Two-component eigenfunction expansion for open systems described by the wave equation I: completeness of expansion

1997 J. Phys. A: Math. Gen. 30 2139

(http://iopscience.iop.org/0305-4470/30/6/034)

J. Phys. A: Math. Gen. **30** (1997) 2139–2151. Printed in the UK PII: S0305-4470(97)73183-4

Two-component eigenfunction expansion for open systems described by the wave equation I: completeness of expansion

P T Leung, S S Tong and K Young

Department of Physics, The Chinese University of Hong Kong, Hong Kong

Received 28 March 1996, in final form 10 September 1996

Abstract. The concept of eigenfunction expansions for the wave equation is generalized to open systems, in which waves escape to the outside. These non-conservative systems are non-Hermitian in the usual sense. It is shown that the natural framework is an eigenfunction expansion within a two-component formalism that treats the wavefunction and its conjugate momentum together. Provided the system approaches spatial infinity rapidly 'without tails', and possesses spatial discontinuities, the expansion in terms of the eigenfunctions (which are now quasinormal modes) is shown to be valid.

1. Introduction

1.1. Closed and open systems

The usual concept of eigenfunction (or normal mode) expansions, central to many areas of mathematical physics, is restricted to conservative systems in which the wave energy (or other conserved quadratic expression, e.g. probability in the case of quantum mechanics) is confined: if a string is clamped at both ends, the waves are reflected, without 'leaking' to the outside. Mathematically, the nodal (or in other systems the antinodal) boundary conditions permit integration by parts without incurring surface terms, and thus ensure hermiticity, which underpins the usual mathematical formalism. Eigenfunction expansions for Hermitian systems are rooted in the Sturm–Liouville theory, of which this work is an extension.

There are two senses of eigenfunction expansion. The first is the representation of an arbitrary function $\phi(x)$ (within a certain class) as the sum of the eigenfunctions $f_n(x)$ of a certain operator H. The second sense is more important for physics: if H is the generator of time evolution, then the expansion of the initial data leads directly to a solution of the dynamics, i.e. $\phi(x,t)$ for $t > 0$, by simply attaching phase factors $\mathrm{e}^{-\mathrm{i}\omega_n t}$ to each term, where ω_n are the corresponding eigenfrequencies.

This paper develops eigenfunction expansions for the wave equation, in both of these senses, for *open* systems from which wave energy 'leaks' to the outside. Without the nodal (or antinodal) conditions, hermiticity in the usual sense is lost. Nevertheless the eigenfunction expansion can be recovered for a large class of such open systems. Many models of this type are interesting and physically relevant. Open strings [1, 2] model dissipative systems interacting with a bath. The formalism is also relevant to electromagnetic waves that can escape from an optical cavity by output coupling [3], or gravitational waves

0305-4470/97/062139+13$19.50

escaping from a region with a nontrivial background curvature, such as the vicinity of a black hole [4].

To set the stage, start with the usual conservative case. For simplicity, consider an interval I in one dimension, and scalar functions ϕ defined on I and vanishing at both ends of the interval. It is straightforward to generalize to other boundary conditions, e.g. $\mathrm{d}\phi/\mathrm{d}x$ vanishing at both ends of the interval. Then, if H is a Hermitian operator bounded from below but unbounded from above, the family of eigenfunctions $\{f_n\}$ defined by $Hf_n(x) = \omega_n f_n(x)$ is complete and orthogonal, with the eigenvalues ω_n being real. Mathematically, completeness in the first sense means that any function ϕ of this class can be expanded, in the distribution sense, as

$$\phi(x) = \sum_n a_n f_n(x) \tag{1.1}$$

whereas orthogonality ensures that the representation is unique, and also allows the coefficients a_n to be found by projecting in the standard way. More importantly, a dynamical system described by, for example, $\partial\phi/\partial t = -\mathrm{i}H\phi$ is then trivially solved for $t > 0$ by attaching phase factors $\mathrm{e}^{-\mathrm{i}\omega_n t}$ to (1.1) in which the coefficients a_n are found by projecting the initial data. We seek expansions in both of these senses for open systems.

These elementary ideas are easily generalized to the wave equation

$$\left[\rho(x)\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}\right]\phi(x,t) = 0 \tag{1.2}$$

where $\rho(x) > 0$. The eigenfunctions and eigenvalues are defined by $[\partial_x^2 + \rho(x)\omega_n^2]f_n(x) = 0$. This wave equation could describe the transverse vibrations of a string of linear density $\rho(x)$ placed under unit tension [1, 2], or the scalar analogue of electromagnetism, with $\rho(x) \equiv n(x)^2$ and $n(x)$ being the refractive index [3]. First, consider these equations defined on a finite interval I, with ϕ vanishing at both ends, so that the system is closed and conservative. The operator $-\partial^2/\partial x^2$ is then Hermitian, positive definite and unbounded from above. By the same arguments, $\{f_n\}$ is complete, and (1.1) holds. The eigenvalues ω_n^2 are real and positive, so one only needs the positive frequencies $0 < \omega_1 \leqslant \omega_2 \leqslant \cdots$ but not the corresponding set $\omega_{-n} = -\omega_n$, and we emphasize this by writing (1.1) in this case as

$$\phi(x) = \sum_{n>0} a_n f_n(x). \tag{1.3}$$

Inner products are defined by $\langle\phi|\psi\rangle = \int_I \mathrm{d}x\, \phi^*(x)\rho(x)\psi(x)$, under which the family $\{f_n\}$ are mutually orthogonal.

