THE COMPUTATION OF THRESHOLDS FOR SCHRÖDINGER OPERATORS

E. B. DAVIES

Abstract

The paper describes an approach to the computation of the zero energy thresholds for the appearance of negative energy eigenvalues of Schrödinger operators.

1. Introduction

Let H be the self-adjoint operator

$$Hf := -\Delta f + sVf$$

acting in $L^2(\mathbf{R}^N)$, where V is a real-valued potential which vanishes as $|x| \to \infty$. Subject to some local regularity assumptions on V it is known that H has continuous spectrum equal to $[0, \infty)$ together with possible negative eigenvalues of finite multiplicity which can only accumulate at zero energy. If $V(x) = o(|x|^{-2})$ at infinity and $N \geqslant 3$, then H has only a finite number of negative eigenvalues, and if s is sufficiently small it has no negative spectrum. We define the thresholds to be the values of s at which the number of negative eigenvalues, counted including multiplicities, changes as s increases. If $V \leqslant 0$ then the number may only increase with s, but we do not make this assumption. We refer to [**17**, Chapter 7] and [**3**] for general inequalites concerning the number of negative eigenvalues. If $N = 1, 2$ then one sometimes has a single negative eigenvalue for all sufficiently small values of s, and the asymptotic expansion for this eigenvalue in the weak coupling limit $s \to 0$ is known under suitable conditions on V [**17**, Chapter 7], [**10**].

Suppose that one has a method which is able to determine any negative eigenvalue of H to within a pre-determined error $\varepsilon > 0$. Unfortunately, this does not provide a means of computing the thresholds of H even approximately unless one has further information. There is the possibility that when a new negative eigenvalue emerges, it remains extremely small for a considerable interval of values of s, before starting to decrease more rapidly.

As an example, suppose that $N = 3$ and

$$V(x) = -s\left((1 + |x|)^{-3} + \delta(1 + |x|)^{-1}\right).$$

For $\delta = 0$ this has finitely many negative eigenvalues for any $s > 0$, and thresholds exist. On the other hand for $\delta > 0$, however small, the operator has infinitely many negative eigenvalues accumulating at zero and it does not make sense to speak of thresholds; see [**7**, Theorem 8.5.5]. A direct numerical application of the Rayleigh–Ritz method to H cannot distinguish between these two cases if $\delta > 0$ is sufficiently small, and so is bound to fail to discover thresholds. The numerical determination of thresholds thus seems to be intractable,

and one has to adopt a strategy which makes use of the particular structure of the operator H.

In one dimension the zero energy thresholds may be determined numerically by a study of the Riccati equation. Although our methods work in one dimension, we do not advocate them in this case for the above reason. There is, unfortunately, no analogous procedure in higher dimensions. The Rayleigh–Ritz (RR) and Temple–Lehmann (TL) inequalities provide standard methods of obtaining rigorous upper and lower bounds on the negative eigenvalues of H [7, Chapter 4], but TL becomes increasingly inaccurate for very small negative eigenvalues, since it requires a substantial spectral gap above the eigenvalue being considered if it is to yield an accurate lower bound [7, Theorem 4.6.3]. While one might be content to accept the upper bounds produced by RR as accurate approximations to the thresholds, our concern in this paper is to provide an idea which can in principle provide rigorous upper and lower bounds on the thresholds.

One method of computing the thresholds is by an application of the Lippmann–Schwinger equation [17, Chapter 7], [9], [11]. The last paper cited obtains detailed low-energy asymptotic expansions of the eigenvalues near thresholds, and shows that these depend sensitively upon the dimension. While potentially capable of providing an effective algorithm for the computation of the thresholds, as currently developed it makes heavy use of the explicit formula for the Green function of Δ. An obvious problem in applying it when H has variable second-order coefficients, or acts in $L^2(\Omega)$ for some unbounded region Ω, is that its Green function cannot then be written down in closed form, and has an infinite singularity along the diagonal $x = y$. A numerical approach along these lines would therefore require considerable thought.

The method which we investigate involves a procedure called 'pre-conditioning' in the numerical literature. We modify certain ideas of [9] by attaching the polynomial weights used there to the operator rather than to the norms of the Banach spaces. We replace the operator of interest by one with more tractable spectral behaviour, but for which the number of negative eigenvalues is unchanged. By varying the relevant parameters one may obtain rigorous upper and lower bounds on the thresholds, provided that the potential is of short enough range.

We describe the method at an abstract level in the next section, without regard to its numerical implementation. It will be seen that it can be adapted to many other operators, but we have refrained from formulating it in the most general conceivable manner. In the later sections we discuss a method of carrying out the computations involved. This involves using a particular class of test functions, which could have a finite element or wavelet basis. The two aspects of the paper are logically separate, and other numerical schemes for implementing the ideas in Section 2 may prove superior.

2. *The pre-conditioning method*

We define the operator H initially on the dense domain C_c^2 of twice continuously differentiable functions with compact support in $L^2(\mathbf{R}^N)$ by

$$Hf(x) := -\Delta f(x) + V_s(x)f(x)$$

where $s \in \mathbf{R}$ and V_s is a bounded real-valued measurable potential which vanishes sufficiently rapidly as $|x| \to \infty$ and depends continuously on s in the L^∞ norm. There is no problem in incorporating local singularities of V into our formalism, but we avoid the extra technicalities in the statements. The cases $N = 1, 2$ and $N \geqslant 3$ require slightly different

treatments for well-known reasons; see [11].

