_2(x_{i-1})$ and the requirement of continuity, and let $\arg\det\hat{\Theta}_2(x)$ be uniquely determined by continuity plus the initial condition $\arg\det\hat{\Theta}_2(x_{i-1}) = \sum_{j=1}^{n}\hat{\beta}_j(x_{i-1})$. Then the connection formula is

$$\arg\det\hat{\Theta}_2(x) = \arg\det\Phi(x) + \sum_{j=1}^{n}\hat{\beta}_j(x) - \sum_{j=1}^{n}\varphi_j(x). \tag{49}$$

If we now substitute this result into (45), we obtain

$$\arg\det\hat{\Theta}_1(x) = \arg\det\Phi(x) - \sum_{j=1}^{n}\varphi_j(x) + \sum_{j=1}^{n}\hat{\alpha}_j(x) + 2\pi M, \tag{50}$$

where M is a fixed integer whose value we will now determine. Recalling the definition (48), we have

$$\arg\det\Phi(x) = \arg\det(C(x) - iS(x))^{-1}(C(x) + iS(x)) + \arg\det\hat{\Theta}_2(x_{i-1}), \tag{51}$$

and thus

$$\arg\det\hat{\Theta}_1(x) = \arg\det\hat{\Theta}_2(x_{i-1}) + \arg\det(C(x) - iS(x))^{-1}(C(x) + iS(x))$$
$$+ \sum_{j=1}^{n}\hat{\alpha}_j(x) - \sum_{j=1}^{n}\varphi_j(x) + 2\pi M. \tag{52}$$

Now set $x = x_{i-1}$; then $C = I$, $S = 0$ and $\Phi = \hat{\Theta}_2$, so that (with appropriate ordering) $\hat{\alpha}_j = \varphi_j$ and, with suitable normalisation, $\arg\det(C - iS)^{-1}(C + iS) = 0$. This shows that the integer M may be taken as zero provided we compute $\arg\det\Phi$ from (51) and take $\arg\det(C - iS)^{-1}(C + iS)$ to be zero when $x = x_{i-1}$. Thus, we have obtained the formula

$$\arg\det\hat{\Theta}_1(x) = \arg\det\hat{\Theta}_2(x_{i-1}) + \arg\det(C(x) - iS(x))^{-1}(C(x) + iS(x))$$
$$+ \sum_{j=1}^{n}\hat{\alpha}_j(x) - \sum_{j=1}^{n}\varphi_j(x) \tag{53}$$

for $\arg\det\hat{\Theta}_1$.

The intention now is to use $\arg\det\hat{\Theta}_1$ as an approximation to $\arg\det\Theta$. In order to use it, we must be able to evaluate all the quantities on the right-hand side of (54). To compute $\arg\det(C - iS)^{-1}(C + iS)$, we recall that C and S are matrix solutions of (46) subject to the initial conditions (47). Since $\hat{Q}$ is symmetric, there exists a diagonal matrix D and an orthogonal matrix R such that $\hat{Q} = RDR^{\mathrm{T}}$; the equation (46) may be written as

$$-U'' + RDR^{\mathrm{T}}U = 0.$$

The matrices C and S are given by $S = R\hat{S}R^{\mathrm{T}}$ and $C = R\hat{C}R^{\mathrm{T}}$, where $\hat{S}$ and $\hat{C}$ are the diagonal matrices given by the equations

$$-\hat{S}'' + D\hat{S} = 0, \quad -\hat{C}'' + D\hat{C} = 0,$$

with initial conditions

$$\hat{S}(x_{i-1}) = 0, \hat{S}'(x_{i-1}) = I; \quad \hat{C}(x_{i-1}) = I, \hat{C}'(x_{i-1}) = 0.$$