We now wish to generalize these notions to *open* systems defined by the wave equation under suitable restrictions on $\rho(x)$ and on the class of functions to be represented. First, let $\rho(x) > 0$, defined on $[0, \infty)$, satisfy two conditions: (a) $\rho(x)$ has a step discontinuity or stronger discontinuity (e.g. a δ-function) at some $x = a$; (b) $\rho(x) = 1$ (or another constant value) for $x > a$. We refer to these as the discontinuity condition and the 'no tail' condition respectively. The discontinuity marks the boundaries of the finite interval $I = [0, a]$, within which an eigenfunction expansion is sought. There are advantages in considering a finite interval. Physically, the interval may describe a laser cavity, and it is desirable to describe its electrodynamics without reference to the outside. There would naturally be a step discontinuity in the dielectric constant, $\rho(x)$, at the boundary of the cavity; in addition, there could be a thin slab of high dielectric constant forming a partially transmitting mirror, described by a δ-function in $\rho(x)$. Relaxation of the discontinuity and the 'no tail' condition will not be dealt with here.

Secondly, restrict attention to differentiable functions $\phi(x)$ satisfying (a) $\phi(x=0)=0$, and (b) the outgoing wavecondition for $x > a$. The escape of the waves to infinity characterizes an open system. The outgoing wavecondition (rather than the nodal condition) at $x = a^+$ renders the operator $-\partial^2/\partial x^2$ non-Hermitian on the interval $I = [0, a]$, and the familiar proofs of completeness and orthogonality break down. The theme of this and the following paper [5] is to show that, under the conditions stated, completeness and orthogonality can be recovered in a natural way. Most of the tools of mathematical-physics constructed theorems (e.g. Rayleigh–Schrödinger perturbation theory) can likewise be recovered. The half-line problem ($x > 0$) so defined is somewhat simpler to start with, and corresponds to many realistic situations, e.g. an optical cavity with a totally reflecting mirror at one end ($x = 0$), or x representing a radial variable. A generalization to a full-line problem will be sketched later.

For such open systems, completeness refers to an expansion in terms of the eigenfunctions, but also with the outgoing wavecondition at $x = a^+$. Thus, the eigenvalues ω_n are complex (with $\mathrm{Im}\,\omega_n < 0$ because the amplitude decays). The eigenfunctions $f_n(x)$ are therefore not normal modes (NMs), but quasinormal modes (QNMs).

Quite generally the QNM frequencies exist in pairs, related by $\omega_{-n} = -\omega_n^*$, where by convention $0 < \mathrm{Re}\,\omega_1 \leqslant \mathrm{Re}\,\omega_2 \leqslant \cdots$. The case where one (or more) QNM frequency falls on the imaginary axis is readily dealt with. But $\omega_{-n}^2 \neq \omega_n^2$, so the eigenfunctions $f_{-n}(x)$ and $f_n(x)$ are linearly independent. Thus, the eigenfunction expansion to be sought is (1.1) rather than (1.3), i.e. we need the full set of eigenfunctions, not just those with $\mathrm{Re}\,\omega > 0$. Many models (e.g. see section 4) contain a parameter, ϵ, characterizing the amount of leakage (e.g. $\epsilon = a/M$ for the model in section 4). The above discussion shows that there is a fundamental difference between the NMs for $\epsilon = 0$ and the QNMs for $\epsilon \neq 0$—the latter are double in number.

Under the conditions stated, the set of all QNMs $\{f_n\}$ of such an open system is complete in the interval I, in the sense of (1.1) [6]. However, earlier results [6] are not entirely satisfactory, for a number of reasons. (a) The family of QNMs $\{f_n\}$ is in fact over-complete. (b) If an inner product is introduced on the interval I, in the naive way, then the family of QNMs $\{f_n\}$ are not mutually orthogonal. It is therefore difficult to project any initial data and obtain the expansion coefficients, and therefore difficult to solve the dynamics in a straightforward manner. (c) The doubling of the eigenfunctions upon the introduction of a small amount of leakage appears perplexing. (d) The outgoing wavecondition for each eigenfunction is $f_n(x) \propto \exp(\mathrm{i}\omega_n x)$, $x > a$, but is more difficult to specify for a general function $\phi(x)$ which is to be expanded—except by saying that it is a linear superposition of such f_n's, which appears unsatisfactory.

1.2. Two-component expansion

In fact, all these difficulties are closely related to one another, and can all be avoided if the right question is posed. This can best be appreciated by considering the dynamical problem, and taking ϕ and $\rho(x)\partial\phi/\partial t$ at $t = 0$:

$$\phi(x, t=0) = \sum_n a_n f_n(x) \tag{1.4}$$

$$\rho(x)\frac{\partial\phi}{\partial t}(x, t=0) = \sum_n a_n(-\mathrm{i}\omega_n)\rho(x) f_n(x). \tag{1.5}$$

Since $\phi(x) \equiv \phi(x, t=0)$ and $\hat{\phi}(x) \equiv \rho(x)\partial\phi(x, t=0)/\partial t$ are arbitrary and can be independently specified (except for one constraint to be discussed below; see (1.11)–(1.13)),

we are led to consider the *simultaneous* expansion of a *pair* of functions $(\phi, \hat{\phi})$ (belonging to a certain class, Γ, to be specified below)

$$\begin{pmatrix} \phi(x) \\ \hat{\phi}(x) \end{pmatrix} = \sum_n a_n \begin{pmatrix} 1 \\ -\mathrm{i}\omega_n \rho(x) \end{pmatrix} f_n(x) \tag{1.6}$$

using the same coefficients, a_n, for both components. If this should be possible, then one would also wish to find a projection formula for determining the coefficients, a_n, from the given function pair $(\phi, \hat{\phi})$. An important point of this paper is that (1.6), rather than (1.1), is the right question to ask.

The inclusion of the factor of $\rho(x)$ in (1.5) turns the second component into the conjugate momentum $\hat{\phi} = \delta L/\delta(\partial_t \phi) = \rho \partial_t \phi$, where the Lagrangian of the system defined by (1.2) is $L = \frac{1}{2} \int \mathrm{d}x \, [\rho(x)(\partial_t \phi)^2 - (\partial_x \phi)^2]$.