We also use the symbol H to refer to the self-adjoint operator associated with the closure of the quadratic form

$$Q(f) := \langle Hf, f \rangle$$

again initially defined on C_c^2; see [4, Chapter 4] for an exposition of the theory of quadratic forms. Actually, H is essentially self-adjoint on its initial domain, so this procedure is not necessary, but all of our subsequent arguments are based on quadratic form methods. A consequence of this is that all of our ideas can be extended to a self-adjoint operator of the form

$$H'f(x) := -\sum_{i,j=1}^{N} \frac{\partial}{\partial x_i} \left\{ a_{i,j}(x) \frac{\partial f}{\partial x_j} \right\} + X(x)f(x) + V_s(x)f(x)$$

where the second-order matrix coefficients are strictly elliptic in the sense that $a(x) \geqslant I$ for all $x \in \mathbf{R}^N$, and X is any non-negative locally bounded potential. One may also apply the ideas to Schrödinger operators acting in $L^2(\Omega)$ subject to Dirichlet boundary conditions, where Ω is a suitable region in $\mathbf{R}^N$.

Now let $\sigma \geqslant 1$ be a C^2 function on $\mathbf{R}^N$, and let $\tilde{H}$ denote the operator initially defined on C_c^2 by

$$\tilde{H}f := \sigma H(\sigma f)$$

and also the self-adjoint operator associated with the closure of its quadratic form $\tilde{Q}$. In our applications, $\sigma(x)$ diverges to $+\infty$ as $|x| \to \infty$. Note that $\mathrm{Dom}(\tilde{Q})$ need not coincide with $\{f \in L^2 : \sigma f \in \mathrm{Dom}(Q)\}$, so some care is needed in applying quadratic form methods. It is clear from its definition that we may write

$$\tilde{H} = \tilde{H}_0 + W_s$$

where $\tilde{H}_0 := -\sigma \Delta \sigma$ and $W_s := \sigma^2 V_s$; this identity is first interpreted in the quadratic form sense on C_c^2 and then extended to the form closure. The key observation of this paper is that although these two operators are not unitarily equivalent (in general), their thresholds are equal.

Lemma 1. *The numbers of negative eigenvalues of H and $\tilde{H}$ are equal, possibly both infinite. If the number is finite then the thresholds for the two operators are the same.*

Proof. Since C_c^2 is a quadratic form core for H, the number of its negative eigenvalues is equal to the supremum of the dimensions of all those finite-dimensional subspaces L of C_c^2 such that $H|_L < 0$ in the sense that $\langle Hf, f \rangle < 0$ for all non-zero $f \in L$, by the variational theorem; see [7, Theorem 4.5.2]. A similar statement applies to $\tilde{H}$. But the operator of multiplication by σ is a linear isomorphism of C_c^2, and so induces a dimension-preserving mapping of the one family of subspaces onto the other. This leads immediately to the first statement of the lemma. The second statement is a direct consequence of the definition of thresholds. $\square$

Assuming that H has only finitely many negative eigenvalues, the advantage of studying $\tilde{H}$ instead of H is that although their thresholds are the same, the bottom of the essential spectrum of $\tilde{H}$ is positive under appropriate assumptions on σ. We show in Section 5 that this enables us to obtain accurate enclosures on the eigenvalue of $\tilde{H}$ closest to zero as s

approaches a threshold by the use of RR and TL, and hence to obtain an accurate enclosure of the threshold itself.

The following theorem provides the basic constraints on the choice of σ and also a condition on V for the number of negative eigenvalues of H to be finite. The proof is related to Birman–Schwinger type results [**17**, Chapter 7], which also provide quantitative bounds on the number of negative eigenvalues.

Theorem 2. *Suppose that*

$$\beta(1 - \beta)|\nabla\sigma|^2 - (1 - \beta)\sigma\,\Delta\sigma \geqslant c + X$$

where c is a positive constant, β is a real constant and X is a continuous function such that $\lim_{|x|\to\infty} X(x) = 0$. Then $\tilde{H}_0 \geqslant cI + X$ in the sense of quadratic forms. If also

$$\lim_{|x|\to\infty} W_s(x) = 0 \tag{1}$$

then X and W_s are relatively compact with respect to $\tilde{H}_0$, and the essential spectrum of $\tilde{H}$ is contained in $[c, \infty)$. In particular the number of negative eigenvalues of H or of $\tilde{H}$ is finite.

Proof. The first statement of the theorem is an application of [**6**, Theorem 1.5.12], once one checks that $\tilde{H}_0\phi \geqslant (c + X)\phi$ for the positive C^2 function $\phi(x) := \sigma^{-\beta}$, which need not lie in L^2. Using equation (1) we next observe that

$$(\tilde{H} + i1)^{-1} - (\tilde{H}_0 - X + i1)^{-1}$$

is a compact operator by an application of [**6**, Lemma 1.6.5] and a limiting argument to deal with the fact that the potentials in question vanish at infinity instead of being of compact support. Since the entire spectrum of $\tilde{H}_0 - X$ lies in $[c, \infty)$, the second statement of the theorem is then standard [**7**, Theorem 8.4.3]. □

Our first corollary is more general but much less detailed than corresponding results in [**9**, Section 3]. Whether the eigenfunctions ϕ described below lie in $L^2(\mathbf{R}^N)$ depends sensitively on the dimension N [**9**], [**11**]: typically the resonance eigenfunction associated with the smallest threshold behaves like $|x|^{2-N}$ as $|x| \to \infty$ if $N \geqslant 3$.

Corollary 3. *Under the conditions of the above theorem every threshold s is associated with a resonance eigenfunction ϕ of H which satisfies*

$$\int_{\mathbf{R}^N} \sigma(x)^{-2}|\phi(x)|^2\mathrm{d}x < \infty.$$

Proof. By standard theorems from analytic perturbation theory, the spectrum of $\tilde{H}$ in $(-\infty, c)$ consists of isolated eigenvalues which depend analytically upon s. The thresholds of $\tilde{H}$ therefore have to be eigenvalues in the usual sense of spectral theory. If ψ is the L^2 eigenfunction of $\tilde{H}$ associated with such a threshold then

$$\langle \tilde{H}\psi, f \rangle = 0$$

for all $f \in C_c^2$. Putting $g := \sigma f$, it follows that

$$\langle H(\sigma\psi), g \rangle = 0$$

for all $g \in C_c^2$. If we put $\phi := \sigma\psi$ then we deduce that $H\phi = 0$ in the weak sense, and the required conclusion follows. □

Our applications of the above theorem are similar to [**6**, Theorem 1.5.14]; see [**5**] for related results on other variable coefficient elliptic operators.