Notice that it is now obvious that C and S commute. Let d_k denote the kth diagonal element of D, c_k the kth diagonal element of $\hat{C}$, and s_k the kth diagonal element of $\hat{S}$. Then s_k is the solution of the initial value problem

$$-s_k'' + d_k s_k = 0, \quad s_k(x_{i-1}) = 0, \quad s_k'(x_{i-1}) = 1,$$

and $c_k(x) = s_k'(x)$. By using the Prüfer transformation described by Pryce in [11], we see that this initial value problem has a solution given by

$$s_k(x) = r_k(x) \sin \psi_k(x), \quad c_k(x) = r_k(x) \cos \psi_k(x), \tag{54}$$

where ψ_k is obtained by solving the initial value problem given by the initial condition $\psi_k(x_{i-1}) = 0$ and the differential equation

$$\psi_k'(x) = \cos^2 \psi_k - d_k \sin^2 \psi_k. \tag{55}$$

There is also an equation for r_k which we do not need here. The equation (55) admits an exact analytic solution which does not lose track of the number of multiples of 2π in ψ_κ; it is given in [9]. From the representation (54), it is easy to see that

$$(c_k - \mathrm{i}s_k)^{-1}(c_k + \mathrm{i}s_k) = \exp(2\mathrm{i}\psi_k),$$

and hence

$$\arg \det(\hat{C} - \mathrm{i}\hat{S})^{-1}(\hat{C} + \mathrm{i}\hat{S}) = 2\sum_{k=1}^{n} \psi_k.$$

Now the eigenvalues of $(C - \mathrm{i}S)^{-1}(C + \mathrm{i}S)$ are precisely the same as the eigenvalues of $(\hat{C} - \mathrm{i}S)^{-1}(\hat{C} + \mathrm{i}S)$, since the two matrices are related by the similarity transformation

$$(C - \mathrm{i}S)^{-1}(C + \mathrm{i}S) = R^{-1}(\hat{C} - \mathrm{i}\hat{S})^{-1}(\hat{C} + \mathrm{i}\hat{S})R;$$

hence

$$\arg \det(C - \mathrm{i}S)^{-1}(C + \mathrm{i}S) = 2\sum_{k=1}^{n} \psi_k.$$

By the remarks above, the right-hand side of this equation may be computed with the correct number of multiples of 2π by using n times the algorithm for a simple scalar Sturm–Liouville equation described in [9]; this involves solving (55). Thus, we have dealt with the problem of computing $\arg \det(C - \mathrm{i}S)^{-1}(C + \mathrm{i}S)$.

The remaining problem, that of computing the $\hat{\alpha}_j$ and φ_j in (54), is much easier to solve; it can be tackled, for example, by transforming the problem into two simultaneous eigenvalue problems for real matrices, and solved using the QZ algorithms.

This completes our description of the coefficient approximation algorithm for the computation of the approximation to $\arg \det \Theta(x_i)$. What we have not discussed

is the question of how the mesh $(x_i)_{i=0}^N$ should be computed. There are at least three different approaches to this problem: one can use extrapolation to obtain an error estimate at each step, allowing the sort of step-size control normally found in codes for initial value problems based on one-step methods; one can derive a perturbation formula for the eigenvalues which gives the change in an eigenvalue consequent upon a change in the coefficients in the differential equation, and use it to equi-distribute the eigenvalue error over the mesh as in [9]; or one can use an approach similar to that adopted by Fulton and Pruess [4], who determine an initial mesh heuristically and then apply successive bisections and extrapolations. This approach is somewhat dependent on the quality of the initial heuristic, but can yield exceptionally good accuracy at low cost; unlike the perturbation formula approach, it does not require an eigenfunction approximation either. That is why it is used in the author's code for Hamiltonian systems. Its main disadvantage is that eigenfunctions can only really be accurately evaluated at a given point if it is included in the mesh before the eigenvalue is even computed.

3.3. USING THE SPECTRAL FUNCTION TO COMPUTE EIGENVALUES

In the text following theorem 1 above, we outlined the use of the spectral function for eigenvalue computation. One starts with an eigenvalue index k and, by evaluating $M(\lambda)$ at a sequence of λ-values determined by a root-finding algorithm, one may locate a point μ such that

$$M(\mu - \varepsilon) + n \leq k, \quad M(\mu + \varepsilon) + n > k,$$

where ε is chosen according to the desired accuracy. If there is no kth eigenvalue, then no value of μ will be found and the root-finding algorithm will fail.

