Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 219 (2006) 296–321
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Computing spectra of linear operators using the Floquet–Fourier–Hill method

Bernard Deconinck *, J. Nathan Kutz

Department of Applied Mathematics, University of Washington, Seattle, WA 98195-2420, USA

Received 6 May 2005; received in revised form 21 March 2006; accepted 21 March 2006
Available online 15 May 2006

Abstract

In order to establish the stability of an equilibrium solution U of an infinite-dimensional dynamical system $\dot{u} = X(u)$, one is interested in the spectrum of the linear operator $L[U]$ obtained by linearizing the dynamical system around U. We use a spectrally accurate method for the computation of the spectrum of the maximal extension of the operator $L[U]$. The method is particularly well-suited to the case of periodic U, although no periodic boundary conditions on the perturbations are imposed. By incorporating the fundamentals of Floquet theory, an almost uniform approximation to the entire spectrum of the maximal extension is obtained, as opposed to an approximation of a few selected elements. The numerical component of the method is limited to: (i) choosing the size of the matrices to be used; and (ii) an eigenvalue solver, such as the QR algorithm. Compared to often-used finite-difference approaches, the method is an order of magnitude faster for comparable accuracy. We illustrate that the method is efficiently extended to problems defined on the whole line.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Linear stability; Spectral stability; Floquet theory; Bloch theory

1. Introduction

Stability plays an essential role in many branches of science and engineering, including several aspects of fluid mechanics [22], pattern formation with reaction-diffusion models [42], high-speed transmission of information [27] and feasibility of MHD fusion devices [14]. The investigation of the stability of solutions of a given mathematical model is an essential aspect of understanding the physical system considered. Stability analysis is important for at least three reasons. First, if a physical phenomenon is observable and persists, then the corresponding solution to a valid mathematical model should be stable. Second, if instability is established, the nature of the unstable modes might hint at what patterns may develop from the unstable solutions. Third, for many problems of physical interest, fundamental mathematical models are well established. However, in many cases these fundamental models are too complicated to allow for detailed analysis, thus leading to the study of simpler approximate models using reductive perturbation methods. Consequently, the stability analysis can be

* Corresponding author. Tel.: +1 2065436069.
E-mail address: bernard@amath.washington.edu (B. Deconinck).

0021-9991/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2006.03.020

used to validate or to suggest modifications to the mathematical models used for a given application. Taken as a whole, stability analysis can yield significant insight into a given physical, biological, or engineering system.

2. Problem formulation

In this paper, we focus on dynamical systems, i.e., systems with a time-like variable. In the theory of dynamical systems (see e.g. [46,18]), equilibrium solutions are singled out because of their relative simplicity within the class of all possible solutions. Furthermore, for non-Hamiltonian systems equilibrium solutions represent the simplest class of potential attractors for the system. Consider a dynamical system

$$\dot{u} = X(u), \tag{1}$$

for a function u, which may have multiple components. Here $\dot{u}$ denotes differentiation of u with respect to a time-like variable t. If $X(u)$ is an n-dimensional vector, then the system is finite dimensional, and the equilibrium solutions are points in $\mathbb{R}^n$. This case is not considered here. Instead, we consider the case where u depends not only on t, but on one or more different independent "spatial" variables, x, y, etc. For simplicity, the rest of this section treats u as a real function of only one spatial variable, thus $u = u(x, t)$ (the complex case is treated by splitting it in real and imaginary parts). Then X is an operator acting on u and its spatial derivatives, and the dynamical system is more explicitly written as

$$\dot{u} = N(x, u, u_x, u_{xx}, \ldots), \tag{2}$$

where lower indices denote differentiation with respect to x.

Equilibrium solutions $U(x)$ are functions that satisfy

$$N(x, U, U_x, U_{xx}, \ldots) = 0. \tag{3}$$

With some conditions on $N(x, U, U_x, U_{xx}, \ldots)$, the above equation has solutions by the basic Cauchy–Kovaleskaya theorem, but only those solutions that satisfy the boundary conditions imposed on (1) are of interest. Of all such possible boundary conditions, we focus specifically on periodic boundary conditions when the equation is defined on a finite domain, and on bounded solutions if the equation is defined on an infinite domain.

Having obtained such a solution, we wish to know about its stability (see for instance [29,46]) what can be said about solutions near the equilibrium solution? Will they wander away from the equilibrium solution? Will they approach it? Or will they stay nearby without approaching it? This question can be answered at different levels. Typically the operator $N(x, u, u_x, u_{xx}, \ldots)$ is nonlinear, and we want to know about stability of the equilibrium solution in the sense of Lyapunov: $U(x)$ is nonlinearly stable if for all $\epsilon > 0$ there is a $\delta > 0$ such that if $\| u(x, t) - U(x)\|_{t=0} < \delta$ then $\|u(x, t) - U(x)\| < \epsilon$ for all $t > 0$, where $\|u\|$ denotes a suitably chosen function norm. Usually this is asking for too much. The list of problems for which the stability has been settled this conclusively is short, and we have to be satisfied with different concepts of stability. One often used concept is that of *spectral stability*:

Definition. (*Spectral stability*). An equilibrium solution $U(x)$ of a dynamical system $\dot{u} = X(u)$ is spectrally stable if the spectrum of the linear operator obtained by linearizing $X(u)$ around $U(x)$ has no strictly positive real part.

Linearizing the dynamical system (1) or (2) with $u(x, t) = U(x) + \epsilon v(x, t)$ gives (assuming analyticity of N at $u(x, t) = U(x)$)

$$\dot{v} = \mathscr{L}[U(x)]v + \mathscr{O}(\epsilon), \quad \text{with } \mathscr{L}[U(x)] = \frac{\partial N}{\partial U} + \frac{\partial N}{\partial U_x}\partial_x + \frac{\partial N}{\partial U_{xx}}\partial_x^2 + \cdots. \tag{4}$$

Here expressions like $\partial N/\partial U_x$ are shorthand for $\partial N/\partial u_x|_{u=U}$. Since the right-hand side of this equation is independent of t, separation of variables (ignoring the $\mathscr{O}(\epsilon)$ term)

$$v(x, t) = w(x)e^{\lambda t} \tag{5}$$

results in

$$\mathscr{L}[U(x)]w(x) = \lambda w(x), \tag{6}$$

justifying the definition of spectral stability. This is the concept of stability we use in this paper. It is different from linear stability (for every $\epsilon > 0$ there is a $\delta > 0$ such that if $\| v(x,t)\|_{t=0} < \delta$, then $\|v(x,t)\| < \epsilon$ for all t) which results from spectral stability if the eigenfunctions of $\mathscr{L}[U(x)]$ are complete, or if all purely imaginary eigenvalues have multiplicity one. Amusingly, spectral stability is a necessary condition for nonlinear stability, whereas linear stability is not [29]. Clearly neither one of these is sufficient to give nonlinear stability.

The method presented in this paper is applicable to the more general problem of computing the spectrum of a locally-acting (no integration operators) linear operator with periodic coefficients

$$\mathscr{L} = \sum_{k=0}^{M} f_k(x)\partial_x^k, \quad f_k(x+L) = f_k(x), \tag{7}$$

which does not necessarily originate from a stability problem. If the functions $f_k(x)$, $k = 1,\ldots,M$ are scalar valued, the operator $\mathscr{L}$ is a scalar linear operator, otherwise it is a vector operator. The spectrum $\sigma(\mathscr{L})$ is defined by

$$\sigma(\mathscr{L}) = \{\lambda \in \mathbb{C} := \|w\| < \infty\}, \tag{8}$$

where w satisfies $\mathscr{L}w = \lambda w$, and is defined in a function space determined by which norm is used.

Eq. (6) is the focal point of this paper. Ideally, one is able to determine the spectrum analytically. This is only possible in some cases, and typically one has to resort to numerical methods. Depending on the system under consideration, some basic properties and bounds on the spectrum may be known [17,36,37]. This is the case for self-adjoint systems (the spectrum is confined to the real line), Hamiltonian systems (eigenvalues occur in quadruplets), etc.

Remark (*Applications to the inverse scattering method*). In addition to investigating the spectral stability of solutions of nonlinear partial differential equations, there are many more applications where the spectrum of a given linear operator is needed. As an example, consider the inverse scattering method for integrable differential equations [3,2]. Associated with every such integrable equation is a Lax pair, which consists of two linear operators. The first step of the inverse scattering program is to solve the forward scattering problem: given initial conditions for the integrable system, one calculates the spectral information associated with the first (usually spatial) element of the Lax pair. The method discussed in this paper may be used to solve this problem numerically. Some examples are given in Section 4.

Below, in Section 3, we present a method which recasts the spectral problem in an equivalent setting, using Fourier modes. Different aspects of this approach have already been used by different authors, see for instance [8,23,26,43]. Those authors were focused on solving specific problems, rather than considering the numerical method in its own right. Numerous classical examples are discussed in Section 4. Those examples were all selected because they give rise to spectra that are well understood, enabling us to separate the use of the method from the problems posed by specific applications.

3. Outline of the Floquet–Fourier–Hill method

In this section, we present a Floquet–Fourier-analysis based method. This method was first used with minor variations by Hill in 1886 [28] to determine the spectrum of the equation that now bears his name. Because of this we refer to the method as the Fourier–Floquet–Hill method (FFHM), or sometimes Hill's method, in short. The method allows for the computation of the spectra of linear operators and has numerous advantages over finite difference methods. Below we show that:

1. The periodicity of the coefficients of the linear operator is used explicitly by using Fourier series.
2. The FFHM is spectrally accurate in the sense that the approximations of the spectral elements it computes are spectrally accurate, as a result of the use of these Fourier series approximations.
3. These spectral elements are obtained as eigenvalues of a truncated difference equation. This difference equation is *equivalent* to the original problem (6). Thus the lattice used for the computation of the spectral elements is inherent in the original Eq. (6) considered: it is the lattice conjugate to the period lattice of Eq. (6).

4. If we settle on the QR algorithm for the computation of the eigenvalues of the matrix obtained from this truncated difference equation, then the only fundamental numerical decision to be made is that of how to truncate the domain of the difference equation. The algorithm in its most general form allows for some additional choices, but these are not essential.
5. The incorporation of Floquet theory into the FFHM allows for an almost uniform approximation of those components of the spectrum that are being approximated, as opposed to isolated elements of it.

For convenience, we assume that all coefficients of the operator $\mathscr{L}$ in (7) are analytic, so that integrals, derivatives and sums, finite or infinite, proper or improper, may be interchanged when so desired. For the sake of explicitness, we also assume below $n = 1$, i.e., the operator $\mathscr{L}$ is scalar. A vector example is considered in Section 4.4.

Since the coefficients of $\mathscr{L}$ are periodic with period L, they may be represented by a Fourier series:

$$f_k(x) = \sum_{j=-\infty}^{\infty} \hat{f}_{k,j} e^{i2\pi jx/L}, \quad k = 0, \ldots, M \tag{9}$$

with

$$\hat{f}_{k,j} = \frac{1}{L} \int_{-L/2}^{L/2} f_k(x) e^{-i2\pi jx/L} \, dx, \quad k = 0, \ldots, M, \ \ j \in \mathbb{Z}, \tag{10}$$

where i denotes the imaginary unit. Additional relationships exist between some of these coefficients if $f_k(x)$ is real-valued, but no such constraint needs to be imposed now.

We start by recalling Floquet's theorem, as found for instance in [4,15,24]. As is usual, the theorem is formulated for first-order systems.

Theorem (*Floquet*). Consider the linear homogeneous differential equation

$$y' = A(x)y, \tag{11}$$

for some square matrix $A(x)$ of complex continuous functions such that $A(x + L) = A(x)$. Then any fundamental matrix $\Phi(x)$ of this system may be decomposed as

$$\Phi(x) = \hat{\Phi}(x) e^{Rx}, \tag{12}$$

where $\hat{\Phi}(x + L) = \hat{\Phi}(x)$, $\hat{\Phi}(x)$ is nonsingular and R is a constant matrix.

If we are only interested in solutions that are bounded for all $x \in \mathbb{R}$, Floquet's theorem allows us to greatly reduce the functional form of any such solutions. The behavior as $x \to \pm\infty$ is governed by the exponential factor, since the first factor is periodic. Further, we may use a similarity transformation to reduce R to its Jordan form. Indeed, the differential equation under consideration is linear, and so only a set of fundamental solutions is of interest. We refer to the eigenvalues of R as Floquet exponents. There are three cases:

1. Suppose there exist Floquet exponents r giving rise to non-diagonal blocks in the Jordan form of R. Then the corresponding solutions of the differential equation are algebraically growing. Such Floquet exponents may be ignored since we are only interested in solutions of the differential equations that are bounded for all x, by the definition of the spectrum of the differential operator.
2. Similarly, if there are Floquet exponents with nonzero real part, then these give rise to exponentially growing solutions as $x \to \infty$ ($\Re r > 0$) or $x \to -\infty$ ($\Re r < 0$).
3. It follows that the only solutions of linear homogeneous differential equations with periodic coefficients that are bounded for all $x \in \mathbb{R}$ correspond to purely imaginary Floquet exponents.