Corollary 4. *If $N \geqslant 3$ and we put $\sigma(x) := (1 + |x|^2)^{1/2}$ and $\beta := N/2$, then we have*

$$\tilde{H}_0 \geqslant \frac{(N-2)^2}{4} I$$

in the sense of quadratic forms.

We include the next corollary for completeness, although other methods are preferable in one dimension.

Corollary 5. *Let*

$$H_0 f(x) := -\frac{d^2 f}{dx^2}$$

acting in $L^2(0, \infty)$ subject to Dirichlet boundary conditions at 0. If we put $\sigma(x) := 1 + x$ and $\beta := 1/2$ then we have

$$\tilde{H}_0 \geqslant \frac{1}{4} I$$

in the sense of quadratic forms.

The proofs are both direct computations.

Lemma 6. *Let $N = 2$ and let σ be a positive C^2 function such that $\sigma(x) = r \log r$ for all x such that $r := |x| \geqslant 2$. Then there exists $X \in C_c$ such that*

$$\tilde{H}_0 \geqslant \frac{1}{4} I + X$$

in the sense of quadratic forms. The essential spectrum of $\tilde{H}_0$ is contained in $[1/4, \infty)$ but 0 is an eigenvalue of $\tilde{H}_0$ with L^2 eigenfunction σ^{-1}.

Proof. If ϕ is a positive smooth function on $\mathbf{R}^2$ such that

$$\phi(x) := r^{-1} \log(r)^{-1/2}$$

for all $r := |x| > 2$, then a direct calculation shows that

$$-\sigma \Delta(\sigma \phi)(x) = \frac{1}{4} \phi(x)$$

for $|x| > 2$. Therefore

$$-\sigma \Delta(\sigma \phi) = \left(\frac{1}{4} + X \right) \phi$$

where X is a continuous function with support in $\{x : |x| \leqslant 2\}$. The first conclusion of the lemma follows as in Theorem 2.

It is elementary that $\sigma^{-1} \in L^2$, but the proof that it lies in the domain of the quadratic form $\tilde{Q}_0$ associated with $\tilde{H}_0$ requires some work. Let $g_n \in C_c^\infty[0, \infty)$ be a sequence of functions such that $0 \leqslant g_n \leqslant 1$, $g_n(r) = 1$ if $r \leqslant 1$ and $\lim_{n\to\infty} g_n(r) = 1$ for all $r \geqslant 0$. Define $f_n \in C_c^2(\mathbf{R}^2)$ by

$$f_n(x) := \sigma(x)^{-1} g_n(|x|).$$

A direct calculation shows that $\| f_n - \sigma^{-1} \|_2 \to 0$ as $n \to \infty$ and that

$$\tilde{Q}_0(f_n) = 2\pi \int_0^\infty |g_n'(r)|^2 r \mathrm{d}r.$$

Since $\tilde{Q}_0$ is closed it follows that if we can choose g_n such that

$$\lim_{n\to\infty} \int_0^\infty |g_n'(r)|^2 r \mathrm{d}r = 0$$

then $\tilde{Q}_0(\sigma^{-1}) = 0$ and the second statement of the lemma follows. Putting $r = \mathrm{e}^s$ and $h_n(s) := g_n(\mathrm{e}^s)$ we have equivalently to construct $h_n \in C^\infty(\mathbf{R})$ such that $0 \leqslant h_n \leqslant 1$, $h_n(s) = 1$ if $s \leqslant 0$, $h_n(s) = 0$ for large enough positive s, $\lim_{n\to\infty} h_n(s) = 1$ for all $s \in \mathbf{R}$ and

$$\lim_{n\to\infty} \int_{-\infty}^\infty |h_n'(s)|^2 \mathrm{d}s = 0.$$

We leave this as an exercise for the reader. $\qquad\square$

The statements in the following theorem which refer to H are due to Klaus and Simon [**17**, Remark 1, p. 93] and [**11**].

Corollary 7. *If $V_s = sV$ where V is not identically zero and*

$$\int_{\mathbf{R}^2} V(x)\mathrm{d}x \leqslant 0$$

then $\tilde{H}$ and hence H have at least one negative eigenvalue for all $s > 0$. The derivative of the smallest negative eigenvalue of $\tilde{H}$ with respect to s is equal to $\int_{\mathbf{R}^2} V$ at $s = 0$, while the derivative of the negative eigenvalue of H with respect to s is zero at $s = 0$, whatever the value of the integral.

Proof. Let us first consider the case where the integral is negative. If $\tilde{\lambda}_1(s)$ is the smallest eigenvalue of $\tilde{H}$, then regular perturbation theory implies that it has the convergent power series expansion

$$\tilde{\lambda}_1(s) = a_1 s + a_2 s^2 + \ldots$$

about $s = 0$, where

$$a_1 = \langle (\sigma^2 V)\sigma^{-1}, \sigma^{-1} \rangle = \int_{\mathbf{R}^2} V(x)\mathrm{d}x.$$

and σ^{-1}, defined in Lemma 6, lies in L^2. This establishes that $\tilde{\lambda}_1(s) < 0$ for small $s > 0$ and for larger s the same follows by the concavity of the function $\tilde{\lambda}_1$, which is a consequence of its variational formula [**7**, Theorem 4.5.2]. For the second statement of the corollary see [**17**, Remark 1, p. 93] and [**11**].