The computed $M(\lambda)$ will be different, in general, from the exact $M(\lambda)$. The difference will consist of a slight shift in the positions of the discontinuities by an amount which may be made arbitrarily small by reducing the integration tolerance *TOL*. Thus, although the sup-norm error in the computed $M(\lambda)$ is always large, eigenvalues can nevertheless be accurately computed.

When searching for eigenvalues of multiplicity 1, the root-finding process can be speeded up substantially by switching from the discontinuous function $M(\lambda)$ to a continuous function once an interval $[\lambda_{min}, \lambda_{max}]$ has been located which contains only the eigenvalue sought. The continuous miss-distance may be obtained from a phase-angle of the miss-matrix $\Theta_R^*(c)\Theta_L(c)$. Just how smooth this function is depends on the choice of c; there are various ways for making a reasonable choice of c.

Most eigenvalues which arise in practice are of multiplicity 1 since the discretisation process will generally break up a multiple eigenvalue into a cluser of close simple eigenvalues.

4. Numerical results

We describe here some tests of the two algorithms of section 3: these have been implemented in FORTRAN 77 codes, called SL12F (coefficient approximation) and SL11F (matrix exponentials). These codes are available from the author and from netlib/aicm/sl12f and netlib/aicm/sl11f, respectively. To obtain some idea of the efficiency of these codes, we compare them with a third code (which we call SL13F), written by the author but based on a method of Dieci et al. [3]: these authors integrate an ODE such as $\Theta' = i\Theta\Omega$ using a method which involves applying a standard integrator at each step followed by a projection to restore the lost unitarity of the matrix Θ. It is very important that this projection be applied at each step and not just at the end of the integration. The standard integrator which we use as the foundation of this projection method is the same RK 5-4 integrator as is used by SL11F to integrate the differential equation for H.

As a first test, consider the case of a scalar Sturm–Liouville problem with periodic boundary conditions:

$$-((p(x)y')' + q(x)y = \lambda w(x)y, \quad x \in [a,b]\ , \tag{56}$$

$$y(b) = y(a), \quad py'(b) = py'(a)\,. \tag{57}$$

This problem is equivalent to the following Hamiltonian problem with separated boundary conditions:

$$u' = \begin{pmatrix} 1/p(x) & 0 \\ 0 & 0 \end{pmatrix} v, \tag{58}$$

$$-v' = \begin{pmatrix} \lambda w(x) - q(x) & 0 \\ 0 & 0 \end{pmatrix} u, \tag{59}$$

$$A_1^{\mathrm{T}} u(a) + A_2^{\mathrm{T}} v(a) = 0\,, \tag{60}$$

$$B_1^{\mathrm{T}} u(b) + B_2^{\mathrm{T}} v(b) = 0\,, \tag{61}$$

where the boundary condition matrices are given by

$$A_1 = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix} = B_1,$$

$$A_2 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = B_2.$$

The eigenfunctions of the original periodic problem are recoverable from the eigenfunctions of the system by noting that if $u = (u_1, u_2)^{\mathrm{T}}$ and similarly for v, then

$$y(x) = u_1(x), \quad py'(x) = v_1(x).$$

As additional information, we also have

$$u_2(x) = y(a) = y(b), \quad v_2(x) = -py'(a) = -py'(b)$$

for all x.

The Hamiltonian system here is one in which the matrix $S_{2,2}$ is not positive definite, so the eigenvalues must be found using SL11F, which implements the first of our two algorithms for computation of $M(\lambda)$. SL11F was used to find the eigenvalues of the Mathieu equation

$$-y'' - \beta\cos(x)y = \lambda y, \quad x \in [0, 2\pi],$$

with periodic boundary conditions. When $\beta = 0$, the zeroth eigenvalue $\lambda_0 = 0$ is simple, and the other eigenvalues $\lambda_k = k^2$ are of multiplicity 2 (the eigenfunctions for λ_k, $k > 0$, live in the two-dimensional space spanned by $\cos kx$ and $\sin kx$). As β is increased, the multiple eigenvalues bifurcate into pairs of simple eigenvalues; this behaviour is evident from table 1, where we list the first five eigenvalues computed at a tolerance of 10^{-6} for various values of β; the numbers in brackets are values calculated at $TOL = 10^{-9}$ for comparison. Generally, the accuracy achieved

Table 1

Test of SL11F on Mathieu's equation (accurate results in brackets).