Thus, every bounded solution of our fundamental equation (6) is of the form

$$w(x) = e^{i\mu x}\phi(x), \tag{13}$$

with $\phi(x + L) = \phi(x)$ for any fixed λ, and $\mu \in [0, 2\pi/L)$. The factor $e^{i\mu x}$ is referred to as a Floquet multiplier and $i\mu$ is the Floquet exponent mentioned above. In what follows, we loosely refer to μ as the Floquet

exponent. Note that in different areas of science, Floquet theory may be known as monodromy theory (Hamiltonian systems, etc.) or Bloch theory (solid state theory, etc). In the context of Bloch theory the Floquet exponent is often referred to as the quasi-momentum.

Like the coefficient functions, the function $\phi(x)$ may be expanded as a Fourier series with period L as well. For future use, we expand $\phi(x)$ as a Fourier series in x of period PL, with $P \in \mathbb{N}$. This effectively limits the possible values of the Floquet exponent μ: if $P = 1$, at most $\mu \in [0, 2\pi/L)$; if $P = 2$, at most $\mu \in [0, \pi/L)$ is needed, etc. Thus

$$w(x) = \mathrm{e}^{\mathrm{i}\mu x} \sum_{j=-\infty}^{\infty} \hat{\phi}_j \mathrm{e}^{\mathrm{i}2\pi jx/PL} = \sum_{j=-\infty}^{\infty} \hat{\phi}_j \mathrm{e}^{\mathrm{i}x(\mu+2\pi j/PL)}, \tag{14}$$

and

$$\hat{\phi}_j = \frac{1}{PL} \int_{-PL/2}^{PL/2} \phi(x) \mathrm{e}^{-\mathrm{i}2\pi jx/PL} \, \mathrm{d}x, \quad j \in \mathbb{Z} \tag{15}$$

is the jth Fourier coefficient of $\phi(x)$. After we multiply (6) (using (7)) by $\mathrm{e}^{-\mathrm{i}\mu x}$, any term of the resulting equation is periodic. Taking the nth Fourier coefficient of this equation, we obtain

$$\begin{aligned}
\lambda \hat{\phi}_n &= \frac{1}{PL} \int_{-PL/2}^{PL/2} \mathrm{e}^{-\mathrm{i}2\pi nx/PL} \left(\mathrm{e}^{-\mathrm{i}\mu x} \sum_{k=0}^{M} f_k(x) \partial_x^k w(x) \right) \mathrm{d}x \\
&= \sum_{k=0}^{M} \sum_{j=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \hat{f}_{k,j} \hat{\phi}_m \frac{1}{PL} \int_{-PL/2}^{PL/2} \mathrm{e}^{-\mathrm{i}2\pi nx/PL - \mathrm{i}\mu x + \mathrm{i}2\pi jx/L} \left(\partial_x^k \mathrm{e}^{\mathrm{i}\mu x + \mathrm{i}2\pi mx/PL} \right) \mathrm{d}x \\
&= \sum_{k=0}^{M} \sum_{j=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \hat{f}_{k,j} \hat{\phi}_m \left[\mathrm{i} \left(\mu + \frac{2\pi m}{PL} \right) \right]^k \frac{1}{PL} \int_{-PL/2}^{PL/2} \mathrm{e}^{\mathrm{i}2\pi x(-n+jP+m)/PL} \, \mathrm{d}x \\
&= \sum_{k=0}^{M} \sum_{j=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \hat{f}_{k,j} \hat{\phi}_m \left[\mathrm{i} \left(\mu + \frac{2\pi m}{PL} \right) \right]^k \delta_{0,jP-n+m},
\end{aligned} \tag{16}$$

where $\delta_{j,k}$ is the Kronecker delta. Thus, if $n - m$ is divisible by P (i.e., $P|n - m$),

$$\sum_{m=-\infty}^{\infty} \left(\sum_{k=0}^{M} \hat{f}_{k,\frac{n-m}{P}} \left[\mathrm{i} \left(\mu + \frac{2\pi m}{PL} \right) \right]^k \right) \hat{\phi}_m = \lambda \hat{\phi}_n, \quad n \in \mathbb{Z}. \tag{17}$$

This equation can be rewritten as a bi-infinite matrix equation:

$$\hat{\mathscr{L}}(\mu) \hat{\phi} = \lambda \hat{\phi}, \tag{18}$$

with $\hat{\phi} = (\ldots, \hat{\phi}_{-2}, \hat{\phi}_{-1}, \hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2 \ldots)^{\mathrm{T}}$ and where the μ-dependence of $\hat{\mathscr{L}}$ is explicitly indicated. This bi-infinite matrix $\hat{\mathscr{L}}(\mu)$ is defined by

$$\hat{\mathscr{L}}(\mu)_{nm} = \begin{cases} 0 & \text{if } P \nmid n - m, \\ \sum_{k=0}^{M} \hat{f}_{k,\frac{n-m}{P}} \left[\mathrm{i} \left(\mu + \frac{2\pi m}{PL} \right) \right]^k & \text{if } P|n - m, \end{cases} \tag{19}$$

The difference Eq. (17) or the matrix equation (18) are *equivalent* to the original problem (6). *No approximations were introduced to obtain either (17) or (18)*. The linear problem (6) has associated with it its period lattice $\Lambda = \{nL : n \in \mathbb{Z}\}$. As a consequence, the equivalent problem in terms of Fourier series (17) or (18) has associated with it the conjugate lattice $\hat{\Lambda} = \{n2\pi/L : n \in \mathbb{Z}\}$ (remember that the integer parameter P was artificially introduced and is not essential to the problem), which determines the set of possible μ values: $\mu \in [0, 2\pi/PL)$ (incorporating P again).

It is conceivable that problems exist where the difference Eq. (17) is more convenient to analyze than (6), but in this paper it is regarded as a tool to obtain numerical approximations to $\sigma(\mathscr{L})$, the spectrum of (6). To this end, we choose a cutoff N on the number of Fourier modes of the eigenfunctions $\phi(x)$, resulting in a matrix system of dimension $2N + 1$:

$$\hat{\mathscr{L}}_N(\mu)\hat{\phi}_N = \lambda_N\hat{\phi}_N, \tag{20}$$

where the eigenvalues λ_N are approximations, in the sense that all $\lambda_N \in \sigma(\hat{\mathscr{L}}_N(\mu)) \to \lambda \in \sigma(\mathscr{L})$ as $N \to \infty$. Similarly, for any fixed μ, the knowledge of the eigenvectors $\hat{\phi}_N$ allows for the reconstruction of the eigenfunction $w(x)$, using (14). Thus, the definition of convergence used here is the following:

Definition (*Convergence of the spectral approximation*). Any eigenvalue λ_N of $\hat{\mathscr{L}}_N(\mu)$ converges to an element of $\sigma(\mathscr{L})$:

$$\lim_{N\to\infty} d(\lambda_N, \sigma(\mathscr{L})) = 0, \tag{21}$$

where $d(\lambda_N, \sigma(\mathscr{L}))$ denotes the distance of λ_N to the set $\sigma(\mathscr{L})$ [35]. Further, the incorporation of the entire interval of Floquet exponents $\mu \in [0, 2\pi/PL)$ gives rise to the entire spectrum:

$$\lim_{N\to\infty} \bigcup_{\mu\in[0,2\pi/PL)} \sigma(\hat{\mathscr{L}}_N(\mu)) = \sigma(\mathscr{L}). \tag{22}$$

It is not proven here that this definition of convergence holds for the numerical method presented. However, it is clear from Example 2 (see below) that this concept of convergence holds for the case of operators with constant coefficients. Further, our numerical examples of Section 4 point to the same conclusion.

At this point we have all the ingredients for a numerical scheme for computing an approximation of $\sigma(\mathscr{L})$. The steps of the FFHM are:

1. Fix P, so that $\phi(x)$ may be expanded in a Fourier series of period PL.
2. Fix the number of Fourier modes N to be used for the approximation of $\phi(x)$.
3. Fix a sequence of μ-values $\{\mu_1, \ldots, \mu_D\}$, $\mu_k \in [0, 2\pi/PL)$, $k = 1, \ldots, D$.
4. For each $k \in \{1, \ldots, D\}$ compute the eigenvalues (and if so desired, the eigenvectors) of the $(2N+1)$-dimensional matrix equation $\hat{\mathscr{L}}_N(\mu_k)\hat{\phi}_N = \lambda_N\hat{\phi}_N$.
5. The collection of all of these eigenvalues gives an approximation to $\sigma(\mathscr{L})$. A more accurate approximation may be obtained by increasing N. An approximation of the same accuracy but with more spectral elements may be obtained by increasing D while keeping N constant.

The black-box nature of this algorithm is the reason that the FFHM is straightforward to use as the computational engine of SpectrUW [19,20].

It may be appropriate to comment on (19). This formula is meant to illustrate the above-mentioned black-box nature of the FFHM: it provides a prescription by which to fill the matrix $\hat{\mathscr{L}}_N(\mu)$ for any scalar finite-order linear differential operator. However, the formula may not appear as transparent as it should. To clarify its use, we elaborate on it from a different point of view. First, it is clear that every term of (7) corresponds to a term in the summation in (19):

$$f_k(x)\partial_x^k \leftrightarrow \hat{\mathscr{L}}^{(k)}(\mu), \tag{23}$$

such that

$$\hat{\mathscr{L}}(\mu) = \sum_{k=1}^{M} \hat{\mathscr{L}}^{(k)}(\mu). \tag{24}$$

As above, the entries $\hat{\mathscr{L}}^{(k)}(\mu)_{nm}$ of $\hat{\mathscr{L}}^{(k)}(\mu)$ are given by

$$\hat{\mathscr{L}}^{(k)}(\mu)_{nm} = \begin{cases} 0 & \text{if } P\nmid n-m, \\ \hat{f}_{k,\frac{n-m}{P}}\left[i\left(\mu+\frac{2\pi m}{PL}\right)\right]^k & \text{if } P|n-m. \end{cases} \tag{25}$$

Thus, when setting up the FFHM matrix $\hat{\mathscr{L}}_N(\mu)$ for a given linear operator, one proceeds term-by-term. For every term, the following steps are executed:

1. Construct the bi-infinite list of Fourier coefficients of $f_k(x)$: $(\ldots, \hat{f}_{k,-2}, \hat{f}_{k,-1}, \hat{f}_{k,0}, \hat{f}_{k,1}, \hat{f}_{k,2}, \ldots)$.
2. A matrix of dimension $(2N+1)\times(2N+1)$ is filled in, with the following structure:

$$\begin{pmatrix}
* & & & * & & & * & & \\
& * & & & * & & & * & \\
& & * & & & * & & & * \\
* & & & * & & & * & & \\
& * & & & * & & & * & \\
& & * & & & * & & & * \\
* & & & * & & & * & & \\
& * & & & * & & & * & \\
& & * & & & * & & & *
\end{pmatrix}, \tag{26}$$

where * denotes a non-zero entry. Thus, the non-zero matrix entries fill the main diagonal and certain sub- and superdiagonals. Between any two filled diagonals, there are exactly $P-1$ zero diagonals.

3. Every non-zero diagonal is multiplied by one of the Fourier coefficients of $f_k(x)$: the main diagonal by $\hat{f}_{k,0}$, the first non-zero superdiagonal by $\hat{f}_{k,-1}$, the first non-zero subdiagonal by $\hat{f}_{k,1}$, the second non-zero superdiagonal by $\hat{f}_{k,-2}$, and so on. Thus generically $\hat{\mathscr{L}}_N(\mu)$ is a full matrix. In the special case where all the coefficients $f_k(x)$, $k \in \{0,\ldots,M\}$ are trigonometric polynomials, the bi-infinite matrix $\hat{\mathscr{L}}(\mu)$ has a finite number of nonzero super and subdiagonals, thus it is a banded matrix.
4. On these non-zero diagonals, every Fourier coefficient entry is multiplied by $\left[\mathrm{i}\left(\mu+\frac{2\pi m}{PL}\right)\right]^k$, where m is the column in which the entry goes. Note that rows and columns are numbered from $-N$ to N. Thus, the zeroth (middle) column is multiplied by $(\mathrm{i}\mu)^k$. The first column (i.e., the column to the right of the central column) is multiplied by $(\mathrm{i}\mu + \mathrm{i}2\pi/PL)^k$, the minus first column (i.e., the column to the left of the central column) is multiplied by $(\mathrm{i}\mu - \mathrm{i}2\pi/PL)^k$. The second column is multiplied by $(\mathrm{i}\mu + \mathrm{i}4\pi/PL)^k$, and so on.

This way, every single one of the matrices $\hat{\mathscr{L}}^{(k)}(\mu)_N$ is easily constructed, using a diagonal-by-diagonal construction, followed by multiplying columns by the appropriate Floquet factor. We illustrate this using an example:

Example 1. Consider the term $f(x)\partial_x^7$, where $f(x) = \sum_{n=-\infty}^{\infty} \mathrm{e}^{\mathrm{i}nx}/2^{n^2}$. Thus we have $L = 2\pi$, and the list of Fourier coefficients of $f(x)$ is given by $(\ldots, \hat{f}_{k,-2}, \hat{f}_{k,-1}, \hat{f}_{k,0}, \hat{f}_{k,1}, \hat{f}_{k,2}, \ldots) = (\ldots, 1/2^4, 1/2^2, 1, 1/2^2, 1/2^4, \ldots)$. Let us choose $P = 3$, so that all filled diagonals of the FFHM matrix corresponding to $f(x)\partial_x^7$ are separated by $P-1 = 2$ zero diagonals. For simplicity, let $N = 4$, so that a 9×9 matrix is constructed, as shown below.