Now suppose that the integral vanishes. The function $s \to \tilde{\lambda}_1(s)$ is concave and real-analytic. It has zero derivative at $s = 0$ and is not identically zero, because it diverges to $-\infty$ as $s \to \infty$. Therefore $\tilde{\lambda}_1(s) < 0$ for all $s \neq 0$. $\qquad\square$

The following variation on Corollary 4 may be applicable in some cases.

Theorem 8. *Let $N \geqslant 3$ and suppose that*

$$|V_s(x)| \leqslant a(1 + |x|^2)^{-\alpha}$$

for all s and all $x \in \mathbf{R}^N$, where $\alpha > 1$. If we put

$$\sigma(x) := (1 + |x|^2)^{\alpha/2}$$

then $\|W_s\|_\infty \leqslant a$ and $\tilde{H}_0$ has purely discrete spectrum. Moreover

$$\tilde{H}_0 \geqslant (1 + |x|^2)^{\alpha-1}(N-2)^2/4$$

in the sense of quadratic forms, so $\mathrm{Spec}(\tilde{H}_0)$ is contained in $[(N-2)^2/4, \infty)$.

Proof. The second statement of the theorem is an application of [6, Theorem 1.5.12] with $\phi(x) := (1 + |x|^2)^{-\gamma}$ and $\gamma := (\alpha-1)/2 + N/4$. The compactness of the resolvent of $\tilde{H}_0$ follows as in [6, Corollary 1.6.7]. □

3. Discretisation

Throughout this section we assume that the hypotheses of Theorem 2 and Corollary 4 are satisfied; modified versions of the results hold under the assumptions of the later results in Section 2. In order to implement the above ideas numerically, we pass to suitable finite-dimensional subspaces in which matrix computations may be performed. We do this in two stages. The first depends upon the choice of a small number of auxiliary functions $\phi_1, \ldots, \phi_k$ and a small scale parameter $h > 0$. The minimum regularity assumption is that each ϕ_r lies in the Sobolev space $W_c^{1,2}$, where the subscript c denotes compact support. Particular choices of ϕ_r are discussed in the examples below.

For every $(m, r) \in \mathbf{Z}^N \times \{1, k\}$ we put

$$\phi_{m,r,h}(x) = h^{-N/2}\phi_r(x/h - m).$$

We then define the infinite matrices

$$A_{m,r;n,t} = \langle \tilde{H}_0 \phi_{m,r,h}, \phi_{n,t,h} \rangle$$
$$B_{s;m,r;n,t} = \langle W_s \phi_{m,r,h}, \phi_{n,t,h} \rangle.$$

Since ϕ_r have compact support, the above matrices are both sparse: their coefficients vanish for large enough values of $|m-n|$. This is very important for numerical computations in dimensions $N \geqslant 2$.

An examination of the definition of $A_{m,r;m,r}$ shows that it diverges to $+\infty$ as $|m| \to \infty$ and this proves that A is not associated with a bounded operator on $l^2(\mathbf{Z}^N \times \{1, k\})$.

Theorem 9. *Under certain generic assumptions on ϕ_r for $r \in \{1, k\}$ specified below, there exist positive constants c_0 and c_1 such that the two matrices A, B_s are associated with self-adjoint operators on $l^2(\mathbf{Z}^N \times \{1, k\})$ which satisfy the inequalities*

$$A \geqslant c_0 I$$
$$\|B_s\| \leqslant c_1.$$

Proof. We define the self-adjoint operator A acting in $l^2(\mathbf{Z}^N \times \{1, k\})$ to be that associated with the closure of the quadratic form

$$Q_A(f) := \sum_{m,r;n,t} A_{m,r;n,t} f_{n,t} \overline{f_{m,r}}$$

defined initially on the set of functions f of finite support.

If J is the linear operator from $l^2(\mathbf{Z}^N \times \{1, k\})$ to $L^2(\mathbf{R}^N)$ defined by

$$Jf(x) := \sum_{m,r} f_{m,r} \phi_{m,r,h}(x)$$

then

$$\|Jf\|^2 = \sum_{m,r;n,t} C_{m-n,r,t} f_{m,r} \overline{f_{n,t}}$$

where

$$C_{m-n,r,t} := \langle \phi_{m,r,h}, \phi_{n,t,h} \rangle$$
$$= \int_{\mathbf{R}^N} \phi_r(x - m + n)\overline{\phi_t(x)}\mathrm{d}x.$$

Putting

$$\hat{f}_r(\theta) := (2\pi)^{-N/2} \sum_{m \in \mathbf{Z}^N} f_{m,r} \mathrm{e}^{-im\cdot\theta}$$

for $r \in \{1, k\}$ and $\theta \in T := (-\pi, \pi)^N$, one has

$$\|Jf\|^2 = \int_T \sum_{r,t} \hat{C}_{r,t}(\theta) \hat{f}_r(\theta) \overline{\hat{f}_t(\theta)}\mathrm{d}\theta$$

where

$$\hat{C}_{r,t}(\theta) = \sum_{k \in \mathbf{Z}^N} C_{k,r,t} \mathrm{e}^{ik\cdot\theta}.$$

The fact that $\|Jf\|^2 \geqslant 0$ for all f implies that $\hat{C}(\theta) \geqslant 0$ as a matrix for all $\theta \in T$, and one should generically expect that $\hat{C}(\theta)$ is invertible for all θ. Whether or not this is the case can be checked explicitly as soon as the test functions ϕ_r have been selected; see the examples below. Making this assumption, let $\{\mu_r(\theta)\}_{r=1}^k$ be its eigenvalues written in increasing order and repeated according to multiplicity. If $\hat{C}(\theta)$ is invertible for all $\theta \in T$ then we may define the positive constants b_0 and b_1 by

$$b_0 := \min_{\theta \in T} \mu_1(\theta)^{1/2}$$
$$b_1 := \max_{\theta \in T} \mu_k(\theta)^{1/2}.$$

It follows immediately that

$$0 < b_0\|f\|_2 \leqslant \|Jf\|_2 \leqslant b_1\|f\|_2 < \infty$$

for all $0 \neq f \in l^2(\mathbf{Z}^N \times \{1, k\})$. Thus J embeds l^2 as a closed subspace of L^2.