β	λ_0	λ_1	λ_2	λ_3	λ_4
0	1.5E-15 (0)	1.000000000 (1)	1.000000000 (1)	3.9999996 (4)	3.9999996 (4)
0.5	-0.113784 (-0.113785)	0.979256 (0.979256)	1.092825 (1.092826)	4.008246 (4.008243)	4.008458 (4.008458)
1.0	-0.378489 (-0.378489)	0.918058 (0.918058)	1.293166 (1.293166)	4.031926 (4.031922)	4.035300 (4.035301)
1.5	-0.708598 (-0.708598)	0.819230 (0.819230)	1.511299 (1.511299)	4.068180 (4.068175)	4.084679 (4.084680)
2.0	-1.070130 (-1.070130)	0.686555 (0.686720)	1.707271 (1.707269)	4.113014 (4.113009)	4.162453 (4.162454)
2.5	-1.450011 (-1.450112)	0.524864 (0.524865)	1.862280 (1.862277)	4.162060 (4.162055)	4.272143 (4.274145)
3.0	-1.842208 (-1.842208)	0.337845 (0.337845)	1.967518 (1.967516)	4.211157 (4.211150)	4.422195 (4.422196)
3.5	-2.243435 (-2.243436)	0.129366 (0.129386)	2.021658 (2.021656)	4.256672 (4.256665)	4.604151 (4.604152)

seems to be reasonable, as would be expected for such modest values of β and k. The CPU times for this test, on a Silicon Graphics Indigo Workstation, ranged from 3 seconds for the zeroth eigenvalue with $\beta = 0$ to 9 seconds for the fourth eigenvalue with $\beta = 3.5$, increasing with β and with eigenvalue index k.

As a second example, we consider a problem proposed by the author in [8]. This is the case of a matrix–vector Sturm–Liouville problem of the form

$$-\frac{1}{x}\frac{\mathrm{d}}{\mathrm{d}x}\left(x\frac{\mathrm{d}y}{\mathrm{d}x}\right)+Q(x)y=\lambda y, \quad x\in[\varepsilon,X], \tag{62}$$

with boundary conditions

$$y(\varepsilon)=0=y(x). \tag{63}$$

Here, the dependent variable is a function $y:[\varepsilon,X]\to\mathbb{R}^n$ and $Q(x)$ is a matrix function $Q:[\varepsilon,X]\to\mathcal{M}_{n\times n}(\mathbb{R})$. There are at least two ways to convert this problem into Hamiltonian form. The first of these is well known: by setting $u=y$ and $v=xy'$, the Sturm–Liouville equation (63) can be re-cast in the form

$$\begin{pmatrix}0 & -I\\ I & 0\end{pmatrix}\begin{pmatrix}u'\\ v'\end{pmatrix}=\begin{pmatrix}x(\lambda I-Q) & 0\\ 0 & x^{-1}I\end{pmatrix}\begin{pmatrix}u\\ v\end{pmatrix}, \quad x\in[\varepsilon,X], \tag{64}$$

with boundary conditions $u(\varepsilon)=0=u(x)$. The second form relies on the observation that the system

$$\begin{pmatrix}0 & -I\\ I & 0\end{pmatrix}\begin{pmatrix}u'\\ v'\end{pmatrix}=\begin{pmatrix}RP^{-1}R+\lambda xI & -RP^{-1}\\ -P^{-1}R^{\mathrm{T}} & P^{-1}\end{pmatrix}\begin{pmatrix}u\\ v\end{pmatrix} \tag{65}$$

is equivalent to the second-order equation

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(P\frac{\mathrm{d}y}{\mathrm{d}x}\right)+Ry'-(R^{\mathrm{T}}y)'=\lambda xy, \tag{66}$$

under the transformation $u=y$, $v=R^{\mathrm{T}}y+Py'$. If we choose $P(x)=xI$ and let R be a symmetric matrix such that

$$-\frac{\mathrm{d}R}{\mathrm{d}x}=xQ(x), \tag{67}$$

then (65) will evidently be equivalent to (62). This gives a more interesting form of Hamiltonian system on which both SL11F and SL12F may be tested.