	-4	-3	-2	-1	0	1	2	3	4	
-4	$(i\mu - 4i/3)^7$	0	0	$\frac{(i\mu - i/3)^7}{2}$	0	0	$\frac{(i\mu+2i/3)^7}{2^4}$	0	0	
-3	0	$(i\mu - 3i/3)^7$	0	0	$\frac{(i\mu)^7}{2}$	0	0	$\frac{(i\mu+3i/3)^7}{2^4}$	0	
-2	0	0	$(i\mu - 2i/3)^7$	0	0	$\frac{(i\mu+i/3)^7}{2}$	0	0	$\frac{(i\mu+4i/3)^7}{2^4}$	
-1	$\frac{(i\mu-4i/3)^7}{2}$	0	0	$(i\mu - i/3)^7$	0	0	$\frac{(i\mu+2i/3)^7}{2}$	0	0	$\nwarrow$
0	0	$\frac{(i\mu-3i/3)^7}{2}$	0	0	$(i\mu)^7$	0	0	$\frac{(i\mu+3i/3)^7}{2}$	0	$\hat{f}_{-2}$
1	0	0	$\frac{(i\mu-2i/3)^7}{2}$	0	0	$(i\mu + i/3)^7$	0	0	$\frac{(i\mu+4i/3)^7}{2}$	
2	$\frac{(i\mu-4i/3)^7}{2^4}$	0	0	$\frac{(i\mu - i/3)^7}{2}$	0	0	$(i\mu + 2i/3)^7$	0	0	$\nwarrow$
3	0	$\frac{(i\mu-3i/3)^7}{2^4}$	0	0	$\frac{(i\mu)^7}{2}$	0	0	$(i\mu + 3i/3)^7$	0	$\hat{f}_{-1}$
4	0	0	$\frac{(i\mu-2i/3)^7}{2^4}$	0	0	$\frac{(i\mu+i/3)^7}{2}$	0	0	$(i\mu + 4i/3)^7$	
				$\nwarrow$ $\hat{f}_2$			$\nwarrow$ $\hat{f}_1$			$\nwarrow$ $\hat{f}_0$
	$\uparrow$ $(i\mu - 4i/3)^7$	$\uparrow$ $(i\mu - 3i/3)^7$	$\uparrow$ $(i\mu - 2i/3)^7$	$\uparrow$ $(i\mu - i/3)^7$	$\uparrow$ $(i\mu)^7$	$\uparrow$ $(i\mu + i/3)^7$	$\uparrow$ $(i\mu + 2i/3)^7$	$\uparrow$ $(i\mu + 3i/3)^7$	$\uparrow$ $(i\mu + 4i/3)^7$	

This example demonstrates how the FFHM matrices are built up, with Fourier coefficients filling in along (sub, super) diagonals, separated by $P-1$ zero diagonals. The Floquet factors $(\mathrm{i}\mu + 2\pi m/PL)$ fill in along columns, as demonstrated above. For clarity, rows and columns have been labeled as they should, from $-N$ to N, resulting in the FFHM matrix in the box.

The following example illustrates that the FFHM is exact on the class of linear operators with constant coefficients. Further, it shows that the spectrum obtained does not consist of a discrete set of points (as would be the case when using periodic eigenfunctions), but rather of a set with continuous components in the complex plane.

Example 2. Consider the important special case of the operator $\mathscr{L}$ having constant coefficients:

$$\mathscr{L} = \sum_{k=0}^{M} a_k \partial_x^k, \quad a_k \in \mathbb{C}, \ \ k = 0, 1, \ldots, M. \tag{27}$$

In this case $L = 0$ and it suffices to choose $N = 0$, so that

$$w(x) = \mathrm{e}^{\mathrm{i}\mu x}\hat{\phi}_0, \quad \mu \in \mathbb{R}. \tag{28}$$

we obtain (using a 1×1 matrix eigenvalue equation)

$$\lambda = \sum_{k=0}^{M} a_k (i\mu)^k, \quad \mu \in \mathbb{R}, \tag{29}$$

which is the correct analytical expression for the locus of $\sigma(\mathscr{L})$. Note that (29) is the dispersion relationship for the linear partial differential equation $u_t = \mathscr{L}u$ (see Appendix of [3], for instance).

Remarks

- *Historical context.* Hill [28] used virtually the same method to examine the spectrum of what is now known as Hill's equation, seven years before the publication of Floquet's theorem [24] (published at his own expense in 1877, and reprinted in [28] nine years later). For computing purposes Hill used $N = 1$, resulting in a 3×3 matrix, $P = 1$, with $\mu = 0$ and $\mu = 1$, corresponding to the edges of the spectrum. Given how crude Hill's approximation was, the results obtained were remarkably accurate. Hill did consider the infinite-dimensional difference system as well, leading to "Hill's infinite determinant", obtaining some theoretical results from it. Since then Hill's method has been used sporadically. Analysts were at odds with the convergence issues of the infinite-dimensional determinant approach, and the generality of Hill's approach for numerical purposes went unrecognized. A good overview of Hill's work may be found in Whittaker and Watson's classic text [50]. There are many instances where Hill's approach was used to numerically approximate the spectrum of a linear problem. The method is somewhat popular in the area of fluid stability, e.g. [8,26,40,16], but its connection to Hill's work and its generality are unacknowledged. The FFHM was rediscovered in the Russian literature by Pavlenko and Petviashvili [43], and successively used by others [23]. These authors appear unaware of Hill's work, but Pavlenko and Petviashvili hint at the generality of the method.
- *Function space and boundary conditions.* Typical eigenfunctions of the form (14) are quasi-periodic due to the presence of two typically non-commensurate periods PL and $2\pi/\mu$. Nevertheless, the eigenfunctions are square-integrable on $[-PL/2, PL/2]$, since the $2\pi/\mu$-periodicity disappears from the integration:

$$\frac{1}{PL}\int_{-PL/2}^{PL/2} |w|^2 \,\mathrm{d}x = \sum_{j=-\infty}^{\infty} |\hat{\psi}_j|^2 < \infty. \tag{30}$$

  In addition, all these eigenfunctions are bounded on the real line, by construction.The boundary conditions can be made more explicit at this point. The eigenfunctions satisfy

$$w(x + PL) = \mathrm{e}^{\mathrm{i}\mu PL} w(x), \tag{31}$$

  which follows from (14) immediately. Such boundary conditions are known as Bloch boundary conditions [5]. Due to the attention devoted to second-order self-adjoint problems, it has been customary to focus especially on periodic (with period L) and anti-periodic (with period $2L$) eigenfunctions, which correspond to $\mu = 0$ with $P = 2$. This results in an infinite, but discrete set, which is a subset of the spectrum we wish to capture. By considering eigenfunctions of larger periods (for instance by choosing $P > 2$), a larger discrete subset is found. This subset provides a point covering of higher density of the spectrum we wish to obtain than the first subset. As the period of the eigenfunctions approaches infinity, an increasingly larger discrete

subset of the spectrum we consider is found. That spectrum may be obtained as the limit spectrum of these discrete spectra, for sequences of eigenfunctions with increasing periods. The linear operator $\mathscr{L}$ with these boundary conditions is referred to as the maximal extension of $\mathscr{L}$ [33].

- *Eigenfunctions with larger periods.* The integer parameter P was introduced artifically. Often it is chosen to be 2, so that the functions $\phi(x)$ are periodic with period $2L$, i.e., $\phi(x)$ is periodic or antiperiodic with the same period as the coefficients $f_k(x)$, $k = 1,\ldots,M$. Thus, even when choosing $D = 1$ all periodic or anti-periodic eigenfunctions of (6) are obtained. This allows for a more direct comparison with classical monodromy methods, such as those used for instance in [13]. Moreover, often the elements of the spectrum corresponding to periodic and anti-periodic eigenfunctions are good approximations to the extremal parts of the spectrum, e.g., the elements with largest or smallest real parts, etc. For second-order self-adjoint problems, it is known that these spectral elements give the edge elements of the spectrum exactly [38]. Further, for a $(2N+1)\times(2N+1)$ FFHM matrix, used to determine the first $2N+1$ Fourier coefficients of the eigenfunctions, choosing $P = 2$ results in the same number of Fourier coefficients being required of the input functions $f_N(x),\ldots,f_0(x)$. In other words, the same amount of information is used as input for the problem as we are aiming to find as output. As the first example above illustrates, choosing $P > 2$, we are trying to recover more information about the eigenfunctions than is used as input for the coefficient functions. In Example 1, the eigenvectors are 9 dimensional, resulting in 9 Fourier coefficients of the eigenfunctions. On the other other hand, only 5 Fourier coefficients of $f(x)$ are used to construct the FFHM matrix. Conversely, with $P = 1$, $4N+1$ Fourier modes of the coefficient functions are used to compute only $2N+1$ Fourier coefficients of the eigenfunctions. Lastly, as the examples below illustrate, different choices of P may lead to faster computations.
- *Choosing Floquet exponents.* For convenience, the sequence $\{\mu_1,\ldots,\mu_D\}$, $\mu_k \in [0, 2\pi/PL)$, $k = 1,\ldots,D$ is usually chosen equally spaced, so that $\mu_k = 2\pi(k-1)/PLD$, $k = 1,\ldots,D$. If so chosen, the functions $w_k(x) = \mathrm{e}^{\mathrm{i}\mu_k x}\phi(x)$ are periodic with period PLD. Truly quasi-periodic functions $w(x)$ may be obtained by including μ-values that are not rationally related to $2\pi/L$. It should be remarked that in applications it is often beneficial to choose $\mu \in [-\pi/PL, \pi/PL)$, for symmetry reasons. This is especially important for computations where N is small (say $N \approx 10$ or less). In those cases, the asymmetric truncation of the bi-infinite difference equation may result in errors which are avoided by performing this asymmetric truncation over a symmetric range of μ-values. If values of μ are chosen outside of the basis cell of the conjugate lattice, i.e., $\mu \notin [0, 2\pi/PL)$, then an equivalent difference Eq. (17) is obtained, but for a shifted bi-infinite eigenvector $\hat{\phi}$. Since n runs over all integers, this results in the same eigenvalues, but with shifted eigenvectors. Thus no new elements of the spectrum are obtained this way, and μ may be restricted to the basis cell of the conjugate lattice, as stated.
- *Heisenberg's matrix formulation* of non-relativistic quantum mechanics [39] is obtained by reformulating the Schrödinger equation in its so-called momentum representation. This is nothing but a transformation to Fourier variables. When operating on the whole real line, it is clear that applying Hill's method is just this transformation from the Schrödinger operator formulation to the Heisenberg matrix formulation. For problems with periodic coefficients, Floquet or Bloch theory has to be taken into account in order to have the Heisenberg and Schrödinger representations be fully equivalent.
- *The effect of truncating the Fourier series* to include a finite number of modes is to not approximate spectral elements in parts of the spectrum that are far removed from the origin of the spectral plane.

3.1. Fourier transforms vs. Fourier series

Taking a Fourier transform of (6) results in the difference equation

$$\lambda\hat{w}(\kappa) = \sum_{k=0}^{M}\sum_{j=-\infty}^{\infty}\hat{f}_{k,j}\left[\mathrm{i}\left(\kappa - \frac{2\pi j}{L}\right)\right]^k \hat{w}(\kappa - 2\pi j/L), \tag{32}$$

where $\hat{w} = \int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}\kappa x}w(x)\,\mathrm{d}x$. This difference equation is clearly defined on the conjugate lattice $\hat{\Lambda} = \{n2\pi/L : n \in \mathbb{Z}\}$. Further, in the setting of Fourier transforms, κ is not restricted to integer values, and may thus

assume all values in $[0, 2\pi/L)$, resulting in a one-parameter family of grids, all shifted versions of the elementary grid $(2\pi/L)\mathbb{Z}$ by shifting by $\mu = k \bmod 2\pi/L$. Then μ plays the role of a Floquet exponent. *This approach may be considered a direct proof of that part of the classical Floquet theorem that deals with bounded solutions*. As before, a numerical method for the approximation of $\sigma(\mathscr{L})$ may be obtained from the difference Eq. (32) by truncating the size of the bi-infinite eigenvectors. This scheme has the same features advertised at the beginning of this section. Although appearing similar, it is not immediately evident that the difference equations (17) and (32) are equivalent. To establish this, we return to the form of the eigenfunction given by the Floquet theorem: $w(x) = e^{i\mu x}\phi(x)$. Equivalently, substituting the Fourier series of $\phi(x)$ results in (14). This shows that the Fourier transform of $w(x)$ resides on a shifted delta-function measure. Explicitly, we obtain

$$\hat{w}(\kappa) = 2\pi \sum_{j=-\infty}^{\infty} \hat{w}_j \delta(\mu - \kappa + 2\pi j/PL). \tag{33}$$

Substituting this result in (32) results in (17), establishing the equivalence of the two approaches. There are settings where the use of Fourier transforms is advantageous over that of Fourier series. In (7), we restricted ourselves to differential operators. When using the Fourier transform approach, it is possible in some cases to incorporate integration operators (not always, as this leads to division by the Fourier variable), and in all cases to incorporate convolution operators. Either one leads to significant complications of finite difference methods. The FFHM also runs into problems, and we can no longer fall back on the Floquet theorem to justify its use.