Since conservative systems represent a limiting case (e.g. by taking $\epsilon \to 0$), we first show that (1.6) reduces to (1.1) when the eigenvalues are real. In this case, $\omega_{-n} = -\omega_n$, and up to an arbitrary constant phase factor which can be chosen to be unity, $f_{-n}(x) = f_n(x)$. Thus, upon grouping pairs of terms in (1.6),

$$\phi(x) = \sum_{n>0} (a_n + a_{-n}) f_n(x) \equiv \sum_{n>0} \beta_n f_n(x) \tag{1.7}$$

$$\hat{\phi}(x) = \rho(x) \sum_{n>0} -\mathrm{i}\omega_n (a_n - a_{-n}) f_n(x) \equiv \rho(x) \sum_{n>0} \gamma_n f_n(x). \tag{1.8}$$

So in this case, the simultaneous expansion of two functions $(\phi, \hat{\phi})$ using the full set of NMs $n = \pm 1, \pm 2, \ldots$ is equivalent to the separate expansion of each function, ϕ and $\hat{\phi}/\rho$, independently (in the sense that β_n and γ_n are independent) using only the NMs $n = 1, 2, \ldots$ in the right half of the frequency plane. Although the latter is the conventional view, the former perspective, i.e. expansion in the sense (1.6), is the appropriate one for generalizing to open systems. Dissipation destroys the symmetry under $t \to -t$, $\omega \to -\omega$, and hence the two groups of eigenfunctions ($\mathrm{Re}\,\omega > 0$ and $\mathrm{Re}\,\omega < 0$) become mixed.

The expansion is meant to apply to the class, Γ, of function pairs $(\phi, \hat{\phi})$ defined on $I = [0, a]$ with a node at $x = 0$ and the outgoing wavecondition at $x = a^+$. Because the system is characterized by $\rho(x) = 1$ for $x > a$, this condition is $\partial\phi/\partial t = -\partial\phi/\partial x$ at $x = a^+$. But since $\hat{\phi}$ represents $\rho \partial\phi/\partial t$, we define the class of pairs, Γ, to be those $(\phi, \hat{\phi})$ such that ϕ and $\hat{\phi}/\rho$ are differentiable, and satisfy (1.9)–(1.11):

$$\phi(x = 0) = 0 \tag{1.9}$$

$$\hat{\phi}(x = 0) = 0 \tag{1.10}$$

$$\hat{\phi}(x = a^+) = -\phi'(x = a^+) \tag{1.11}$$

where $' = \mathrm{d}/\mathrm{d}x$. Because $\hat{\phi}$ is not continuous across $x = a$ and ϕ' may also be discontinuous, we need to specify the point of evaluation as a^+ in (1.11). If $\rho(x)$ contains a δ-function, then $\mathrm{d}^2\phi/\mathrm{d}x^2$ would also contain a δ-function, so in general one only requires the existence of one-sided derivatives. Besides, we have already used the fact that $\rho(a^+) = 1$. Here we see another reason for dealing with a pair of functions—the outgoing wavecondition cannot be defined in terms of a single function.

Completeness, in the sense of using *all* the QNMs to expand a *single* function in the sense of (1.1) has previously been demonstrated [6], and indeed some special cases have been known earlier [7]. However, because this earlier perspective is unsatisfactory, we choose to make this paper self-contained. The rest of this paper is organized as follows. In section 2 the proof of completeness is sketched and the two-component formalism is

developed. Section 3 compares several possible expansion methods showing that the two-component formalism is the most natural and effective. Numerical examples are given in section 4. In order to focus on the main issues, the main parts of the paper discuss only the simplest possible case (a one-dimensional system defined on a half-line, with one nodal condition at $x = 0$ and one discontinuity at $x = a$, and with $\rho(x) - 1$ being strictly zero outside the discontinuity). The possibility of relaxing some of these conditions (e.g. a full-line problem, multiple discontinuities) is examined in section 5. Concluding remarks are given in section 6. The following paper [5] further develops the associated linear space structure, to prove the uniqueness of the expansion and for application to time-independent perturbation theory.

2. Completeness

2.1. Green's function

Since the eigenfunction expansion is intimately related to the dynamics, we consider the causal Green's function, defined by $[\rho(x)\partial_t^2 - \partial_x^2]G(x, y; t) = \delta(t)\delta(x - y)$, together with the initial condition $G(x, y; t) = 0$ for $t \leqslant 0$. Our starting point is the result [6] that under the discontinuity and 'no tail' conditions stated in section 1, G can be represented in terms of the eigenfunctions f_n as

$$G(x, y; t) = \frac{\mathrm{i}}{2} \sum_n \frac{1}{\omega_n} f_n(x) f_n(y) \mathrm{e}^{-\mathrm{i}\omega_n t} \tag{2.1}$$

for $x, y \in I$, $t > 0$. Specifically, this holds even for $x, y = a$, provided that $t > 0$. In (2.1), the normalization convention is $\langle f_n | f_n \rangle = 2\omega_n$, where the generalized norm is defined by

$$\langle f_n | f_n \rangle = 2\omega_n \int_0^R \mathrm{d}x\, \rho(x) f_n(x)^2 + \mathrm{i} f_n(R)^2 \tag{2.2}$$

evaluated for any $R > a$. This differs by a factor of $2\omega_n$ from the expression introduced previously [6], and is more convenient. Several remarks about this generalized norm are in order. (a) It involves f^2 rather than $|f|^2$, and is therefore in general complex. (b) It involves a surface term. (c) Each of the two terms on the right-hand side of (2.2) depends on R, but the sum is independent of R provided $R > a$. (d) The eigenvalue appears explicitly, so it is not immediately apparent how to generalize it to the norm of an arbitrary function $\langle \phi | \phi \rangle$, or to an inner product $\langle \phi | \psi \rangle$. This problem will be addressed in the next paper [5] and its solution again relates to the two-component formalism.