The particular case of (62) which we consider here arises from an application of separation of variables to a Laplacian eigenvalue problem; details are given in [8]. It is traditional at this stage to use r (a radial coordinate) rather than x to denote the independent variable. We consider a problem of the form (62) in which the matrix Q is given by

$$Q(r) = \frac{1}{r^2} D + a_0(r) Q_c,$$

where a_0 is a function to be chosen and Q_c and D are constant matrices, D being the diagonal matrix

$$D = \text{diag}(0, 1, 1, 4, 4, 9, 9, \ldots,),$$

and Q_c the matrix whose first few rows and columns are given by

$$Q_c = \begin{pmatrix} 1 & 0 & 2^{-3/2} & 0 & 2^{-2} & 0 & 2^{-5/2} \\ 0 & 1-2^{-3} & 0 & 2^{-2}-2^{-4} & 0 & 2^{-3}-2^{-5} & 0 \\ 2^{-3/2} & 0 & 2^{-1}+2^{-4} & 0 & 2^{-2}+2^{-5} & 0 & 2^{-3}+2^{-6} \\ 0 & 2^{-2}-2^{-4} & 0 & 1-2^{-5} & 0 & 2^{-2}-2^{-6} & 0 \\ 2^{-2}0 & 0 & 2^{-2}-2^{-5} & 0 & 2^{-1}+2^{-6} & 0 & 2^{-2}+2^{-7} \\ 0 & 2^{-3}-2^{-5} & 0 & 2^{-2}-2^{-6} & 0 & 1-2^{-7} & 0 \\ 2^{-5/2} & 0 & 2^{-3}+2^{-6} & 0 & 2^{-2}+2^{-7} & 0 & 2^{-1}+2^{-8} \end{pmatrix}.$$

Note that the dimension of the system is variable and must be fixed to create a suitable test problem. Here, we choose $n = 5$.

It remains only to choose the function $a_0(r)$. We consider the following three possibilities for $a_0(r)$:

Choice 1. $a_0(r) = \frac{2}{r^2} - \frac{1}{r}$;

Choice 2. $a_0(r) = \frac{2}{r^2} - 2000(2 - e^{-1.7(r-1.3)})e^{-1.7(r-1.3)}$;

Choice 3. $a_0(r) = \frac{2}{r^{12}} - \frac{1}{r^6} - \frac{1}{r}$.

In order to compute the matrix R in (67) (recall that x is now denoted by r), we define for each choice of a_0 the function

$$b_0(r) = \int^r t a_0(t) \mathrm{d}t;$$

then clearly

$$R = -D \log(r) - b_0(r) Q_c. \tag{68}$$

Corresponding to the three choices of a_0 above, we have

Choice 4. $b_0(r) = 2 \log(r) - r$;

Choice 5. $b_0(r) = 2\log(r) - \frac{2000}{2.89}\left(2(1-1.7r) - \frac{1}{4}(1-3.4r)e^{-1.7(r-1.3)}\right)e^{-1.7(r-1.3)}$;

Choice 6. $b_0(r) = -\frac{1}{5r^{10}} + \frac{1}{4r^4} - r.$

Thus, we have two equivalent Hamiltonian systems, namely

$$\begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} r(\lambda I - Dr^{-2} - a_0(r)Q_c) & 0 \\ 0 & r^{-1}I \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}, \tag{69}$$

with $u(\varepsilon) = 0 = u(X)$, and

$$\begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} r^{-1}R^2 + \lambda rW & -r^{-1}R \\ -r^{-1}R & r^{-1}I \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}, \tag{70}$$

with the same boundary conditions and R given by (68). This allows a total of six test problems, which we define as follows: problems 1, 2 and 3 consist of the system (69) with $a_0(r)$ given, respectively, by choices 1, 2 and 3 above; and problems 4, 5 and 6 consist of the system (70) with $b_0(r)$ given, respectively, by choices 4, 5 and 6 above. The eigenvalues of problem 1 are the same as those of problem 4, the eigenvalues of problem 2 are the same as those of problem 5, and the eigenvalues of problem 3 are the same as those of problem 6. The intervals $[\varepsilon, X]$ are chosen in each case to give an eigenvalue approximation which is close to the eigenvalue of the original singular problem on $(0, \infty)$ except for problems 3 and 6, where an interval of moderate length is chosen for curiosity.