3.2. Computational cost

One of the main features of these FFHMs is the possibility to obtain a better, more uniform approximation of $\sigma(\mathscr{L})$ by increasing D. Thus instead of computing eigenvalues and eigenvectors of a larger matrix, we compute eigenvalues and eigenvectors of matrices of the same size, but of more of them. This results in a significant reduction of computation time required for the approximation of $\sigma(\mathscr{L})$: The cost for the computation of $D(2N+1)$ eigenvalues and eigenvectors (accuracy determined by N), using the QR algorithm, is proportional to $D(2N+1)^3$. Using a standard finite difference method, this cost would be proportional to $D^3(2N+1)^3$, thus finite difference methods are a factor D^2 more expensive. This quick comparison does not take into account the actual location of the eigenvalues computed, which is more uniform for the FFHMs than for the finite difference methods. Furthermore, in order to obtain the same accuracy as the FFHMs, finite difference methods usually require a higher value of N, as can be seen from Table 1.

3.3. Incorporating Floquet theory in finite difference methods (FDMs)

The method of finite differences is the most commonly used technique for computing an approximation to the spectrum of a given linear operator. FDMs [6] are black-box methods in that they use little of the specifics of the problem at hand. It is possible and even likely that for any given problem a more tailored approach will result in a "better" (faster, more accurate, etc.) answer. The main advantage of FDMs is their applicability to a large class of problems, without their having to involve the mathematical details of the specific problem.

Table 1
Comparing the FDM (2nd and 4th order) and FFHM for computing the lowest eigenvalue $\tilde{a} \approx -1.513956885056448$ of (35) with $q = 2$

	Accuracy: 10^{-3}		Accuracy: 10^{-6}		Accuracy: 10^{-9}	
	Matrix size	CPU time	Matrix size	CPU time	Matrix size	CPU time
FFHM ($P=1$)	5	0.5	7	0.6	9	0.5
FFHM ($P=2$)	7	1	13	1	17	1
FFHM ($P=4$)	9	1.5	25	2.6	33	3.3
FDM2	239	1075	8000	5.5E6	N/A	N/A
FDM4	52	25	293	1100	1630	1.5E5

All CPU times are given relative to those of the FFHM with $P = 2$, which are 0.0004, 0.0007 and 0.0010 s, respectively. Here N/A means "not applicable" and denotes a numerical run requiring a matrix size beyond Matlab7s capabilities.

In addition, finite-difference methods are relatively straightforward to implement. Further, the local nature of FDMs leads to banded matrices, with a few additional corner elements, due to the boundary conditions. Thus the matrices obtained from the FDM are often sparse, which does offer some advantages for the computation of the eigenvalues [47].

FDMs are based on Taylor series expansions, and as such allow for explicit estimates of the computational error. For a given $h = \Delta x = L/N$ (stepsize), N eigenvalues and eigenvectors are computed. In order to approximate more elements of $\sigma(\mathscr{L})$, one increases N, which at once results in more accurate approximate spectral elements, and more of them. These are obtained at the cost of increased computational time, proportional to $(nN)^3$, if the QR algorithm is used for the eigenvalue/eigenvector solver [47].

It is possible to augment FDMs by taking Floquet theory into account. Using FDMS with periodic boundary conditions as a starting point, we illustrate below how FDMs may be modified rather trivially to obtain more uniform approximations of $\sigma(\mathscr{L})$, at a low implementation cost and a computational increase *linear* in the number of Floquet shifts D. At that point, the most important remaining difference between FDMs and FFHMs is the higher (spectral) accuracy of the latter.

We show this using the FDM approximation of fourth order to the second-derivative operator. Instead of the more standard periodic boundary conditions $y_{N+k} = y_k$ with $k = -2,-1,0,1,2$, we may use Floquet or Bloch boundary conditions: from $y(x) = \mathrm{e}^{\mathrm{i}\mu x}\phi(x)$, we obtain $y(x+L) = \mathrm{e}^{\mathrm{i}\mu L}y(x)$. Thus, we may impose $y_{N+k} = \mathrm{e}^{\mathrm{i}\theta}y_k$ for $k = -2,-1,0,1,2$, for $\theta \in [0,2\pi)$ which corresponds to $\mu \in [0,2\pi/L)$. Then the FDM to approximate ∂^2 is

$$D_2^{(4,h)} = \frac{1}{12h^2}\begin{pmatrix} -30 & 16 & -1 & 0 & \cdots & \cdots & 0 & -\mathrm{e}^{-\mathrm{i}\theta} & 16\mathrm{e}^{-\mathrm{i}\theta} \\ 16 & -30 & 16 & -1 & 0 & \cdots & \cdots & 0 & -\mathrm{e}^{-\mathrm{i}\theta} \\ -1 & 16 & -30 & 16 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 16 & -30 & 16 & -1 & 0 & \cdots & \cdots \\ & & & \ddots & \ddots & \ddots & & & \\ \cdots & \cdots & 0 & -1 & 16 & -30 & 16 & -1 & 0 \\ 0 & \cdots & \cdots & 0 & -1 & 16 & -30 & 16 & -1 \\ -\mathrm{e}^{\mathrm{i}\theta} & 0 & \cdots & \cdots & 0 & -1 & 16 & -30 & 16 \\ 16\mathrm{e}^{\mathrm{i}\theta} & -\mathrm{e}^{\mathrm{i}\theta} & 0 & \cdots & \cdots & 0 & -1 & 16 & 30 \end{pmatrix}, \tag{34}$$

which should be contrasted with the same matrix obtained using periodic boundary conditions, corresponding to $\theta \equiv 0$. As a general rule, every non-zero element in the upper right corner is multiplied by $\mathrm{e}^{-\mathrm{i}\theta}$ and the corresponding elements in the lower left corner are multiplied by $\mathrm{e}^{\mathrm{i}\theta}$. If the matrix $D_2^{(4,h)}$ with $\theta \equiv 0$ is self-adjoint, so is the one above. This way, using all $\theta \in [0,2\pi)$, an approximation to the entire spectrum $\sigma(\mathscr{L})$ may be obtained. Then more equally accurate elements of this spectrum are obtained by varying θ, without decreasing h. Thus, as in the FFHM, more elements of the spectrum are obtained at the cost of computing eigenvalues of matrices of the same size. For instance, to approximate spectral elements corresponding to both periodic and antiperiodic eigenfunctions, the above matrix is used twice: once with $\theta = 0$ (periodic) and once with $\theta = \pi$ (antiperiodic).

4. Examples

In this section we consider several examples. These examples were chosen because their spectra are well-understood. Thus we can demonstrate the virtues and pitfalls of each method without getting tangled up in the intricacies of the specific problem we are attempting to solve. All numerical runs were done in Matlab, on a Mac G4 single processor (1G Hz) laptop with 1GB memory.

4.1. Mathieu equation

The Mathieu functions [1] are the bounded solutions of the Mathieu equation

$$y'' + (a - 2q\cos(2x))y = 0 \iff -y'' + 2q\cos(2x)y = ay, \tag{35}$$

one of the classical linear differential equations defining a class of special functions. The equation originates from the Helmholtz equation through the use of separation of variables. Here we are interested in determining all a values for which bounded solutions to (35) exist. Many texts on perturbation methods (e.g. [7,34]) use this equation as one of their prototypical examples. One of their goals is to determine the edges of the spectrum for varying q. The operator $-\partial_x^2 + 2q\cos(2x)$ is self-adjoint, thus the spectrum is confined to the real line. Also, since the operator $-\partial_x^2$ is positive and $2q\cos(2x)$ is a bounded operator (for fixed q), the spectrum of (35) is bounded from below, and extends to infinity. Thus, as announced, the spectrum of this problem is well-understood. We illustrate here how an accurate approximation to the spectrum may be obtained using either the FDM or the FFHM. Both methods are compared using different sets of criteria.

Using (14) with $L = \pi$ and equating the coefficients of different harmonics results in the difference equation

$$q\hat{\phi}_{n-P} + \left(\mu + \frac{2n}{P}\right)^2 \hat{\phi}_n + q\hat{\phi}_{n+P} = a\hat{\phi}_n, \quad n \in \mathbb{Z}, \tag{36}$$

with $\mu \in [0, 2/P)$. Because of the symmetry of the coefficient function of (35), it is possible to reduce the list of possible μ values further. Since we wish to emphasize the black-box qualities of the FFHM this is not pursued here.

Restricting the above equation to $n = -N, \ldots, N$ results in a class of matrix equation (20), one for each choice of P. Here $\hat{\mathscr{L}}_N$ is of size $(2N+1)\times(2N+1)$. Its diagonal is $((\mu - N/P)^2, \ldots, (\mu - 2/P)^2, \mu^2, (\mu + 2/P)^2, \ldots, (\mu + N/P)^2)$. Its Pth sub- and super diagonals consist of qs. All other elements are zero. This matrix is self-adjoint for any choice of $\mu \in [0, 2/P)$, and any choice of $q \in \mathbb{R}$, reflecting the self-adjointness of the original problem.

Consider the truncated matrix equation with D equispaced μ-values: $\mu_k = 2(k-1)/P$, $k = 1, \ldots, D$. Let $(a_1, a_2, \ldots, a_{D(2N+1)})$ be the ordered list of eigenvalues of the D linear problems associated with $\hat{\mathscr{L}}_N$ for these different μ values. Then $a_{k+1} \geqslant a_k$, for $k = 1, \ldots, D(2N+1) - 1$.

First, we compute approximations to the lowest eigenvalue $\tilde{a}$ of (35). We choose $q = 2$, then $\tilde{a} \approx -1.513956885056448$. The results are displayed in Table 1. For all runs, $D = 1$ and $\mu = 0$, thus we are computing eigenvalues corresponding to eigenfunctions with period πP. It is known [38] that for (35) the endpoints of the spectral bands correspond to periodic or anti-periodic (periodic with twice the period) eigenfunctions. Thus the FFHM and its Fourier transform analogue are ideally suited to the task of computing $\tilde{a}$ with high accuracy. Moreover, as observed from Table 1, a linear increase in the number of Fourier modes results in an exponential improvement in the accuracy of $\tilde{a}$, as is expected of a spectral method. A comparable accuracy improvement using the FDM requires an exponential size increase of the FDM matrix. It is important to notice the dependence on P of the computational time of the Fourier methods.

A second comparison focuses on entire components of the spectrum, as opposed to a single element of it. Define the numerical spectral deficit of a spectral band σ_B to be

$$\delta_B = \frac{\max\limits_{a_k, a_{k+1} \in \sigma_B} (a_{k+1} - a_k)}{|\sigma_B|}. \tag{37}$$

In other words, δ_B is the fraction of the maximal distance between any two successive approximations to spectral elements in σ_B to the length of the spectral band σ_B. Thus as $\delta_B \to 0$, a uniformly better approximation to σ_B is obtained. It follows from Floquet theory applied to the Matthieu equation [38] that $\delta_B \to 0$ simultaneously across all spectral bands where approximations to spectral elements are computed. Thus, δ_B has a constant value across all spectral bands that are approximated, and is zero for all others. As N is increased, the number of bands that are approximated by the FFHM increases. In Table 2, a comparison is shown for the FFHM and the FDM (using the second band). For all FFHM runs, the N value from Table 1 is used for which $\tilde{a}$ is approximated with an accuracy of 10^{-3}. Then D is increased to obtain more spectral element approximations. This was also done for the FDM (4th order), as explained at the end of the last section. All these results are contrasted with the 4th order FDM without using Floquet theory. One should notice that the CPU time is inversely proportional to the required δ_B. This is a vast improvement over regular FDMs without Floquet theory. Except for the largest allowed spectral deficit, these methods require matrixes of a size larger than what Matlab7 can handle (about 8192×8192). For the most efficient FFHM with $P = 2$,

Table 2
Comparing the FDM (4th order) and FFHM for computing a uniform approximation to the second spectral band of (35) with $q = 2$

	$\delta_2 = 0.25$		$\delta_2 = 0.025$		$\delta_2 = 0.0025$	
	D	CPU time	D	CPU time	D	CPU time
FFHM ($P = 1$)	12	0.4	124	1	1190	1.3
FFHM ($P = 2$)	6	1	62	1	590	1
FFHM ($P = 4$)	3	2	31	1.3	295	1.3
FDM4 matrix size=1172	1	5.4E5	N/A	N/A	N/A	N/A
FDM4, $D > 1$	4	8100	32	2100	310	1700

All CPU times are given relative to those of the FFHM with $P = 2$, which are 0.001, 0.03 and 0.35 s, respectively. Here N/A means "not applicable" and denotes a numerical run that was not done due to lack of memory on our platform.

the matrix size is 7, which is a tiny matrix for Matlab computations. Smaller δ_B does not increase this matrix size (until $|\sigma_B|\delta_B$ becomes comparable to the allowed error on $\tilde{a}$), but rather requires the consideration of more matrices of this same size.

Fig. 1 illustrates the effect of increased δ (theoretically independent of which band is being considered), to obtain approximations to entire parts of the spectrum, as opposed to isolated points of it. Lastly, we regenerate the famous picture of the spectrum of the Mathieu equation (35) for varying q, see [7,1]. This picture displays a vs. q, for $0 \leqslant q \leqslant 15$, using 300 different equally spaced q values. The same matrix size (25×25) was used for all q values, so as to automate the numerics more easily. Similarly, $D = 62$ for all q values. Fig. 2 was generated in 25.5 s, using the FFHM with $P = 2$. Any point in Fig. 2 has an absolute error of at most 10^{-3}.