Making this regularity assumption on J, and hence on ϕ, we now conclude from Corollary 4 that

$$\langle Af, f \rangle \geqslant (N-2)^2\|Jf\|^2/4 \geqslant (N-2)^2 b_0^2\|f\|^2/4$$

which proves the first statement of the theorem with

$$c_0 := (N-2)^2 b_0^2/4.$$

Similarly

$$|\langle B_s f, f \rangle| \leqslant \|W_s\|_\infty \|Jf\|_2^2$$

from which we deduce the second statement of the theorem with

$$c_1 := \|W_s\|_\infty b_1^2.$$

□

It is clear from the proof of the above theorem that we have a loss of numerical control depending on the condition number of the multiplication operator $\hat{C}$; that is, the size of

$$\mathrm{cond}(\hat{C}) := \frac{\max\{\mu_k(\theta) : \theta \in T\}}{\min\{\mu_1(\theta) : \theta \in T\}}.$$

This loss is reduced if the functions $\phi_{m,r,h}$ have a very little overlap for different values of m. However, this choice reduces the relationship between the spectral behaviour of $\tilde{H}_0$ and the associated operator $A = J^*\tilde{H}_0 J$ on l^2, as one sees by considering the extreme case in which $\phi_{m,r,h}$ have disjoint supports for different values of m.

Example 10. This example is related to the use of continuous compactly supported piecewise linear test functions. If we put $k := 1$ and

$$\phi(x) := \begin{cases} 1 - |x| & \text{if } |x| \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

then a direct calculation shows that

$$\langle \phi_{m,h}, \phi_{n,h} \rangle = \begin{cases} 2/3 & \text{if } m = n \\ 1/6 & \text{if } |m-n| = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore

$$\hat{C}(\theta) = 2/3 + 1/3\cos(\theta)$$

and

$$\mathrm{cond}(\hat{C}) = 3.$$

Example 11. This example is related to the use of compactly supported piecewise cubic test functions whose values and first derivatives are continuous everywhere. We put $k := 2$ and

$$\phi_1(x) := \begin{cases} (1-|x|)^2(1+2|x|) & \text{if } |x| \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\phi_2(x) := \begin{cases} 7x(1-|x|)^2 & \text{if } |x| \leqslant 1 \\ 0 & \text{otherwise.} \end{cases}$$

A routine calculation using Maple shows that

$$\hat{C}(\theta) = \begin{bmatrix} \frac{9}{35}\cos(\theta) + \frac{26}{35} & -\frac{13i}{30}\sin(\theta) \\ \frac{13i}{30}\sin(\theta) & -\frac{7}{10}\cos(\theta) + \frac{14}{15} \end{bmatrix}$$

from which one deduces that

$$\mathrm{cond}(\hat{C}) = 7.$$

The second stage in the discretisation involves choosing a finite-dimensional subspace L of $L^2(\mathbf{R}^N)$. If we simply put $L := M_0$ where

$$M_0 := \mathrm{lin}\{\phi_{m,r,h} : m \in S \text{ and } r \in \{1, k\}\}$$

for some finite subset S of $\mathbf{Z}^N$, the numerical results obtained will not be accurate unless we take a very large set S. The reason is that although threshold eigenfunctions of $\tilde{H}$ lie in L^2, they decrease rather slowly as $|x| \to \infty$. See Section 6 for an amplification of this point. It is therefore essential to adjoin to L_0 a few functions which allow one to approximate the tails of the eigenfunctions. If $V_s(x)$ is very small for $|x| \geqslant R$ then a threshold eigenfunction satisfies $\Delta(\sigma\phi) = 0$ approximately for $|x| > R$ as well as $\|\phi\|_2 < \infty$. Hence ϕ can be expanded approximately in a series, the first few terms of which are

$$\phi(x) = \alpha|x|^{1-N} + \sum_{r=1}^{N} \beta_r x_r |x|^{-1-N} + \cdots.$$

Generically, the first term dominates for large $|x|$ unless symmetry considerations force the coefficient α to vanish [11]. We therefore choose $L := M_0 \oplus M_1$ where

$$M_1 := \text{lin}\{\psi_1, \ldots, \psi_R\}$$

and $\{\psi_r\}_{r=1}^{R}$ is a small set of smooth L^2 functions which are expected to give a good approximation to the tails of a threshold eigenfunction. If H has variable second-order coefficients and V decreases slowly at infinity, then the determination of suitable ψ_r may involve substantial effort.

We next orthogonalise the functions ψ_r with respect to each other and with respect to M_0 to obtain functions ξ_r which generate a subspace M_2 for which one has an orthogonal direct sum

$$L = M_0 \oplus M_2.$$

Since R is small this is not expensive.

We define the operator $\tilde{J} : l^2(S \times \{1, k\}) \oplus M_2 \to L$ by

$$\tilde{J}(f \oplus g) = (Jf) \oplus g$$

and observe that

$$\|f \oplus g\|_2 \leqslant \|\tilde{J}(f \oplus g)\|_2 \leqslant \max\{b_1, 1\}\|f \oplus g\|_2.$$

Therefore the conclusions of Theorem 9 also apply in this context, with $l^2(\mathbf{Z}^N \times \{1, k\})$ replaced by $l^2(S \times \{1, k\}) \oplus L_2$.

Having chosen the subspace L, the restriction $C|_L$ of any operator C is defined to be the operator $C|_L : L \to L$ such that

$$\langle C|_L f, g \rangle = \langle Cf, g \rangle$$

for all $f, g \in L$. If h, ϕ_r and S are well enough chosen one may expect the thresholds s_n' of $(A + B_s)|_L$ to be close to the thresholds s_n of H. In some cases we can be more precise.