Table 2 shows some eigenvalues for these problems computed by SL11F, comparing the performance with the method of Dieci et al. as implemented in SL13F. Table 3 shows the same eigenvalues as computed by SL12F. The "exact" results are the results computed by the author for problems 1, 2 and 3 using the code SL09F described in [8], and verified by using a simple finite difference method. From table 2, we see that SL11F is faster than SL13F on problems 1, 3, 4 and 6, but slower on problems 2 and 5. The reason for the poor performance on the latter two problems is stiffness. The Dieci et al. algorithm is immune to stiffness. On the non-stiff problems, SL11F runs faster, particularly on the higher-index eigenvalue. This is not surprising since one expects the projection method to do less well where the differential equation is oscillatory. One feature which is not apparent from table 2 is that the cost of a step for SL11F is also less than the cost of a step for SL13F, by about 25%: SL11F is achieving lower run-times even though it is taking more steps. The reason for taking more steps seems to be that the error estimator used by SL11F is less optimistic than that used by SL13F, and indeed while all the results for SL11F were obtained using a tolerance of 10^{-6}, it was

Table 2

Test of SL11F on PDE-derived problems 1 to 6 ($n = 5$, $TOL = 10^{-6}$).

Problem	k	$[\varepsilon, X]$	λ_k			CPU (sec)	
			SL11F	SL13F	accurate	SL11F	SL13F
1	0	[0.01,150]	-7.82575E-2	-7.82575E-2	-7.82575E-2	75	75
1	9	[6.25E-3,576]	-8.9524E-3	-8.9524E-3	-8.9524E-3	114	162
2	0	[0.2,18]	-2587.207	-2587.208	-2587.207	543	499
2	9	[0.2,18]	-1632.913	-1632.913	-1632.913	834	513
3	0	[0.35,1]	76.328164	76.328154	76.328164	86	104
3	9	[0.35,1]	346.5955	346.5950	346.5954	151	177
4	0	[0.01,150]	-7.82575E-2	-7.82575E-2	-7.82575E-2	94	98
4	9	[6.25E-3,576]	-8.9524E-3	-8.9524E-3	-8.9524E-3	360	370
5	0	[0.2,18]	-2587.199	-2587.204	-2587.207	2027	1334
5	9	[0.2,18]	-1632.354	-1632.767	-1632.913	5360	2574
6	0	[0.35,1]	76.328165	76.328160	76.328164	138	158
6	9	[0.35,1]	346.5954	346.5955	346.5954	267	352

Table 3

Test of SL12F on PDE-derived problems 1 to 6 ($n = 5$, $TOL = 10^{-7}$).

Problem	k	$[\varepsilon, X]$	λ_k approx.	N_{mesh} (extrap.)	CPU (sec)
1	0	[0.01,150]	-7.82575E-2	224(3)	79
1	9	[0.00625,576]	-8.9524E-3	348(4)	163
2	0	[0.2,18]	-2587.207	312(2)	118
2	9	[0.2,18]	-1632.913	432(3)	158
3	0	[0.35,1]	76.328164	219(2)	165
3	9	[0.35,1]	346.59536	292(3)	258
4	0	[0.01,150]	-7.82575E-2	876(4)	701
4	9	[0.00625,576]	-8.9529E-3	1770(6)	2158
5	0	[0.2,18]	-2587.187	1630(6)	2929
5	9	[0.2,18]	-1632.903	1640(6)	2749
6	0	[0.35,1]	76.328164	612(2)	157
6	9	[0.35,1]	346.59532	612(2)	549

occasionally necessary to tighten the tolerance for SL13F to obtain comparable accuracy. As a final note on this table, we observe that the high run-times on problem 5 were partly due to the inability of the code to find a suitable matching point for the shooting, with the result that root-finding was on a much less smooth function than would have been desirable.