4.2. Hill's equation with a finite number of gaps

4.2.1. The periodic case

As our next example, we consider two Hill equations with a finite number of gaps:

$$L_-(k)u = -u'' + (2k^2\mathrm{sn}^2(x,k) - k^2)u = \lambda u, \tag{38a}$$

$$L_+(k)v = -v'' + (6k^2\mathrm{sn}^2(x,k) - 4 - k^2)v = \lambda v. \tag{38b}$$

Here $\mathrm{sn}(x,k)$ denotes the Jacobi sine function, limiting to $\sin(x)$ as the elliptic modulus approaches 0, and limiting to $\tanh(x)$ as the elliptic modulus approaches 1. Further, $\mathrm{sn}(x,k)$ resp. $\mathrm{sn}^2(x,k)$ is periodic with period $4K(k)$ resp. $2K(k)$, where $K(k) = \int_0^{\pi/2}(1 - k^2\sin^2 x)^{-1/2}\,\mathrm{d}x$ is the complete elliptic integral of the first kind

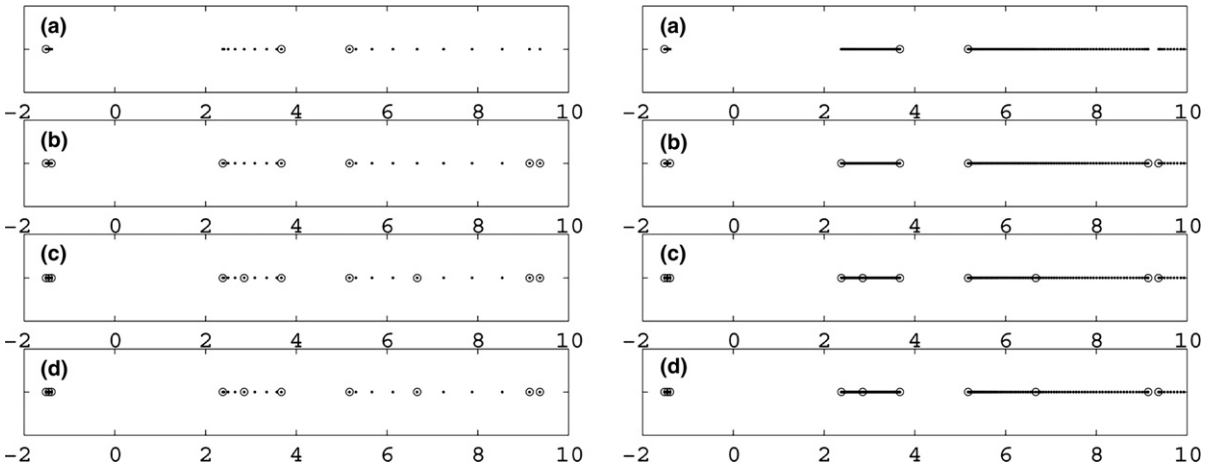


Fig. 1. Numerical approximations of the spectrum for the Mathieu equation (35) with $q = 2$. For the figure on the left $\delta_B = 0.25$, while $\delta_B = 0.025$ on the right. Four different methods were used, all approximating $\tilde{a}$ with an accuracy of 10^{-6}. For (a)–(c) the FFHM was used with $P = 1$, $P = 2$ and $P = 4$, respectively. For (d) the FDM4 was used. In these figures, approximate spectral elements computed with $D = 1$ are indicated by circles. The a-value is real and is shown on the horizontal axis.

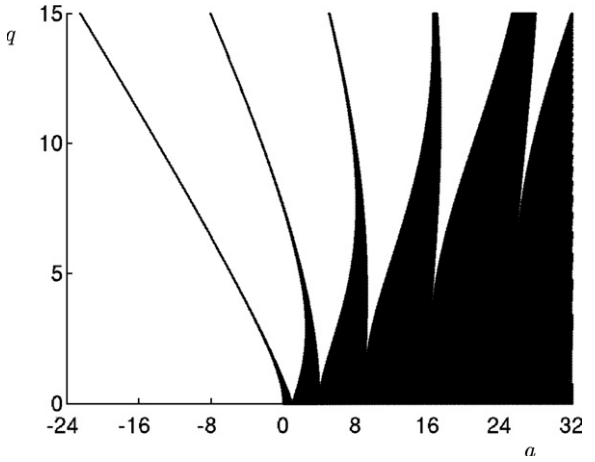


Fig. 2. The spectrum for the Mathieu equation (35), using the FFHM with $P = 2$. Every black point is an approximate point of the spectrum. No filling in of spectral regions was done. All plotted points were the result of a computation.

[50]. Finally, $L_-(k)$ and $L_+(k)$ are the linear Schrödinger operators determined by the left-hand sides of these two Hill equations. It is known (see e.g., Whittaker and Watson [50]) that all equations of the form (38a) and (38b) give rise to a real spectrum (due to the equation being self adjoint), with a spectrum that is bound from below, but not from above. Further, there are a finite number of finite-sized gaps in the spectrum to the right of the lowest eigenvalue if and only if the coefficient of $2k^2\mathrm{sn}^2(x,k)$ is a triangular number, i.e., a number of the form $n(n+1)/2$, n a positive integer. The number of spectral gaps is exactly n. For (38a), $n = 1$, while for (38b) $n = 2$. The special case of Hill's equation considered here is also known as the Lamé equation [30]. All other coefficients in (38a) and (38b) are chosen for later convenience. The spectra of these equations and all their eigenfunctions are known exactly. The spectra are drawn in Fig. 3.

The Fourier series of $sn^2(x,k)$ is given by

$$\mathrm{sn}^2(x,k) = \frac{1}{k^2}\left(1 - \frac{E(k)}{K(k)}\right) - \frac{2\pi^2}{k^2K^2(k)}\sum_{n=1}^{\infty}\frac{nq^n}{1-q^{2n}}\cos\left(\frac{n\pi x}{K(k)}\right), \tag{39}$$

where $q = \mathrm{e}^{-\pi K(k')/K(k)}$, $k' = \sqrt{1-k^2}$, and $E(k) = \int_0^{\pi/2}\sqrt{1-k^2\sin^2 x}\,\mathrm{d}x$ is the complete integral of the second kind. Note that this Fourier series appears to be an obscure result, not even mentioned in [25]. An erroneous Fourier series for $\mathrm{sn}^2(x,k)$ is mentioned in [9]. Jacobi published a correct one in his treatise on elliptic functions [31]. Since the Fourier series of $\mathrm{sn}^2(x,k)$ does not terminate, the matrices resulting from the FFHM for (38a) and (38b) are not banded.

These cases of Hill's equation with their analytically known spectra are excellent benchmarks for the numerical methods discussed in this paper. Fig. 4 illustrates the exponential convergence rate of the FFHM, compared to the finite-order convergence rate of the FDM4. Another advantage of the FFHM is the ease with which eigenfunctions may be reconstructed, as the eigenvectors of the eigenvalue problem solved give truncated vectors of Fourier coefficients of the eigenfunctions (possibly after multiplication by the appropriate $\mathrm{e}^{\mathrm{i}\mu x}$). There is a single eigenfunction of (38b) corresponding to $\lambda = -3(1-k^2)$, namely (up to a multiplicative constant)

$$y = \mathrm{sn}(x,k)\mathrm{dn}(x,k). \tag{40}$$

The numerical eigenfunction is easily normalized by requiring that $y(K(k)) = k'$. Fig. 5a displays the numerical approximation to the eigenfunction for $k = 0.9$, while Fig. 5b shows the numerically computed L^2-norm of the

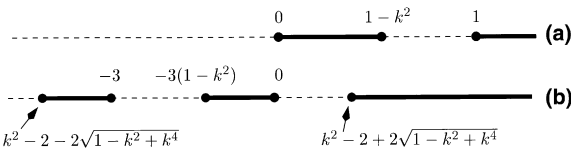


Fig. 3. The spectra of the two Hill equations (38a) and (38b).

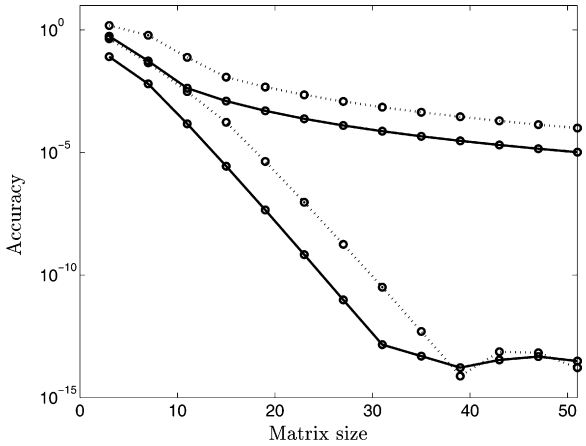


Fig. 4. The convergence of the FFHM (two bottom curves) and of the FDM4 (two top curves). The points on the solid curves show the accuracy of numerically computing the lowest eigenvalue of (38a). The points on the dotted curves show the same for (38b). These computations demonstrate the claimed exponential convergence rate of the FFHM, compared to the power-law (4th order here) convergence for FDMs. Using the FFHM with double-precision arithmetic, it is pointless to proceed beyond matrices of size 35 (i.e., 17 Fourier modes), as the obtained accuracy is on the level of machine round-off error.

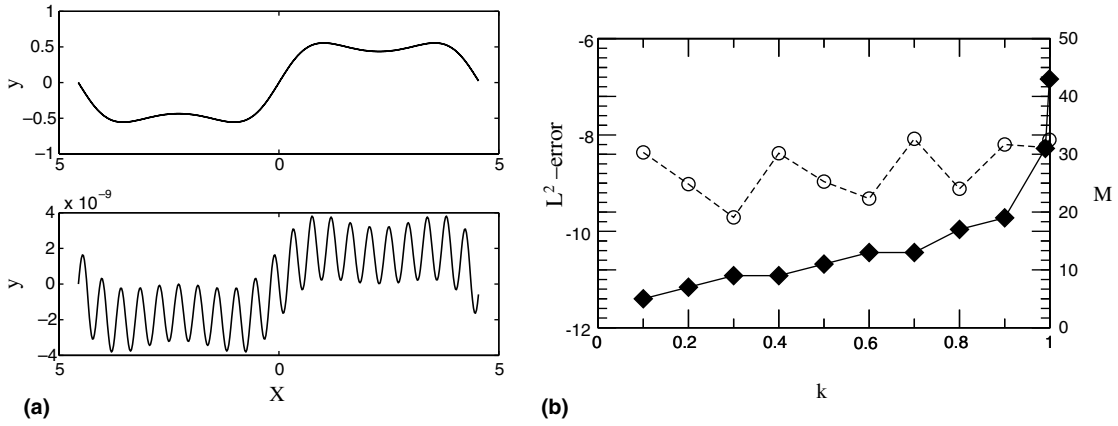


Fig. 5. The numerical approximation using FFHM of the eigenfunction (40) (a, top), and the pointwise error of this approximation (a, bottom). Also shown (b) are the varying matrix sizes $M = 2N + 1$ as a function of the elliptic modulus k, for $k = 0.1$ to $k = 0.99$, required to maintain an L^2-error $< 10^{-8}$ on the numerical approximation of the eigenfunction (40).

difference between (40) and its approximation using the FFHM as a function of k, as well as the matrix sizes giving rise to this difference. By the nature of its construction, the approximation of the eigenfunction is given by a trigonometric polynomial and is uniformly valid as a function of x. This should be contrasted with the piecewise in x approximation obtained from the FDM4.

4.2.2. The soliton limit

The FFHM is ideally suited for problems with periodic coefficients. In problems (38a) and (38b), the period of the coefficients increases as $k \to 1$, since $\lim_{k\to 1} K(k) = \infty$. In this limit, the two Hill equations become

$$L_-(1)u = -u'' + (1 - 2\mathrm{sech}^2 x)u = \lambda u, \tag{41a}$$

$$L_+(1)v = -v'' + (1 - 6\mathrm{sech}^2 x)v = \lambda v. \tag{41b}$$

The spectrum of the first equation consists of a double eigenvalue at $\lambda = 0$ and a continuous band $\lambda \in [1, \infty)$. Similarly, the spectrum of the second equation is $\lambda \in [-3] \cup [0] \cup [1, \infty)$. This limit of the periodic problem will be referred to as the soliton limit. There are two ways to approach these spectral problems using the FFHM:

- *Large k limit*. Consider (38a) and (38b) for large (i.e., close to 1) values of k. Here close to 1 should be on the order of $k = 0.9999$ or more. Then $2K(0.9999) = 11.29$, and the functional form of $\mathrm{sn}^2(x,k)$ starts to resemble $\mathrm{sech}^2 x$. In this case reasonable accuracy requires many terms of the series (39). For instance, $k = 0.9999$ requires $N = 25$, resulting in matrices of size 51×51. For numerical purposes this is still a small matrix, especially when compared to the matrix size that would be required using equally spaced FDMs, in order to discretize a periodic function on an interval of length 11.29 with high accuracy. Table 1 gives an idea of the order of magnitude of the matrix sizes and the CPU times that FDMs would require to obtain comparable results.
- *Soliton cutoff*. A second approach is to consider (41a) and (41b) directly, but cutoff the exponential tails of the coefficients. Then we may consider a periodic extension of the resulting problem. This extension will have an exponentially small mismatch at the edges. The advantage of this approach is that large periods are easily achieved. The Fourier coefficients of a truncated $\mathrm{sech}^2 x$ give rise to integrals that cannot be evaluated analytically, but their high-accuracy numerical evaluation poses no problem. This second approach is the one of choice when handed a typical problem defined on the whole real line, for the simple reason that it is typically not obvious how to construct a sequence of periodic problems that does not have a mismatch at the edges of its period interval which limits to the given problem on the whole real line.