Lemma 12. *If $V_s = sV \leqslant 0$ then $s_n' \geqslant s_n$ for all n.*

Proof. For any $s > s_n'$, $(A + sB)|_L$ has at least n negative eigenvalues. By the variational theorem $\tilde{H}_0 + sW$ also has at least n negative eigenvalues, and by Lemma 1, H has at least n negative eigenvalues. Hence $s_n \leqslant s$. We now let s converge to s_n'. $\square$

4. *The matrix computation*

In this section we assume, for simplicity, that the potential depends linearly on s. It will be seen that the ideas can be adapted to the more general case. The task is to find the

eigenvalues (thresholds) of the problem

$$(A + sB)f = 0$$

when A, B are very large, sparse, self-adjoint matrices satisfying the conditions of Theorem 9, and B may have both positive and negative eigenvalues. We assume more precisely that we are given explicit positive constants c_i such that

$$c_2 I \geqslant A \geqslant c_0 I \tag{2}$$

$$\|B\| \leqslant c_1. \tag{3}$$

As s increases, the computation becomes more expensive, both because the dimension of L increases and because the size of c_2 increases. On the basis of the analysis in Section 2, we expect that for $s \neq 0$, the matrix $A + sB$ has only a few eigenvalues in the interval $[c_0 - |s| \|B\|, c_0)$, and wish to find the values of s for which it has a zero eigenvalue.

The method of determination of the thresholds which we describe below depends upon being able to compute particular eigenvalues of $A + sB$ and the corresponding eigenvectors for values of s close to the chosen threshold. Since $A + sB$ will be a very large matrix its eigenvalue closest to $\mu < c_0$ may be determined by inverse power iteration where $f \to (A + sB - \mu I)^{-1} f$ is calculated by the conjugate gradient method [**2**, Chapter 11]. Since A and B are sparse, the steps in this iteration are cheap. Since there are only a few eigenvalues less than c_0 and they are typically well separated, the inverse power iteration method is efficient. Other methods exist [**2**], [**16**], and one might consider using the subspace version of inverse power iteration.

We first describe an idealised version of our method for finding the threshold value of s for which the smallest eigenvalue $\lambda_1(s)$ of $A + sB$ vanishes. Starting from a number t_1 for which $A + t_1 B$ has a negative eigenvalue, let f_1 be the normalised eigenvector associated with $\lambda_1(t_1)$. Standard perturbation theory states that

$$\lambda_1'(t_1) = \langle Bf_1, f_1 \rangle.$$

Newton's method suggests that we put

$$t_2 := t_1 - \lambda_1(t_1)/\langle Bf_1, f_1 \rangle$$

or, equivalently, define t_2 to be the real solution of

$$\langle (A + t_2 B) f_1, f_1 \rangle = 0.$$

The variational theorem implies that $A + t_2 B$ has a non-positive eigenvalue, and we assume that it is negative since otherwise the iteration can be terminated. Applying the same idea inductively yields a sequence t_n which converges rapidly to the solution of $\lambda_1(s) = 0$ by Newton's method. The convergence of Newton's method is guaranteed by the concavity of $\lambda_1(s)$ as a function of s. If one stops the computation when $\lambda_1(t_n)$ is close to zero, then Theorem 13 below gives an explicit estimate of the difference beteen t_n and the zero of $\lambda_1(s)$.

The discussion above assumes that the smallest eigenvalue and eigenvector of $A + t_i B$ are computed exactly for each relevant value t_i. We now describe how to implement the method computationally. Starting from an initial vector f_0 and a number t_1 for which $A + t_1 B$ has a negative eigenvalue, let f_1 be a normalised vector such that

$$\langle (A + t_1 B) f_1, f_1 \rangle < 0. \tag{4}$$

The vector f_1 is constructed as an approximation to the eigenvector of $A + t_1 B$ associated with its smallest eigenvalue, but it need not be an accurate approximation, and so may be determined fairly quickly. We now define t_2 to be the solution of

$$\langle (A + t_2 B) f_1, f_1 \rangle = 0 \qquad (5)$$

so the variational theorem implies that $A + t_2 B$ has a non-positive eigenvalue. Note that inequality (4) implies that $\langle B f_1, f_1 \rangle < 0$, so that the solution t_2 of (5) satisfies $t_2 < t_1$. We now repeat the procedure starting from f_1 and t_2 and continue inductively. In principle, the decreasing sequence t_n might converge to a value larger than the solution of $\lambda_1(s) = 0$, but this can be detected and the error at any stage controlled using Theorem 13 below.

The sequence f_n of approximate eigenfunctions may be constructed within smaller subspaces than L, with great advantage. We know that the true zero energy eigenfunction lies in $L^2(\mathbf{R}^N)$ and hence is localised in space. Instead of using a single subspace L associated with a fairly large set $S \subseteq \mathbf{Z}^N$ we may consider a sequence of smaller subspaces L_n associated with an increasing sequence of subsets S_n, and let f_n be the solution of the appropriate eigenvalue problem in L_n for each n. The initial computations will then be cheaper because the matrices involved are smaller, and only the final stages in the computation, which lead to the most accurate approximation to the zero energy eigenfunction, will be expensive.

Newton's method can also be used to determine higher thresholds, even though the other eigenvalues $\lambda_n(s)$ need not be concave functions of s, provided that one starts with a value of t_1 close enough to the relevant threshold. One can obtain an initial approximation to the positions of the thresholds by computing the number of negative eigenvalues of $A + sB$ for a range of values of s. The Chebyshev iteration method provides a possible procedure for doing this [16].