Given this representation of G, and the initial conditions $G(x, y; t = 0) = 0$, $\rho(x)\partial_t G(x, y; t = 0) = \delta(x - y)$, we immediately obtain (subject to the validity of term-by-term differentiation and the limit $t \to 0$, to be discussed below)

$$\frac{\mathrm{i}}{2} \sum_n \frac{1}{\omega_n} f_n(x) f_n(y) = 0 \tag{2.3}$$

$$\tfrac{1}{2}\rho(x) \sum_n f_n(x) f_n(y) = \delta(x - y). \tag{2.4}$$

Both (2.3) and (2.4) hold only for $x, y \in I$.

At first sight it may seem that (2.4), through the resolution of the identity, leads to an expansion of any function ϕ in terms of the QNMs f_n. But this naive expansion (which coincides with method C in section 3), is neither natural nor the most rapidly convergent. Instead, in section 2.2, we shall introduce the natural expansion starting from (2.1) itself.

However, before presenting this development, it is useful to at least provide some heuristic arguments for the necessity of the discontinuity and 'no tail' conditions, and also sketch the main elements of the derivation leading to (2.1) [6].

First, a representation in terms of a discrete basis can, at best, work over a finite spatial interval; without a discontinuity, there would be no natural marking of doundaries of such an interval. Secondly, consider the dynamical expansion for all times, especially large values of t. If $\rho(x)$ has a tail, then waves from a source point y can propagate to a distant point x', be scattered by the tail $\rho(x') - 1$, and return to the observation point $x \ll x'$. Since x' can be arbitrarily large, this signal would arrive at an arbitrarily late time $t \approx 2x'$ with an amplitude $\propto \rho(x') - 1 \approx \rho(t/2) - 1$. If $\rho(x)$ has a significant tail, such a wave cannot be represented as the discrete sum of QNMs, each of which decays as an exponential.

The main elements of the derivation leading to (2.1) are as follows [6]. The Fourier transform of G can be expressed as

$$\tilde{G}(x, y; \omega) = \frac{f(\omega, x)g(\omega, y)}{W(\omega)} \tag{2.5}$$

for $x < y$, where f and g are homogeneous solutions at frequency ω, with f satisfying the left boundary condition (i.e. $f = 0$ at $x = 0$) and g satisfying the right boundary condition (i.e the outgoing wavecondition for $x > a$); $W(\omega)$ is their Wronskian. If f and g are proportional (in other words at a zero of W) a QNM satisfies both boundary conditions.

Consider the inverse Fourier transform to obtain G from (2.5); the original integration is along the real axis in the ω-plane. If we distort the contour to a large semicircle in the lower half-plane, this then identifies three contributions to G.

First, the contribution from the large semicircle vanishes for all $t \geqslant 0$ if $\rho(x)$ has a discontinuity. Without the discontinuity condition, there would be a 'prompt' contribution limited to a finite time. In ray optics language, the 'prompt' contribution contains signals that travel directly from the source to the observation point without being reflected by $\rho(x)$ [8].

Secondly, there could be singularities in $f(\omega, x)$ and $g(\omega, x)$. However, provided $\rho(x)$ has 'no tail', the boundary conditions can be imposed respectively at $x = 0$ and $x = a^+$; as a result, at x, f and g are obtained by integrating the defining equations through a finite distance in x, which cannot generate any singularities in ω [9]. However, if $\rho(x)$ has a tail, $g(\omega, x)$ has a cut along the negative $\mathrm{Im}\,\omega$-axis, extending to the origin. This then leads to late time signals [8].

Under the discontinuity and 'no tail' conditions, both the prompt and the late time contributions vanish, leaving only the residues at the zeros of W. Thus, one obtains an expression for G in terms of QNMs, and some arithmetic [6] then leads to (2.1), where $f_n(x) \equiv f(\omega_n, x)$. The norm (2.1) emerges naturally from the factor of $\partial W/\partial\omega$ that appears in evaluating the residue.

For systems violating these conditions, say blackholes, the Green's function generally consists of contributions from QNMs, the prompt and the late time tail [8]. However, the expansion (2.1) will still hold within certain finite spacetime domain provided that the signals of the prompt and the tail vanish (or become insignificant) in that domain [10].

2.2. Two-component formalism

The main aim of this paper is to obtain an eigenfunction expansion from (2.1). Given functions $\phi(x)$, $\hat{\phi}(x)$ defined on $[0, a]$ and belonging to Γ, i.e. satisfying (1.9)–(1.11), we wish to demonstrate the completeness sum (1.6). For this purpose, it is convenient to

consider an initial value problem. We first extend $\phi(x)$ to $[0,\infty)$; the extension can be arbitrary except that $\phi(x)$ is differentiable and $\phi(x)\to 0$ rapidly as $x\to\infty$ (or has finite support). We also extend $\hat{\phi}(x)$ by

$$\hat{\phi}(x) = -\phi'(x) \qquad x > a. \tag{2.6}$$

This condition is compatible with continuity because $(\phi,\hat{\phi})\in\Gamma$. These extensions are purely mathematical and need not coincide with the actual wavefunction outside the interval. Moreover, the derivation below goes through even without requiring $\phi(x)$ to vanish as $x\to\infty$.