For both of these scenarios the accuracy of the periodic approximation may be judged by the extent to which the finite spectral bands have collapsed to points. No comparison with the FDM approach will be done, as we are only considering black-box FDMs that have equally spaced grids, which is unnecessarily computationally intensive and wasteful. Letting $k = 1 - \epsilon$, using the analytical expressions for the band edges giving in Fig. 4 we find that the length of the spectral bands approaches zero at a quadratic rate in ϵ. Table 3 shows that the Soliton cutoff approach is capable of obtaining equally accurate results compared to the large k limit approach. Note that in general, for problems that are not self-adjoint, a large-period approximation to a whole-line problem will have small circles of continuous spectrum, surrounding the location of the isolated eigenvalues of the limit problem. Self-adjoint problems are one example where these small circles degenerate to small line-segments (“bands”). An overview of results like this may be found in [45], where one also finds statements about the convergence of the radius of the small circles to zero, as the period of the periodic generalization approaches infinity. Results from large k limit numerics and the soliton cutoff approach are given in Table 3. These results illustrate that the exponential mismatch at the edges of the period box result in a loss of accuracy for an equal number of Fourier modes, so that a larger number of modes has to be used to obtain the same accuracy. Another example of the soliton cutoff method is given in the next section.

4.3. The forward scattering problem for the Korteweg-deVries equation

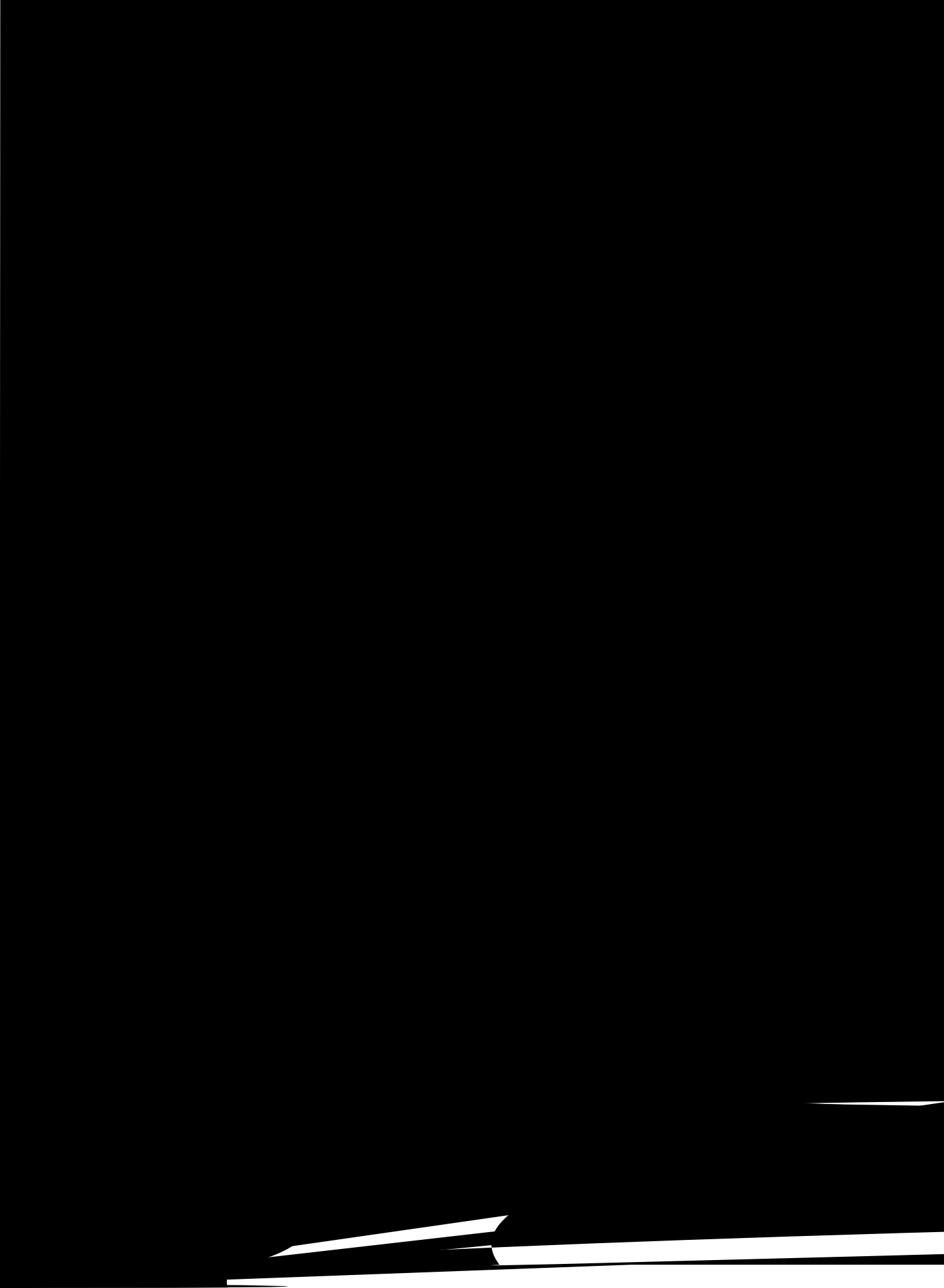
As mentioned in the introduction, the FFHM may be used to numerically solve the forward scattering problem for any integrable equation. A special case of this is illustrated here for the Korteweg-deVries (KdV) equation with sech^2 potential. The KdV equation

$$u_t = 6uu_x + u_{xxx} \tag{42}$$

Table 3
A comparison of the large k approach with the soliton cutoff method

Elliptic modulus k	Period $2K(k)$	N
1×10^{-3}	9.0	22
1×10^{-4}	11.3	27
1×10^{-5}	13.6	32
1×10^{-6}	15.9	37
1×10^{-7}	18.2	42
1×10^{-8}	20.5	47
1×10^{-9}	22.8	52
Soliton cutoff	22	100

For all runs, the number of modes with a given spatial period was chosen to ensure a band length of any finite-sized band less than 10^{-6}. This table illustrates there is a price to pay for the exponential mismatch at the edges of the period box, using the soliton cutoff method.



where $\omega = 2 - k^2$ for solutions of the form (46), or $\omega = 1$ for the soliton solution (48). Substitution of this ansatz in (45) and retaining first-order terms in ϵ results in the linear matrix equation

$$\begin{pmatrix} 0 & L_-(k) \\ -L_+(k) & 0 \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix} = \lambda \begin{pmatrix} U \\ V \end{pmatrix} \tag{50}$$

with the linear operators $L_-(k)$ and $L_+(k)$ as given in (38a) and (38b) for $k \in [0,1)$, and (41a) and (41b) for $k = 1$. The problem at hand is to determine the spectrum of (50), i.e., the set of all λ for which there are bounded perturbations $U(x)$ and $V(x)$. If any such λ has nonzero real part (the spectrum has reflection symmetry across both the real and imaginary λ axis, due to the Hamiltonian character of the NLS equation), the NLS solution is unstable.

Since the linear stability problem (50) has constant coefficients if $k = 0$, the spectrum is easily found in this case. It consists of the entire imaginary axis and the real interval $[-2, 2]$. For the soliton case ($k = 1$) the spectrum is known analytically as well, as are many of the eigenvectors $(U, V)^{\mathrm{T}}$ [49]. The spectrum is contained on the imaginary λ-axis and consists of two semi-infinite line segments from $\pm\mathrm{i}$ to ∞ along the imaginary axis, and an eigenvalue of multiplicity four at the origin. The literature on the stability analysis of the solutions with $0 < k < 1$ is limited. The most recent and most extensive results may be found in [32]. To the best of our knowledge, nobody has been able to describe the spectrum fully for $0 < k < 1$, either analytically or numerically.

In order to use the FFHM we proceed as before, using the Fourier series of the coefficient functions of (50), and substituting an unknown Fourier series for U and V. The procedure differs slightly from what was done before, as (50) is a matrix system of differential equations. In essence, all is similar to what was done for the scalar case. Instead of a scalar bi-infinite difference equation, a block matrix equation with bi-infinite blocks is constructed. The block matrix entries correspond to the elements of the 2×2 matrix in (50). The zero entries give rise to bi-infinite zero blocks, whereas the $L_-(k)$ and $-L_+(k)$ entries result in the same bi-infinite blocks as for the scalar case. Specifically, let

$$U = \mathrm{e}^{\mathrm{i}\mu x} \sum_{j=-\infty}^{\infty} U_j \mathrm{e}^{\mathrm{i}\frac{2\pi jx}{PL}}, \tag{51a}$$

$$V = \mathrm{e}^{\mathrm{i}\mu x} \sum_{j=-\infty}^{\infty} V_j \mathrm{e}^{\mathrm{i}\frac{2\pi jx}{PL}}, \tag{51b}$$

$$\phi^2(x) = \sum_{j=-\infty}^{\infty} \gamma_j \mathrm{e}^{\mathrm{i}\frac{2\pi jx}{L}}, \tag{51c}$$

with

$$U_j = \frac{1}{PL} \int_{-PL/2}^{PL/2} (U(x)\mathrm{e}^{-\mathrm{i}\mu x}) \mathrm{e}^{-\mathrm{i}\frac{2\pi jx}{PL}} \,\mathrm{d}x, \tag{52a}$$

$$V_j = \frac{1}{PL} \int_{-PL/2}^{PL/2} (V(x)\mathrm{e}^{-\mathrm{i}\mu x}) \mathrm{e}^{-\mathrm{i}\frac{2\pi jx}{PL}} \,\mathrm{d}x, \tag{52b}$$

$$\gamma_j = \frac{1}{L} \int_{-L/2}^{L/2} \phi^2(x) \mathrm{e}^{-\mathrm{i}\frac{2\pi jx}{L}} \,\mathrm{d}x, \tag{52c}$$

for $j \in \mathbb{Z}$. Due to the linearity of the equations, the truncated finite-difference equation obtained from (50) using the above is (50) with the following replacements:

$$U \to (U_j)_{j=-N}^{N}, \tag{53a}$$

$$V \to (V_j)_{j=-N}^{N}, \tag{53b}$$

$$L_-(k) \to \left(\left[\left(\frac{2\pi n}{PL} + \mu \right)^2 + \omega \right] \delta_{n,m} - 2\gamma_{(n-m)/P} \delta_{p|n-m} \right)_{n,m=-N}^{N}, \tag{53c}$$

$$L_+(k) \to \left(\left[\left(\frac{2\pi n}{PL} + \mu \right)^2 + \omega \right] \delta_{n,m} - 6\gamma_{(n-m)/P} \delta_{p|n-m} \right)_{n,m=-N}^{N}, \tag{53d}$$

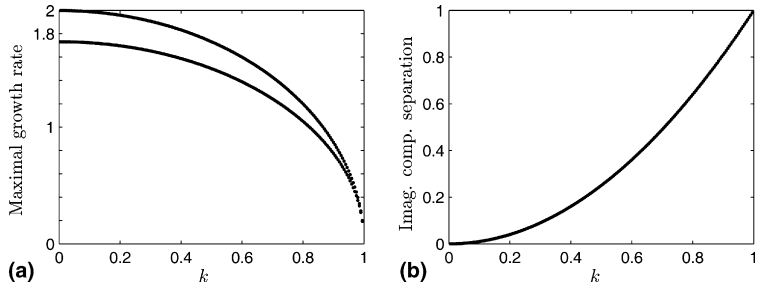


Fig. 8. (a) The maximal growth rate of (46) and its limit cases (47) and (48), as a function of k. For the bottom curve, only periodic and anti-periodic perturbations are considered. The top curve accounts for a variety of μ values, and thus corresponds to perturbations of arbitrary period, or even quasi-periodic ones. (b) The separation of the imaginary components of the spectrum from the real axis. In this case, using periodic or anti-periodic perturbations suffices to find the points of these imaginary components. We note that this separation is a *perfect* fit with the curve k^2, as remarked in [32]. However, we disagree with Kartashov et al. [32] that the top curve in (a) is well described by $2(1-k^2)$. Two hundred points are plotted for each curve. All points plotted are the result of computing a spectrum, with $N = 50$ and $D = 10$.

tion away from the origin of the imaginary component of the spectrum, again for equal-period and other perturbations. Fig. 8 extends Figs. 1a and 2a of [32]. This figure illustrates that it is essential to consider the entire range of μ-values, not just the values on the boundary of its domain: merely considering periodic or anti-periodic modes results in an underestimation of the maximal growth rate by more than 10%.

4.5. The quantum harmonic oscillator

In Section 4.2.2, an example was given where the FFHM was applied to a non-periodic problem. In that case, an exponentially small mismatch existed at the boundaries of the periodic domain, as a consequence of using the soliton cutoff method. In this section, we examine the performance of the FFHM when applied to the one-dimensional quantum harmonic oscillator, i.e., the spectral problem

$$-\psi_{xx} + x^2\psi = \lambda\psi, \quad x \in \mathbb{R}. \tag{54}$$

In this case, the penalty for using the soliton cutoff method is enormous: the x^2 potential is limited to a finite area $x \in [-L/2, L/2]$, resulting in a large derivative mismatch at the boundaries $x = \pm L/2$. The purpose of this section is to illustrate what the FFHM allows one to get away with: the method is not designed to work on a problem like (54). On the other hand, (54) is well-understood and its eigenvalues and eigenfunctions are known and can be found in any quantum mechanics text [39]. Therefore it is a good test case for any numerical method. The results of applying the FFHM are shown in Table 4 and Fig. 9, with parameter values as indicated.