Suppose that one expects that for a given value of s the matrix $A + sB$ has a small number k of negative eigenvalues, and wants to confirm this. Denote the actual number of negative eigenvalues of $A + sB$ by k'. Define the operator K by

$$Kf := f - \frac{2}{c_2 + c_1 |s|} (A + sB) f$$

so that the spectrum of K lies within $[-1, 1]$ except for k' eigenvalues larger than 1. Now take an initial linear subspace M_0 of dimension $k + 4$ (or some such number) and define

$$M_n := p_n(K) M_0$$

where p_n is the nth Chebyshev polynomial. The actual code uses the iterative formula

$$f_n := 2K f_{n-1} - f_{n-2},$$

where f_0 runs through a basis of M_0 and $f_1 := K f_0$. Generically, the subspace M_n contains very good approximations to the eigenvectors of H associated with its k' negative eigenvalues as $n \to \infty$. Hence if we restrict K to M_n it will have exactly k' eigenvalues larger than 1 for all large enough n. Note that we are only trying to determine k', not the actual magnitudes of the negative eigenvalues of $A + sB$, which are of no interest in the context of this paper.

If all of the eigenvalues of K restricted to M_n are greater than 1 for large values of n, then this indicates that $k' \geqslant k + 4$, and we have to restart the algorithm with a larger value of k.

The method is not rigorous because eigenvalues of H very close to zero will only become visible by taking n extraordinarily large, so if we terminate the iteration at some preassigned

value of n we may not detect those eigenvalues. However, we are only advocating this method as a way of getting an initial view of the likely location of the higher thresholds.

5. Obtaining enclosures on thresholds

Everything written so far leads to the construction of certain approximate solutions of $\tilde{H} f = 0$ for values of s which we expect to be close to the thresholds of $\tilde{H}$, or equivalently of H. The computations involved would presumably be done in double-precision floating-point arithmetic. In this section we describe how to obtain rigorous enclosures on the first threshold from this information. The computations of this section should be done in interval arithmetic. The following theorem reduces the problem of obtaining an accurate enclosure of s_1 to that of obtaining an accurate enclosure of $\lambda_1(s)$ for a value of s such that $\lambda_1(s)$ is very close to zero.

Theorem 13. *Let*

$$\tilde{H} := \tilde{H}_0 + s W$$

in the quadratic form sense, where

$$\tilde{H}_0 = \tilde{H}_0^* \geqslant cI > 0$$

and W is a self-adjoint relatively compact form perturbation of $\tilde{H}_0$. Then the smallest eigenvalue $\lambda_1(s)$ of $\tilde{H}$ is a concave, decreasing function for $s \geqslant 0$. The threshold s_1 of s for which it vanishes is unique. If $0 \leqslant \lambda_1(s) \leqslant \varepsilon < c$ then

$$s \leqslant s_1 \leqslant s(1 - \varepsilon/c)^{-1}.$$

On the other hand, if $-\delta \leqslant \lambda_1(s) \leqslant 0$ then

$$s(1 + \delta/c)^{-1} \leqslant s_1 \leqslant s.$$

It follows from the variational formula for the smallest eigenvalue that it is a concave function of s, and from regular perturbation theory that it is analytic. The proofs of the inequalities are elementary consequences of concavity.

Given the approximate eigenfunction f constructed in the previous sections, we obtain an upper bound on $\lambda_1(s)$ easily using RR. The lower bound can only be obtained by means of TL once we have crude lower bounds on some higher eigenvalues of $\tilde{H}$. As in [8], [13], [14], [15] this depends upon a homotopy with a comparison operator for which the eigenvalues are known precisely, as described next.

Given H and the rotationally invariant weight σ, let V_0 be a rotationally invariant potential such that $V_0(x) \leqslant V(x)$ for all $x \in \mathbf{R}^N$. We define

$$W_t(x) := \sigma(x)^2 [(1 - t)V_0(x) + t V(x)]$$

so that W_t is an increasing function of t. We then put $K_t := \tilde{H}_0 + W_t$ for $0 \leqslant t \leqslant 1$.

Since K_0 is rotationally invariant, its spectrum may be analysed independently in each angular momentum sector. This reduces one to the study of a sequence of half-line problems, and we assume that accurate enclosures of the eigenvalues of such operators can be obtained by a study of the solutions of the Riccati equation or some related method [1]. This frequently necessitates obtaining asymptotic control of the solutions of the differential equation at infinity, and solving an initial value problem for a range of parameters, which can be turned into a rigorous procedure if one uses interval arithmetic [12].

Once this has been achieved, it provides crude lower bounds on the eigenvalues of K_t by a variational comparison theorem, and this enables one to obtain accurate and rigorous lower bounds on the eigenvalues of K_{t_1} for some small enough $t_1 > 0$. An inductive procedure leads to accurate enclosures on the eigenvalues of K_{t_r} for an increasing sequence t_r and eventually for $t_n = 1$. We refer to [8], [13], [14], [15] for the elaboration of this idea.

The extension of the method to higher thresholds cannot use Theorem 13, since higher eigenvalues need not be concave functions of s. However, one may proceed by applying the intermediate value theorem after obtaining enclosures of the relevant eigenvalue at two values of s which straddle the threshold.

6. A numerical example

If H is an $M \times M$ self-adjoint matrix, the ease of computation of its smallest eigenvalue λ_1 depends not only upon how well that is separated from the remainder of the spectrum but also upon how easy it is to invert the matrix. We define the relative condition number of H to be

$$\mathrm{rcn}(H) := \frac{\lambda_M - \lambda_1}{\lambda_2 - \lambda_1}.$$

The closer this is to 1, the cheaper is the computation of λ_1 by any iterative method. In particular, if H_s depends continuously on a parameter s, then the ease of the computation of a threshold $\bar{s}$ at which $\lambda_1(\bar{s}) = 0$ depends upon the size of $\lambda_M(\bar{s})/\lambda_2(\bar{s})$.