Then consider the initial value problem of finding $\phi(x,t)$ satisfying (a) the differential equation (1.2); (b) the boundary conditions $\phi(x=0,t)=0$ and $\phi(x,t)$ being an outgoing wave as $x\to\infty$; and (c) the initial conditions that

$$\phi(x,t=0)=\phi(x) \tag{2.7}$$

$$\rho(x)\frac{\partial\phi}{\partial t}(x,t=0)=\hat{\phi}(x) \tag{2.8}$$

for all x. Quite generally, the solution to this problem is

$$\phi(x,t)=\int_0^\infty \mathrm{d}y\,[G(x,y;t)\hat{\phi}(y)+\partial_t G(x,y;t)\rho(y)\phi(y)]. \tag{2.9}$$

Split the integral into the inside part $(0,a^+)$ and the outside part (a^+,∞). For the former, we use the representation of the Green's function (2.1) to obtain $\sum_n a_n(\mathrm{in})f_n(x)\mathrm{e}^{-\mathrm{i}\omega_n t}$, where the 'inside' contribution to the coefficient is

$$a_n(\mathrm{in})=\frac{\mathrm{i}}{2\omega_n}\int_0^{a^+}\mathrm{d}y\,[f_n(y)\hat{\phi}(y)+\hat{f}_n(y)\phi(y)] \tag{2.10}$$

in which we have introduced $\hat{f}_n(y)=-\mathrm{i}\omega_n\rho(y)f_n(y)$ to form the pair $(f_n,\hat{f}_n)$. This definition is consistent with viewing the second component as the initial conjugate momentum, and also with the notion that an eigenfunction corresponds to a single term in expansion (1.6).

With the outgoing wavecondition, the arbitrary initial data assumed for the 'outside' should not propagate into the 'inside', and reference to it should disappear, apart from possible dependence on the wavefunction at the interface $x=a^+$. To see this, we note that the outgoing condition on G and the property $\rho(x)=1$ for $x>a$ give $\partial_t G=-\partial_y G$, and together with (2.6), the 'outside' part of the integral in (2.9) becomes

$$-\int_a^\infty \mathrm{d}y\,[G(x,y;t)\partial_y\phi(y)+\partial_y G(x,y;t)\phi(y)]=G(x,a;t)\phi(a). \tag{2.11}$$

Using representation (2.1) at the point $y=a$, we obtain $\sum_n a_n(\mathrm{out})f_n(x)\mathrm{e}^{-\mathrm{i}\omega_n t}$, where

$$a_n(\mathrm{out})=\frac{\mathrm{i}}{2\omega_n}f_n(a)\phi(a). \tag{2.12}$$

Thus, for $x\in I$, $t>0$, the initial value problem is solved by the discrete series

$$\phi(x,t)=\sum_n a_n f_n(x)\mathrm{e}^{-\mathrm{i}\omega_n t} \tag{2.13}$$

and consequently

$$\frac{\partial\phi}{\partial t}(x,t)=\sum_n a_n(-\mathrm{i}\omega_n)f_n(x)\mathrm{e}^{-\mathrm{i}\omega_n t} \tag{2.14}$$

where a_n is given by the projection

$$a_n = \frac{\mathrm{i}}{2\omega_n}\left\{\int_0^{a^+} \mathrm{d}y\,[f_n(y)\hat{\phi}(y) + \hat{f}_n(y)\phi(y)] + f_n(a)\phi(a)\right\}. \tag{2.15}$$

These coefficients do not depend on the arbitrary extension of the initial data to the 'outside' region. The above then proves that by projecting the initial data using (2.15) will provide the solution to the later ($t > 0$) dynamical development using the sums (2.13) and (2.14)—in other words, completeness in the second sense.

The derivation shows two features. (a) The two-component structure arises from two sets of initial data. (b) All the 'outside' initial data can be collapsed to the point $x = a$, which is the physical origin of the surface term.

2.3. The zero-time limit

Taking $t \to 0^+$ in (2.13) and (2.14) yields an eigenfunction expansion like the one in (1.6), i.e. completeness in the first sense. This derivation, in particular the need to keep $t > 0$ in the intermediate steps, shows that in general the right-hand side of (1.6) should be understood as a regulated sum: attach factors $\mathrm{e}^{-\mathrm{i}\omega_n t}$, then take $t \to 0^+$ in the final result. The next section, however, shows that for step discontinuities, there is in fact no need for regularization, and we can simply differentiate term-by-term and set $t = 0$.

3. Comparison with other expansions

There are many different ways of expanding a wavefunction $\phi(x)$, so it is necessary to spell out the unique features and advantages of the method developed here. For this purpose, we compare three different expansion schemes.

(A) The first is the scheme developed here in terms of the QNMs f_n, using both ϕ and $\hat{\phi}$, and with the coefficients a_n given by (2.15).

(B) If one were to abandon the second component $\hat{\phi}$, the expansion will involve a set of coefficients, b_n, calculated from (2.15) without the $\hat{\phi}$ contribution. In other words, using $\hat{f}_n(y) = -\mathrm{i}\omega_n \rho(y) f_n(y)$, we have

$$b_n = \frac{1}{2}\int_0^{a^+} \mathrm{d}y\,\rho(y) f_n(y)\phi(y) + \frac{\mathrm{i}}{2\omega_n} f_n(a)\phi(a). \tag{3.1}$$

We can regard this as the natural expansion (i.e. method A) applied to the pair $(\phi, \hat{\phi}) = (\phi, 0)$, so the sum will certainly give $\phi(x)$ correctly.

(C) The third method makes use of the resolution of the identity in (2.4), which leads directly to an expansion with the coefficients

$$c_n = \tfrac{1}{2}\int_0^{a^+} \mathrm{d}y\,\rho(y) f_n(y)\phi(y). \tag{3.2}$$

In other words the surface term in (3.1) is ignored. It may seem surprising that neither the $\hat{\phi}$ contribution nor the surface term is necessary for representing $\phi(x)$; we shall, however, show below that there are definite advantages when these are retained, as in method A.

We now wish to compare these methods of expansion, and in doing so explain the advantages of the method developed here.