These results illustrate that the computations of the eigenvalues remains highly accurate, even when the pointwise error of the approximation of the eigenfunctions increases. This is not surprising, as this same principle underlies the success of the Rayleigh–Ritz principle, both in quantum mechanics [10] and in nonlinear

Table 4
The first 10 eigenvalues of (54) and their absolute errors using the FFHM with $N = 100$, $L = 16$

Exact eigenvalue	Absolute error using FFHM
1	4.011724286101526e − 11
3	2.640969221090472e − 09
5	6.515937123197091e − 09
7	6.373176653085011e − 09
9	2.706672219687789e − 09
11	2.547496436022811e − 09
13	9.293426472822830e − 09
15	7.243926347655361e − 09
17	1.454309739301607e − 08
19	3.501478929024415e − 08

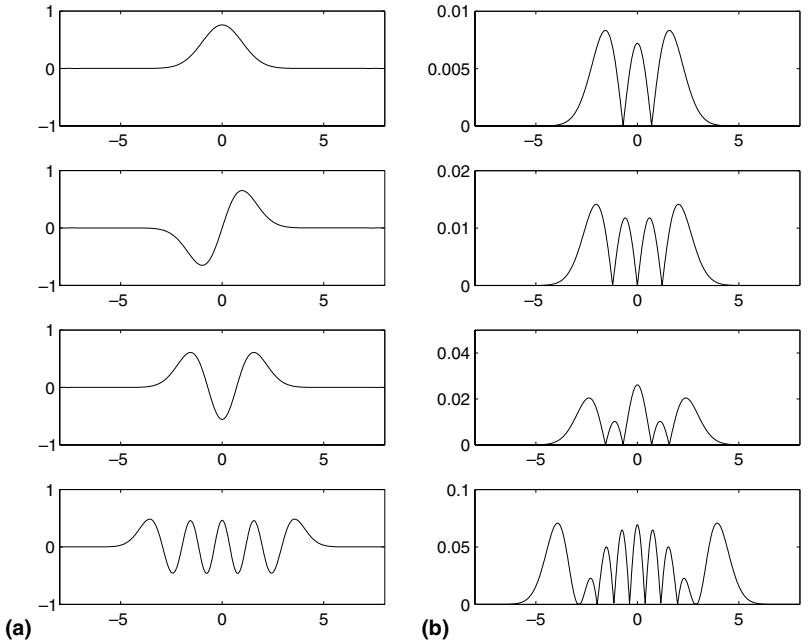


Fig. 9. (a) The first, second, third and tenth eigenfunction of (54) using the FFHM with $N = 100$ and $L = 16$. (b) The absolute value of the pointwise error comparing with the exact eigenfunctions using Hermite functions.

problems [48]. As before, without knowing the exact values of the eigenvalues, the convergence of the method may be estimated by considering the degeneration of small periodic bands or circles of approximate spectral elements to their discrete limit.

Although the approximation of the eigenvalues is significantly more accurate than that of their corresponding eigenfunctions, the approximation of especially the first few eigenfunctions is reasonable. When comparing the effort required here with that which might be required using an FDM to obtain equal accuracy (see Table 1, last two columns), the results are more than satisfactory.

4.6. A two-dimensional Schrödinger equation with periodic coefficients

Just like the FDM, the FFHM is easily adapted to multidimensional problems. Where previously the coefficients of the linear operator were expanded in a one-dimensional Fourier series, now they are expanded in a multidimensional Fourier series. Similarly, the eigenfunctions are expanded as products of multidimensional Fourier series and an exponential factor containing a multidimensional Floquet multiplier. For the sake of explicitness, in the remainder of this section we examine the two-dimensional case. As an example, below we consider a specific two-dimensional Schrödinger equation with a standing light-wave potential.

Any coefficient function of a two-dimensional linear operator with periodic coefficients is expanded as

$$f(x,y) = \sum_{j_x=-\infty}^{\infty} \sum_{j_y=-\infty}^{\infty} \hat{f}_{j_x j_y} \mathrm{e}^{\mathrm{i}2\pi(j_x x/L_x + j_y y/L_y)}, \tag{55}$$

with

$$\hat{f}_{j_x j_y} = \frac{1}{L_x L_y} \int_{-L_x/2}^{L_x/2} \int_{-L_y/2}^{L_y/2} f(x,y) \mathrm{e}^{-\mathrm{i}2\pi(j_x x/L_x + j_y y/L_y)} \,\mathrm{d}x\,\mathrm{d}y, \quad (j_x, j_y) \in \mathbb{Z} \times \mathbb{Z}, \tag{56}$$

where L_x and L_y denote the periods in the x and y directions. Using a two-dimensional extension of Floquet theory (perhaps more appropriately referred to as Bloch theory in the multidimensional context), the eigenfunctions are expanded as

$$w(x,y) = \mathrm{e}^{\mathrm{i}\mu_x x + \mathrm{i}\mu_y y} \sum_{j_x=-\infty}^{\infty} \sum_{j_y=-\infty}^{\infty} \hat{\phi}_{j_x,j_y} \mathrm{e}^{\mathrm{i}2\pi(j_x x/P_x L_x + j_y y/P_y L_y)} = \sum_{j_x=-\infty}^{\infty} \sum_{j_y=-\infty}^{\infty} \hat{\phi}_{j_x,j_y} \mathrm{e}^{\mathrm{i}x(\mu_x + 2\pi j_x/P_x L_x) + \mathrm{i}y(\mu_y + 2\pi j_y/P_y L_y)}, \tag{57}$$

where the expansion coefficients $\hat{\phi}_{j_x,j_y}$, $(j_x, j_y) \in \mathbb{Z} \times \mathbb{Z}$ are of the same dimension as $w(x, y)$. Further, the Floquet exponents (μ_x, μ_y) are limited to the dual lattice of the original spectral problem, with an extra scaling due to the integer factors P_x and P_y: $(\mu_x, \mu_y) \in [0, 2\pi/P_x\, L_x) \times [0, 2\pi/P_y L_y)$. Substitution of the Fourier expansions of the coefficient functions and of the eigenfunction expansions in the spectral problem results in a two-dimensional bi-infinite difference equation for the expansion coefficients $\hat{\phi}_{j_x,j_y}$ of the eigenfunctions. In fact, a two-parameter family of such difference equations is obtained, as the difference equation is parametrized by μ_x and μ_y.

Let us consider the specific example of a two-dimensional Schrödinger equation with a potential that is periodic in both x and y:

$$-\frac{1}{2}(\psi_{xx} + \psi_{yy}) + f(x,y)\psi = \lambda\psi. \tag{58}$$

Substituting (57) in this equation we find the two-dimensional bi-infinite difference equation

$$\frac{1}{2}\hat{\phi}_{k,l}\left[\left(\mu_x + \frac{2\pi k}{P_x L_x}\right)^2 + \left(\mu_y + \frac{2\pi l}{P_y L_y}\right)^2\right] + \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \hat{f}_{\frac{m-k}{P_x},\frac{n-l}{P_y}} \hat{\phi}_{m,n} = \lambda\phi_{k,l}, \tag{59}$$

for $(k, l) \in \mathbb{Z} \times \mathbb{Z}$. Further, as before $\hat{f}_{\frac{m-k}{P_x},\frac{n-l}{P_y}}$ is defined to be zero if either one of $(m-k)/P_x$ or $(n-l)/P_y$ is noninteger.

In order to use (59) as the basis for a numerical scheme for computing the spectrum of the two-dimensional Schrödinger equation (58), the number of modes in each dimension is truncated. Depending on the periods in those dimensions, different truncations may be required for different dimensions, to obtain a certain accuracy. It is clear at this stage how the implementation of such a scheme will differ from that of the one-dimensional setting. Instead of a single loop over all retained Fourier modes in one-dimension, a double nested loop is required, each inner loop corresponding to a loop of Fourier modes in that dimension, depending on the Fourier mode of the outer loop. Generalizing this to more than two dimensions should be obvious. Such constructions are also required for implementations of FDMs in higher dimensions. The essential difference between the two methods is that the number of Fourier modes required for a given accuracy is typically orders of magnitude less than the number of grid points required in an FDM setting, as is seen in Table 1 and Fig. 5. As the number of dimensions increases, this difference is paramount. Suppose a problem with periodic coefficients is considered in one spatial dimension. Assume that to obtain a given accuracy, 10 Fourier modes need to be retained, resulting in a matrix of size 21×21. Using Table 1 as a guide, an FDM matrix of size about 200×200 is required to obtain comparable accuracy, if a 4th order method is used. Now consider a two-dimensional version of this same problem with equal periods in both dimensions. One may expect that 10 Fourier modes in each dimension is sufficient to obtain an accuracy comparable to that obtained in the one-dimensional case. Thus, a matrix of size $21^2 \times 21^2 = 441 \times 441$ is required. For numerical purposes this is still a relatively small matrix. Using FDM4, it appears a matrix of size $200^2 \times 200^2 = 40{,}000 \times 40{,}000$ is needed, which is large by any standard. These comparisons are slated even more in favor of the FFHM if non-equal period modes are considered and many matrices of the above sizes are necessary. This assumes that the FDM4 is incorporating Floquet theory as specified before, and not by considering larger periods. This would increase the required matrix size even more, resulting in prohibitively large matrices. This discussion may be repeated for three or more dimensions, becoming more one-sided as the number of dimensions increases.

It should not be a surprise that the study of spectra of linear Schrödinger operators has received a lot of attention, given their importance in both physics and mathematics. Specifically, the Schrödinger equation with a periodic potential is the fundamental equation of solid state physics. Any text on solid state physics (e.g., the classic [5]) will touch on several ways in which the spectrum of this equation may be understood. One of the canonical ways is the so-called orthogonal plane-wave expansion method, which is identical to Hill's method as described here. This method is the cornerstone for the computations performed by the commercial package RSoft [44]. As stated before, particular aspects of the FFHM have been used before by different authors on specific examples, and the plane-wave expansion method is another instance of this.

In solid state physics, the spectrum is usually plotted as a function of the Floquet parameters $(\mu_x, \mu_y) \in [0, 2\pi/P_x L_x) \times [0, 2\pi/P_y\, L_y)$. This is useful in this setting, as (μ_x, μ_y) is the mathematical representation of the quasi-momentum. With this knowledge, one determines at which values of the quasi-momentum different Brillouin zones exist [5]. We will do the same below, for the case of a standing-light wave potential modeled by

$$f(x,y) = A \sin^2 x \sin^2 y, \tag{60}$$

so that $L_x = L_y = \pi$. Thus, using $P_x = P_y = 2$, $(\mu_x, \mu_y) \in [0,1) \times [0,1)$. Then

$$(\hat{f}_{i,j})_{i,j=-\infty}^{\infty} = A \begin{pmatrix} & \vdots & \vdots & \vdots & \\ \cdots & 1/4 & -1/2 & 1/4 & \cdots \\ \cdots & -1/2 & 1 & -1/2 & \cdots \\ \cdots & 1/4 & -1/2 & 1/4 & \cdots \\ & \vdots & \vdots & \vdots & \end{pmatrix}, \tag{61}$$

and all Fourier coefficients of (60) not explicitly written are zero. The results for two different values of A are shown in Fig. 10. For $A = 0$, the spectrum is $\lambda \in [0, \infty)$, but for other values of A band gaps may appear in this spectrum, as shown. To obtain the results shown in Fig. 10, $N = 3$ modes were used, for 40 different values of the quasi-momentum in each dimension. Thus 49 eigenvalues of $40^2 = 1600$ different matrices of size $(2N+1)^2 \times (2N+1)^2 = 49 \times 49$ were computed. To generate the data for each of these figures took less than 9 s to generate. Lastly, Figs. 11 and 12 display a few eigenfunctions, for parameter values as indicated.

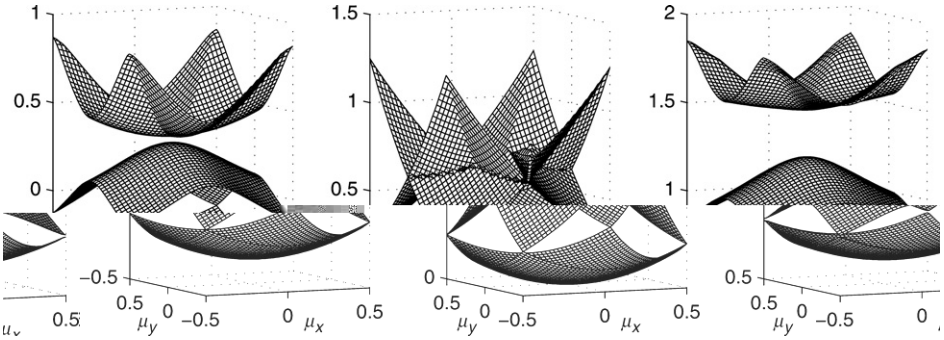


Fig. 10. The eigenvalue λ as a function of the quasi-momentum (μ_x, μ_y), for different values of A: $A = -0.3$ (left), $A = 0$ (middle) and $A = 1$ (right). Note that the spectrum of (58) is obtained by projecting these figures on the vertical axis. For all these runs, $N = 3$ proved sufficient, which is not surprising given that the potential (60) is trigonometric. The data used to generate these figures is accurate to within 10^{-3}, as in the case of the Mathieu equation (see Table 1).