The example which we consider in this section is a discretisation of the ordinary differential operator

$$Af(x) := -f''(x) - s \cos(\pi x) \exp(-x^2/2) f(x)$$

acting in $L^2(0, \infty)$, but we consider it as a model in its own right. We determined the eigenvalues of the various matrices using sparse matrix routines from Matlab, since the example is only one-dimensional. The operator H on $l^2\{1, \infty\}$ is defined with the aid of a positive integer k, representing the number of points per unit length, and is associated with the infinite matrix

$$H_s := H_0 - s V$$

where

$$H_{0,m,n} := \begin{cases} 2k^2 & \text{if } m = n \\ -k^2 & \text{if } |m - n| = 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$V_{m,n} := \delta_{m,n} \cos(\pi n/k) \exp(-n^2/2k^2).$$

This operator has continuous spectrum $[0, 4k^2]$, and may also have a finite number of eigenvalues outside this interval.

In order to compute the first threshold we use the weight

$$\sigma(n) := 1 + n/k.$$

Given the positive integer b, representing the length of the interval, we define M_0 to be the subspace of sequences with support in $\{1, N\}$ where $N := bk$. We let M_2 be the one-

dimensional subspace spanned by the sequence

$$u(n) := \begin{cases} 0 & \text{if } 1 \leqslant n \leqslant N \\ (1 + n/k)^{-1} & \text{if } n \geqslant N + 1 \end{cases}$$

so that u approximates the tail of any threshold eigenfunction and M_2 is orthogonal to M_0. If $h = 20$ and $b = 10$ the restriction of $\tilde{H}$ to $L := M_0 \oplus M_2$ has a zero energy threshold for $s \approx 4.2558517176$, this value being essentially independent of b for $b \geqslant 10$ (see Appendix A).

   The relative condition number of $\tilde{H}$ as defined above is rather large, 1.4735×10^5 even for $h = 20$ and $b = 10$, and it increases rapidly with b. This emphasises the importance of obtaining a good tail expansion to reduce the size of the set S ($\{1, N\}$ in this example) and hence the cost of the computation. If one does not include the vector u in the subspace L, the results obtained are very inaccurate: for $b = 10$ one gets $s \approx 4.56997$ and even for $b = 100$ one only gets $s \approx 4.28473$ (see Appendix B).

Appendix A. *Computation of threshold with 'tail' test function*

   The Maple program to find the eigenvalues of the modified operator THT when the coupling constant is set at the value for which there is a zero energy threshold is available to subscribers to the journal at:
http://www.lms.ac.uk/jcm/2/lms99003/appendix-a/.

Appendix B. *Computation of threshold with sharp cut-off on test functions*

   The Maple program to find values of s for which one has zero energy thresholds for a self-adjoint operator with a coupling constant s is available to subscribers to the journal at:
http://www.lms.ac.uk/jcm/2/lms99003/appendix-b/.

References

1. A. A. ABRAMOV, K. BALLA and N. B. KONYUKHOVA, 'Stable initial manifolds and singular boundary value problems for sets of ordinary differential equations', *Comput. Math. Banach Centre Publs* 13 (1984) 319–351. 151

2. O. AXELSSON, *Iterative solution methods* (Cambridge University Press, 1994). 149, 149

3. M. SH. BIRMAN and M. Z. SOLOMYAK, 'Estimates of the number of negative eigenvalues of the Schrödinger operator and its generalizations', *Estimates and asymptotics for discrete spectra of integral and differential equations*, Advances in Soviet Mathematics 7. (ed. M. Sh. Birman, Amer. Math. Soc., Providence, RI, 1991) 1–55. 139

4. E. B. DAVIES, *One-parameter semigroups* (Academic Press, London/New York, 1980). 141

5. E. B. DAVIES, 'L^1 properties of second order elliptic operators', *Bull. London Math. Soc.* 17 (1985) 417–436. 143

6. E. B. DAVIES, *Heat kernels and spectral theory* (Cambridge University Press, 1989). 142, 142, 143, 145, 145

7. E. B. DAVIES, *Spectral theory and differential operators* (Cambridge University Press, 1995). 139, 140, 140, 141, 142, 144

8. F. GOERISCH, 'Ein Stufenverfahren zur Berechnung von Eigenwertschranken', *Numerical treatment of eigenvalue problems* vol. 4, ISNM 83 (ed. J. Albrecht *et al.*, Birkhauser-Verlag, Basel, 1987) 104–114. 151, 152

9. A. JENSEN and T. KATO, 'Spectral properties of Schrödinger operators and time decay of the wave equation', *Duke Math. J.* 46 (1979) 583–611. 140, 140, 142, 142

10. M. KLAUS, 'On the bound state of Schrödinger operators in one dimension', *Ann. Phys.* 108 (1977) 288–300. 139

11. M. KLAUS and B. SIMON, 'Coupling constant thresholds in nonrelativistic quantum mechanics, I. Short range two body case', *Ann. Phys.* 130 (1980) 251–281. 140, 141, 142, 144, 144, 148

12. R. LOHNER, 'Verified solution of eigenvalue problems in ordinary differential equations', Unpublished manuscript, 1990. 151

13. M. PLUM, 'Eigenvalue inclusions for second order ordinary differential operators by a numerical homotopy method', *J. Appl. Math. and Phys.* 41 (1990) 205–226. 151, 152

14. M. PLUM, 'Bounds for eigenvalues of second order elliptic operators', *J. Appl. Math. and Phys.* 42 (1991) 848–863. 151, 152

15. M. PLUM, 'Guaranteed numerical bounds for eigenvalues', *Spectral theory and computational methods of Sturm–Liouville problems* (ed. D. Hinton and P. W. Schaefer, Marcel Dekker, New York/Basel, 1997). 151, 152

16. Y. SAAD, *Numerical methods for large eigenvalue problems* (Manchester University Press, 1992). 149, 150

17. B. SIMON, *Trace ideals and their applications*, London Math. Soc. Lecture Notes 35 (Cambridge University Press, 1979). 139, 139, 140, 142, 144, 144

E. B. Davies   `E.Brian.Davies@kcl.ac.uk`

Department of Mathematics
King's College
Strand
London
WC2R 2LS