First and foremost, methods B and C will not solve the dynamical evolution for $t > 0$ by simply attaching phase factors $\mathrm{e}^{-\mathrm{i}\omega_n t}$, i.e. they do not provide a complete expansion in the second sense. This is hardly surprising since there is no knowledge of the initial $\partial_t \phi$.

Methods A, B and C all provide explicit projection formulae for the coefficients, respectively (2.15), (3.1) and (3.2). By a straightforward WKB analysis, it is easy to demonstrate several properties of these coefficients for functions ϕ with bounded derivatives up to ϕ'' and $\hat{\phi}/\rho$ being differentiable. (i) For a system with a step discontinuity, these coefficients behave asymptotically as $|a_n f_n(x)| \sim n^{-3}$, $|b_n f_n(x)| \sim n^{-2}$, $|c_n f_n(x)| \sim n^{-1}$, showing that method A is the most rapidly convergent and thus the most effective in practice. (ii) The analogous sums for $\phi'(x)$ (or in the case of method A, also $\hat{\phi}(x)$) would involve $f_n'(x)$ (or $\hat{f}_n(x)$) rather than $f_n(x)$, differing by a factor of $-\mathrm{i}\omega_n$ (or $-\mathrm{i}\rho(x)\omega_n$) which is asymptotically $\propto n$. Thus, the terms in the sum go as $|a_n f_n'(x)| \sim n^{-2}$, $|b_n f_n'(x)| \sim n^{-1}$, $|c_n f_n'(x)| \sim n^0$. Method C fails completely for $\phi'(x)$. (iii) Whether method C converges for $\phi(x)$ and whether method B converges for $\phi'(x)$ depends on the phase of the summands, since their magnitudes go as n^{-1}. For $x < a$, the phases are oscillatory in n (in fact linearly increasing with n, by an amount not equal to an integral multiple of 2π), so the sum is conditionally convergent. But for $x = a$, the phase is asymptotically constant, so these sums are logarithmically divergent. These properties are illustrated numerically in the next section.

In other words, method C (the naive projection using (2.4)) is the least convergent. Incorporation of the surface term (method B) improves it by one power; including the second component (method A) improves it by one more power provided $(\phi, \hat{\phi}) \in \Gamma$. Thus, quite apart from dynamical evolution, method A is the best. This is a principal improvement over previous results [6].

The differences can be understood in another way. Suppose we start with method A. Then method B can be obtained if $\hat{\phi}$ is identically zero. But in order for the pair to be in Γ, we would then have to force $\phi'(a^+) = 0$, even though $\phi'(a^-) \neq 0$. The discontinuous derivative naturally causes the rate of convergence to get worse. Next, method C, which drops the surface term, can be regarded as method B applied to a function $\phi(x)$ which is forced to have $\phi(a^+) = 0$ (to which the surface term is proportional), even though $\phi(a^-) \neq 0$. The discontinuity in the function (compared with the discontinuity in the derivative) uses another power.

Incidentally, the foregoing arguments also show that for method A, the rate of convergence allows the series to be differentiated and the limit $t \to 0$ taken term-by-term.

Moreover, method A leads to a unique expansion, as will be proved in the next paper [5]. On the other hand, because methods B and C are not constrained by the second component, they are not unique. To see this explicitly, one may start from identity (2.3), and project on any function $\psi(y)$ to give $\sum_n e_n f_n(x) = 0$, where

$$e_n = \frac{1}{\omega_n} \int_0^a \mathrm{d}y \; f_n(y)\psi(y) \tag{3.3}$$

and the series for methods B or C can be modified respectively according to $b_n \to b_n + e_n$ and $c_n \to c_n + e_n$, without changing the value of the sum. In particular, suppose the function ϕ to be expanded is itself a QNM: $\phi(x) = f_m(x)$. Then, as will be shown in the next paper, we would have $a_n = \delta_{nm}$, but this would not be the case for b_n and c_n.

The large number of degrees of freedom represented by the coefficients e_n can be understood in two ways. First, consider a system with negligible leakage ($\epsilon \to 0$), so that all Im $\omega_n \approx 0$, and $\omega_{-n} \approx -\omega_n$, $f_{-n} \approx f_n$. Then from (3.3), we see $e_{-n} \approx -e_n$. Thus, in this limit, only $b_n + b_{-n}$ and $c_n + c_{-n}$ are determined, not $b_n - b_{-n}$ and $c_n - c_{-n}$. This is as expected from (1.7) and (1.8). Secondly, for an open system, we note by comparison with (2.15) that $\psi(y)$ can be identified with $(\mathrm{i}/2)\hat{\phi}(y)$. In other words, methods B and C are not unique because we can add any $\hat{\phi}$, which is unspecified when one considers only the first component.

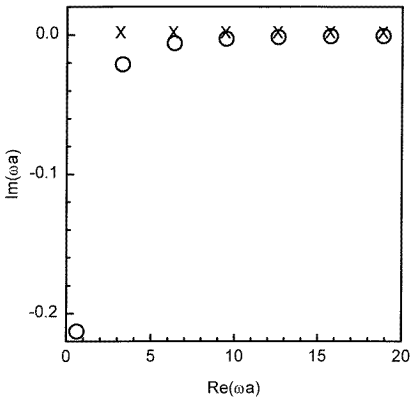

Figure 1. A schematic representation of the pole positions in the complex ω-plane for the model system (4.1). Here $\epsilon = 0.5$ (circles) and $\epsilon = 0$ (crosses).

All these remarks show that method A, based on the two-component formalism, is the best expansion method.