Turning to table 3, we see that SL12F gives a very similar performance to SL11F. The main advantage of SL12F is in its performance on problem 2 and, to a lesser extent, on problem 5. This is because SL12F is obviously immune to stiffness. The reason that the difference between the two is much more pronounced on problem 2 than on problem 5 is that on the latter the eigenfunctions and the coefficient matrices do not decay as fast for large r as in problem 2, so the coefficients have to be resolved more accurately near the endpoints to obtain good eigenvalue approximations. The lack of decay in the coefficients for large r on problems 4, 5 and 6 also explains why they are all more expensive to solve than their equivalents 1, 2 and 3, whatever the method used.

5. Concluding remarks

Differential equations of the form

$$\Theta' = \mathrm{i}\Theta\Omega(x, \Theta),$$

or in the equivalent form

$$Y' = A(x, Y)Y,$$

where A is anti-Hermitian, have been the subject of much research recently. They arise in continuous orthonormalisation processes for the solution of two-point boundary value problems and in the construction of analytic singular value decompositions. We have already mentioned the paper of Dieci et al. [3] and indeed we have implemented one of their methods. In fact, these authors also considered implicit RK methods based on Gauss quadrature. These were found to be very expensive because of the need to solve a system of nonlinear equations in a number of matrix unknowns. The projection approach was preferred, although its "brute force" nature makes it less aesthetically pleasing.

This paper has concentrated on the numerical solution of differential equation eigenvalue problems, where the differential equation can be cast as a linear Hamiltonian system. Because no assumptions have been made about the structure of the matrices $S_{i,j}$ $(i, j = 1, 2)$, the method described may not be the most efficient in individual cases. The formally self-adjoint elliptic $2m$th order differential equations examined by Greenberg [5] are a case where the coefficients $S_{i,j}$ have a very special structure. These cases will require further research if truly efficient solution algorithms are to be developed.

Acknowledgements

The author would like to thank Professor Leon Greenberg of the University of Maryland, who kindly supplied the reports [5] in advance of publication. Thanks are also due to Dr. J.D. Pryce of the Royal Military College of Science, for many interesting and fruitful discussions, and to my student Chris Graves, who wrote an improved root-finding routine for SL11F.

References

[1] M.H. Alexander and D.E. Manolopoulos, A stable linear reference potential algorithm for solution of the quantum close-coupled equations in molecular scattering theory, J. Chem. Phys. 86(1987) 2044–2050.

[2] F.V. Atkinson, *Discrete and Continuous Boundary Value Problems* (Academic Press, 1964).

[3] L. Dieci, R.D. Russell and E.S. Van Vleck, Unitary integrators and applications to continuous orthonormalization techniques, Preprint (1992).

[4] C. Fulton and S. Pruess, Mathematical software for Sturm–Liouville problems, NSF Final Report for Grants DMS88-13113 and DMS88-00839, Computational Mathematics Division (1991).

[5] L. Greenberg, A Prüfer method for calculating eigenvalues of self-adjoint systems of ordinary differential equations, Parts 1 and 2, Technical Report TR91-24, University of Maryland (1991).

[6] W. Magnus, On the exponential solution of differential equations for a linear operator, Comm. Pure Appl. Math. 7(1954)649–673.

[7] M. Marletta, Theory and implementation of algorithms for Sturm–Liouville computations, Ph.D. Thesis, Royal Military College of Science (1991).

[8] M. Marletta, Computation of eigenvalues of regular and singular vector Sturm–Liouville systems, Numer. Algor. 4(1993)65–99.

[9] M. Marletta and J.D. Pryce, Automatic solution of Sturm–Liouville problems using the Pruess method, J. Comp. Appl. Math. 39(1992)57–78.

[10] S. Pruess, Estimating the eigenvalues of Sturm–Liouville problems by approximating the differential equation, SIAM J. Numer. Anal. 10(1973)55–68.

[11] J.D. Pryce, Error control of phase-function shooting methods for Sturm–Liouville problems, IMA J. Numer. Anal. 6(1986)103–123.

[12] W.T. Reid, A continuity property of principal solutions of linear Hamiltonian differential systems, Scripta Mathematica 29(1973)337–350.

[13] W.T. Reid, *Sturmian Theory for Ordinary Differential Equations*, Applied Mathematical Sciences 31 (Springer, 1980).

[14] J.M. Sanz-Serna, Runge–Kutta schemes for Hamiltonian systems, BIT 28(1988)877–883.