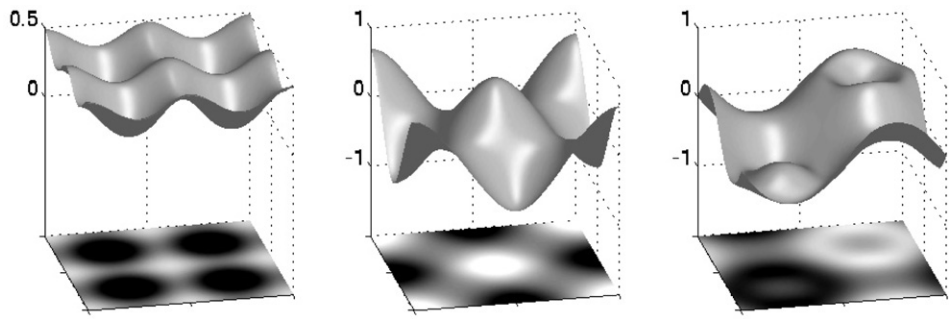


Fig. 11. Eigenfunctions of (58) with $A = 1$ corresponding to zero quasi-momentum, *i.e.*, $(\mu_x, \mu_y) = (0,0)$. These eigenfunctions are periodic with period π in both directions. They correspond to the central points on the three eigenvalue sheets plotted in Fig. 10 on the right.

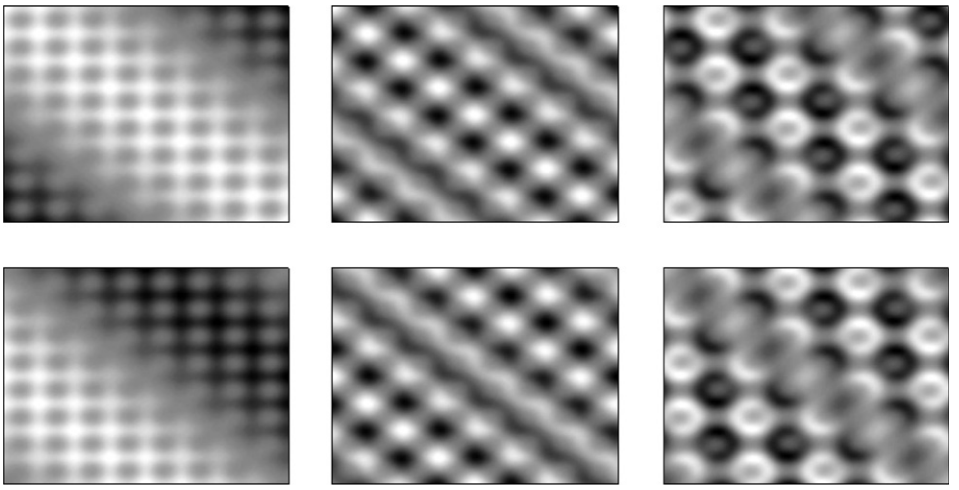


Fig. 12. Contour plots of the real (top) and imaginary (bottom) parts of eigenfunctions of (58) with $A = 1$ corresponding to quasi-momentum $(\mu_x, \mu_y) = (1/4,1/4)$. These eigenfunctions are periodic with period 4π in both directions. They correspond left-to-right to the points at (1/4, 1/4) on the three eigenvalue sheets bottom-to-top plotted in Fig. 10 on the right. The x and y variables range from -2π to 2π.

5. Summary

In 1886, Hill incorporated the theories of both Fourier and Floquet to investigate the spectrum of what is now known as Hill's equation [28]. In this paper, we have used Hill's ideas to develop the Fourier–Floquet–Hill method (FFHM) for the numerical computation of spectra of arbitrary linear operators with periodic coefficients. This method may be applied to scalar or vector operators, in one or more dimensions. The incorporation of Fourier series leads to a spectrally accurate method. Using Floquet theory results in a uniform approximation of entire components of the spectrum, not a selected few points in it. We have also demonstrated that finite-difference methods (FDMs) may be modified almost trivially so as to reap these benefits of incorporating Floquet theory. All of these issues were amply discussed and illustrated in the examples.

In addition to its core application areas, we have illustrated that the FFHM may be applied to problems with coefficients defined on unbounded domains, such as soliton problems or classical problems involving special functions, such as the quantum harmonic oscillator. This greatly extends the applicability of the FFHM.

The black-box method most commonly used for the computation of spectra of linear operators is the FDM. We have demonstrated that for comparable accuracy the FFHM is orders of magnitude faster, due to its spectral accuracy, and its ability to compute approximations to more spectral elements by using multiple smaller matrices instead of a single large matrix. This enables us to accurately compute spectra of linear operators that were previously outside the reach of any computation effort.

Lastly, we have highlighted some interesting features of spectral problems which are often disregarded in the literature: all too often the space of eigenfunctions is restricted to contain only functions that are periodic with the same period as the coefficients, or small integer multiples thereof. As demonstrated in the examples, unless the linear operator is scalar and of second order, this approach is usually flawed, in that it will result in an underestimation of the extremal parts of the spectrum. Perhaps most importantly, for problems originating from a linear stability analysis this leads to an incorrect prediction of: (i) the maximal growth rate; and (ii) the shape of the most unstable mode.

Acknowledgements

We acknowledge useful discussions with J. Rademacher, P.J. Schmid and J. Stockie. We also thank D.E. Pelinovski for pointing us towards [23]. The investigations in this paper were supported by NSF-DMS

0139093 (B.D.) and NSF-DMS 0092682 (N.K.). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] M. Abramowitz, I.A. Stengun, Handbook of Mathematical Functions, With Formulas, Graphs and Mathematical Tables, National Bureau of Standards, Washington, DC, 1966.
[2] M.J. Ablowitz, P.A. Clarkson, Solitons, Nonlinear Evolution Equations and Inverse Scattering, Cambridge University Press, Cambridge, 1991.
[3] M.J. Ablowitz, H. Segur, Solitons and the Inverse Scattering Transform, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1981.
[4] H. Amann, Ordinary Differential Equationsde Gruyter Studies in Mathematics, vol. 13, Walter de Gruyter & Co., Berlin, 1990.
[5] N.W. Ashcroft, N.D. Mermin, Solid State Physics, Saunders College Publishing, Philadelphia, 1976.
[6] Kendall E. Atkinson, An Introduction to Numerical Analysis, Wiley, New York, 1989.
[7] C.M. Bender, S.A. Orszag, Advanced Mathematical Methods for Scientists and Engineers. I, Springer-Verlag, New York, 1999.
[8] P.J. Blennerhassett, A.P. Bassom, The linear stability of flat Stokes layers, J. Fluid Mech. 464 (2002) 393–410.
[9] P.F. Byrd, M.D. Friedman, Handbook of Elliptic Integrals for Engineers and Scientists. Die Grundlehren der mathematischen Wissenschaften, Band 67, second ed., revised., Springer-Verlag, New York, 1971.
[10] A.Z. Capri, Nonrelativistic quantum mechanics, in: Lecture Notes and Supplements in Physics, Benjamin-Cummings, California, 1985.
[11] J.D. Carter. Private communication, 2004.
[12] J.D. Carter, B. Deconinck, Instabilities of one-dimensional trivial-phase solutions of the two-dimensional cubic nonlinear Schrödinger equation, Physica D 214 (2006) 42–54.
[13] J.D. Carter, H. Segur, Instabilities in the two-dimensional cubic nonlinear Schrödinger equation, Phys. Rev. E 68 (2003) 045601.
[14] F. Chen, Introduction to Plasma Physics and Controlled Fusion, Plenum Press, New York, 1984.
[15] E.A. Coddington, N. Levinson, Theory of Ordinary Differential Equations, McGraw-Hill Book Company, Inc., New York-Toronto-London, 1955.
[16] R. Cortez, C.S. Peskin, J.M. Stockie, D.A. Varela, Parametric resonance in immersed elastic boundaries, SIAM J. Appl. Math. 65 (2) (2004) 494–520.
[17] R. Courant, D. Hilbert Methods of Mathematical Physics, vol. I, Interscience Publishers, Inc., New York, NY, 1953.
[18] M.C. Cross, P.C. Hohenberg, Pattern formation outside of equilibrium, Rev. Modern Phys. 65 (3) (1993) 851–1112.
[19] B. Deconinck, F. Kiyak, J.D. Carter, J.N. Kutz, SpectrUW, a software package for the computation of spectra of linear operators. Available from: <http://www.amath.washington.edu/hill/spectruw.html>..
[20] B. Deconinck, F. Kiyak, J.D. Carter, J.N. Kutz. SpectrUW: a laboratory for the numerical exploration of spectra of linear operators (submitted).
[21] P.G. Drazin, R.S. Johnson, Solitons: An Introduction, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 1989.
[22] P.G. Drazin, W.H. Reid, Hydrodynamic Stability, Cambridge University Press, Cambridge, 1981.
[23] S.E. Fil'chenkov, G.M. Fraiman, A.D. Yunakovskii, Instability of periodic solutions of the nonlinear Schrödinger equation, Sov. J. Plasma Phys. 13 (8) (1987) 554–557.
[24] G. Floquet, Sur les équations différentielles linéaires à coefficients périodiques, Ann. École Norm. Ser. 2 (12) (1883) 47–89.
[25] I.S. Gradshteyn, I.M. Ryzhik, Table of Integrals, Series, and Products, fifth ed., Academic Press, Inc., Boston, MA, 1994.
[26] P. Hall, The linear stability of flat Stokes layers, Proc. Roy. Soc. London Ser. A 359 (1697) (1978) 151–166.
[27] A. Hasegawa, Optical Solitons in Fibers, second ed., Springer-Verlag, Berlin, 1990.
[28] G.W. Hill, On the part of the lunar perigee which is a function of the mean motions of the sun and the moon, Acta Math. 8 (1886) 1–36.
[29] D.D. Holm, J.E. Marsden, T. Ratiu, A. Weinstein, Nonlinear stability of fluids and plasma equilibria, Phys. Rep. 123 (1985) 1–116.
[30] E.L. Ince, Ordinary Differential Equations, Dover Publications, New York, 1944.
[31] C.G.J. Jacobi, Fundamenta Nova Theoriae Functionum Ellipticarum, Königsberg (1829).
[32] Yaroslav V. Kartashov, Victor A. Aleshkevich, Victor A. Vysloukh, Alexey A. Egorov, Anna S. Zelenina, Stability analysis of (1+1)-dimensional cnoidal waves in media with cubic nonlinearity, Phys. Rev. E 67 (2003) 036613.
[33] T. Kato, Perturbation Theory for Linear Operators, Springer-Verlag, Berlin, 1966.
[34] J. Kevorkian, J.D. Cole, Multiple Scale and Singular Perturbation Methods, Applied Mathematical Sciences, vol. 114, Springer, Verlag, New York, 1996.
[35] A.N. Kolmogorov, S.V. Fomin, Introductory Real Analysis, Dover Publications Inc., New York, 1975.
[36] M.G. Krein, A generalization of some investigations of A.M. Lyapunov on linear differential equations with periodic coefficients, Doklady Akad. Nauk SSSR (N.S.) 73 (1950) 445–448.
[37] B.M. Levitan, I.S. Sargsjan, Introduction to Spectral Theory: Selfadjoint Ordinary Differential Operators, American Mathematical Society, Providence, RI, 1975.
[38] W. Magnus, S. Winkler, Hill's Equation, Dover Publications Inc., New York, 1979.

[39] E. Merzbacher, Quantum Mechanics, second ed., Wiley International, Singapore, 1970.
[40] R.P. Mied, The occurence of parametric instabilities in finite-amplitude internal gravity waves, J. Fluid Mech. 78 (4) (1976) 763–784.
[41] R.M. Miura, The Korteweg-de Vries equation: a survey of results, SIAM Rev. 18 (3) (1976) 412–459.
[42] J.D. Murray, Mathematical Biology II: Spatial Models and Biomedical Applications, Springer-Verlag, Berlin, 2003.
[43] V.I. Pavlenko, V.I. Petviashvili, Band theory for the stability of nonlinear periodic waves in plasmas, Sov. J. Plasma Phys. 8 (1) (1982) 117–120.
[44] The RSoft Design Group. Available from: <http://www.rsoftdesign.com>/.
[45] B. Sandstede, A. Scheel, On the stability of periodic travelling waves with large spatial period, J. Differen. Equat. 172 (1) (2001) 134–188.
[46] R. Temam, Infinite-dimensional Dynamical Systems in Mechanics and Physics, second ed., .Applied Mathematical Sciences, vol. 68, Springer, Verlag, New York, 1997.
[47] L.N. Trefethen, D. Bau III, Numerical Linear Algebra, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.
[48] T. Tsrumi, H. Morise, M. Wadati, Stability of Bose–Einstein condensates confined in traps, Int. J. Modern Phys. 14 (7) (2000) 655–719.
[49] M.I. Weinstein, Modulational stability of ground states of nonlinear Schrödinger equations, SIAM J. Math. Anal. 16 (3) (1985) 472–491.
[50] E.T. Whittaker, G.N. Watson, A Course of Modern Analysis, Cambridge University Press, Cambridge, 1902.