4. Examples

This section presents some simple examples [6] to put the discussion into context, and to illustrate the eigenfunction expansions. For the first model system, let $\rho(x) = 1+M\delta(x-a)$, $x > 0$, which can be thought of as a string of unit linear density on which a mass, M, is attached to mark off a finite interval $[1, 2]$; in the optical analogue [3], the δ-function at $x = a$ represents a thin slab of high dielectric constant which forms a partially transmitting mirror defining one end of an optical cavity, with a totally reflecting mirror at the other end $x = 0$. In the case of the Schrödinger equation, where $\rho(x)$ plays the role of potential energy function, the δ-function also mimics the effect of a very thin potential barrier [11]. This model is interesting because $\epsilon \equiv a/M$ controls the amount of leakage from the interval $I = [0, a]$, and the limit $\epsilon \to 0$ renders the system conservative, with $\phi(x)$ forced to have a node at $x = a$. The QNM frequencies are shown schematically by the circles in figure 1; for small ϵ, they are $\omega_n a \approx n\pi$, $n = \pm 1, \pm 2, \ldots$, together with a pair of 'zero modes' at $\omega a \approx \pm\epsilon^{1/2}$. As $\epsilon \to 0$, the first family turn into the NMs of the corresponding conservative system, shown by the crosses in figure 1, while the 'zero modes' disappear.

The second and even simpler example is a one-dimensional dielectric rod of index $n_0 > 1$, i.e. $\rho(x) = 1 + (n_0^2 - 1)\Theta(a - x)$, where Θ is the unit step function. Exactly this system has been discussed as a much simplified model of gravitational radiation from stellar objects [7]. The QNMs are regularly spaced along a line parallel to the real ω-axis. Because Im ω_n is constant and Re ω_n are regularly spaced, the eigenfunctions can be related directly to a standard trigonometric series, and in this manner the completeness of the QNMs (in the sense of representing *one* function rather than two) has been noticed [7], and the generality beyond the specific example has been conjectured, though the connection with a spatial discontinuity was not emphasized. This example is important in that any model in which $\rho(x)$ has a step discontinuity is asymptotically (i.e. for $n \to \infty$) very similar.

The numerical results below all refer to the dielectric rod model. Choose $\phi(x) = (x/a)\sin(\pi\alpha)$, $\hat{\phi}(x) = -(\gamma/a)\rho(x)\sin(\pi\alpha x/a)$, where α and γ are two arbitrary parameters. We have evaluated the expansion coefficients a_n by the projection formula (2.15). The partial sums on the right-hand side of (1.6) are then calculated up to $|n| \leqslant N$, and the absolute values of the differences from the original functions ϕ and $\hat{\phi}$ are then denoted as $\Delta_1(N, x)$ and $\rho(x)\Delta_2(N, x)$ respectively. First, consider the case $\gamma = 1$, for

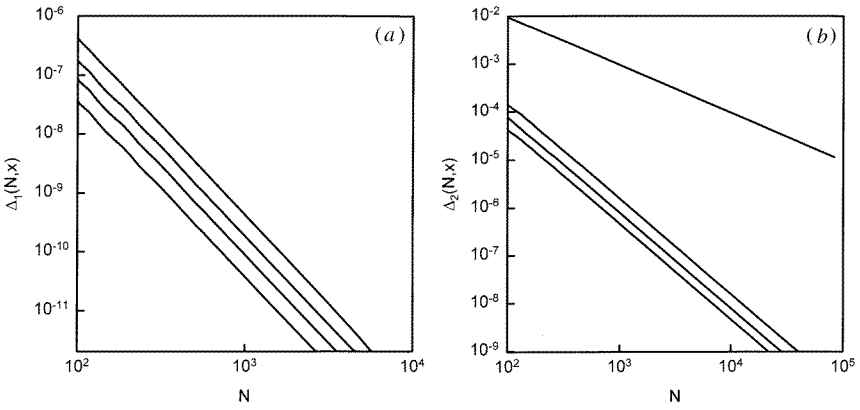

Figure 2. (*a*) $\Delta_1(N,x)$ versus N for $a=1$, $\alpha=1.7$, and $\gamma=1$. From bottom to top, $x=0.2$, $x=0.4$, $x=0.6$, $x=1$. (*b*) $\Delta_2(N,x)$ versus N for $a=1$, $\alpha=1.7$, and $\gamma=1$. From bottom to top, $x=0.2$, $x=0.6$, $x=0.8$, $x=1$.

which the function pair satisfies the outgoing wavecondition and hence belongs to Γ. We choose $a=1$ and $\alpha=1.7$, so that the point $x=a$ is not a node of ϕ. Figure 2(*a*) shows $\Delta_1(N,x)$ versus N, and figure 2(*b*) shows $\Delta_2(N,x)$ versus N, for various choices of $x\in I$. Clearly the remaining error converges rapidly to zero, in the pointwise sense, even at $x=a$, a feature already discussed in section 3.

Next suppose a different value of γ is chosen; then the pair does not satisfy the outgoing wavecondition. In this case (i) the partial sum for ϕ still converges pointwise to the correct value, for all $0\leqslant x\leqslant a$; (ii) the partial sum for $\hat{\phi}/\rho$ still converges pointwise to the correct value, for all $0\leqslant x<a$; but (iii) the partial sum for $\hat{\phi}/\rho$ does *not* converge to the correct value at $x=a$. This agrees with the discussion in the last section, namely that for $(\phi,\hat{\phi})\notin\Gamma$, a crucial cancellation does not occur, and method A is no better than method B.

We have also computed the coefficients b_n and c_n corresponding to methods B and C as described in section 3. (Since these do not depend on $\hat{\phi}$, the result is the same for any choice of γ.) Figure 3 shows the magnitudes of the summands versus n for the point $x=a$; the predicted behaviour is indeed verified. We have also verified that, despite the slower rates of convergence, the sums in methods B and C still converge to the correct $\phi(x)$ (in the case of method C excluding the point $x=a$).

5. Generalization

In this section the possibility of relaxing some of the conditions will be sketched.

5.1. The full-line problem

The formalism is readily generalized to one-dimensional systems defined on an interval I, but allowed to leak at both ends to $\pm\infty$, so that one has to consider the full line $-\infty<x<\infty$. In this case, the auxiliary function $g(\omega,x)$ in (2.5) is defined as before, and we still define $f(\omega,x)$ as the solution satisfying the left boundary condition; but instead of $f(\omega,x=0)=0$, this is now given by the outgoing wavecondition $f(\omega,x)=\mathrm{e}^{-\mathrm{i}\omega x}$, $x\to-\infty$. The 'no tail' condition now consists of two parts. First, the condition that ρ

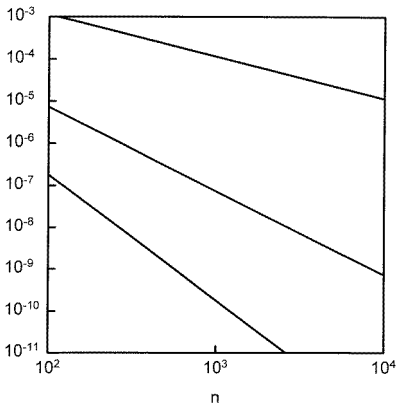

Figure 3. Magnitudes of the summands in (3.3) versus n. From bottom to top, $|a_n f_n(a)|$, $|b_n f_n(a)|$, $|c_n f_n(a)|$ where $a = 1$.

has no tail as $x \to +\infty$ guarantees that g is analytic in ω. Secondly, the condition that ρ should also have no tail as $x \to -\infty$ ensures that f is analytic in ω.

For the full-line problem, the system must have *two* discontinuities. (In fact there could be more than two; see below.) If the discontinuities occur at a_1 and a_2, with $a_1 < a_2$, then the expansion is valid inside the interval $I = [a_1, a_2]$. The projection formula (2.15) acquires two surface terms, respectively at the two end-points a_1 and a_2. All the derivation is straightforward, and will not be recorded.

In fact, the half-line case can be considered to be a special case, with the origin regarded as one of the discontinuities.

5.2. Multiple discontinuities

There could also be more than two discontinuities, say $a_1, a_2, \ldots, a_N$ assumed to be in ascending order. By matching the WKB approximations across each of these discontinuities, it is readily shown that the integral along the large semicircle in the ω-plane vanishes provided x, y lie between the outermost discontinuities. The eigenfunction expansion then holds in the interval $I = [a_1, a_N]$.

6. Conclusion

In this paper we have presented a formalism of eigenfunction expansions for open systems. For systems with discontinuities and 'no tail', the QNMs provide a complete basis for functions satisfying the outgoing waveboundary condition. Completeness is to be understood both in the usual mathematical sense of expanding a certain class of functions in terms of eigenfunctions, and also in the physical sense of solving the dynamics by attaching phase factors $e^{-i\omega_n t}$ to each term in the sum.

The two-component formalism provides the expansion in both senses—it should be obvious that any formulation that does not involve the initial $\partial_t \phi$ has no hope of solving the dynamics. But even for time-independent problems, i.e. expansion in the first sense, the two-component formalism is the most rapidly convergent and useful. The rate of convergence is such that for step or stronger discontinuities the infinite sums for both ϕ and $\hat{\phi}$ are valid (including at the end-point) without the need for regularization. The advantage is confined to function pairs $\in \Gamma$. This is not surprising, since the basis employed describes outgoing waves.

Projection (2.15) for the expansion coefficients suggests the introduction of an inner product, and hopefully the notion of orthogonality of the eigenfunctions and the self-adjointness of the time-evolution operator. These issues are addressed in the following paper [5].

Acknowledgments

We thank E S C Ching, H M Lai, S Y Liu and W M Suen for discussions. C K Au kindly commented on the manuscript. This work is supported in part by grant no 452/95P from the Hong Kong Research Grants Council.

References

[1] Dekker H 1984a *Phys. Lett.* **104A** 72
Dekker H 1984b *Phys. Lett.* **105A** 395
Dekker H 1984c *Phys. Lett.* **105A** 401
Dekker H 1985 *Phys. Rev.* A **31** 1067
[2] Lai H M, Leung P T and Young K 1987 *Phys. Lett.* **119A** 337
[3] Lang R, Scully M O and Lamb W E 1973a *Phys. Rev.* A **7** 1788
Lang R and Scully M O 1973b *Opt. Commun.* **9** 331
Penaforte J C and Baseia B 1984 *Phys. Rev.* A **30** 1401
[4] See, e.g. Chandrasekhar S 1983 *The Mathematical Theory of Black Holes* (Oxford: Oxford University Press)
[5] Leung P T, Tong S S and Young K 1997 *J. Phys. A: Math. Gen.* **30** 2153
[6] Leung P T, Liu S Y and Young K 1994a *Phys. Rev.* A **49** 3057
Leung P T, Liu S Y, Tong S S and Young K 1994b *Phys. Rev.* A **49** 3068
Leung P T, Liu S Y and Young K 1994c *Phys. Rev.* A **49** 3982
[7] Price R H and Husain V 1992 *Phys. Rev. Lett.* **68** 1973
[8] Ching E S C, Leung P T, Suen W M and Young K 1995a *Phys. Rev. Lett.* **74** 2414
Ching E S C, Leung P T, Suen W M and Young K 1995b *Phys. Rev.* D **52** 2118
[9] Newton R G 1960 *J. Math. Phys.* **1** 319
Newton R G 1982 *Scattering Theory of Waves and Particles* 2nd edn (New York: Springer)
[10] Bachelot A and Motet-Bachelot A 1993 *Ann. Henri Poincaré Phys. Theor.* **59** 3
[11] Massman H 1985 *Am. J. Phys.* **53** 679