Spectral Pollution

Alex H. Room

February 9, 2024

Abstract Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Contents

I | INTRODUCTION: THE SPECTRUM AND ITS APPROXIMATION

The computation of spectra can boldly be considered the 'fundamental problem of operator theory' [1]. The spectrum of an operator holds the same key to the entire structure as eigenvalues do for linear algebra, however (as often occurs when passing from the finite- to infinite-dimensional case) we lose the ease of representation that allows the creation of such algorithms and formulae as those used for matrices.

Throughout, our examples will mainly focus on differential and multiplication operators. As a result, we will not generally assume that the operator is bounded (many texts use 'operator' to mean 'bounded operator'). However, we will focus mainly on Hilbert spaces, and assume that the operator has the properties required for the spectrum and adjoint to exist and make sense - these properties are subtle and will not be a hassle or an excessive restriction (merely allowing us to avoid fiddly arguments to deal with pathological examples).

1.1 Spectra

We must first define our quantity of interest: the spectrum of an operator.

Definition. ***(Resolvent and spectrum)*** *(Adapted from [2]) Let T be a linear operator on a Banach space.*

- *The resolvent of T is the set $\rho(T) := \{\eta \in \mathbb{C} : (T - \eta I)$ has a bounded inverse$\}$, where I is the identity operator. If it exists, we call its inverse $(T - \eta I)^{-1}$ the 'resolvent operator' for η and T.*
- *The spectrum of T, denoted $\mathrm{Spec}(T)$, is $\mathbb{C} \setminus \rho(T)$, i.e. the set of all complex numbers λ such that the operator $(T - \lambda I)$ does not have a bounded inverse.*

In the remainder of the text, the identity operator I will be implicit in the operator - i.e. we will simply write $(T - \lambda I)$ as $(T - \lambda)$.

We must keep this generality (rather than just defining eigenvalues of a linear operator like we do for matrices) because in infinite dimensions, the failure of invertibility can come about in a variety of ways. In finite-dimensional spaces, the rank-nullity theorem asserts that if $(T - \lambda)$ is not invertible, then it is not injective. As a result, $(T - \lambda)u = (T - \lambda)v$ for some $u \neq v$; so $T(u - v) = \lambda(u - v)$, and thus any point in the spectrum is an eigenvalue. The variety of ways in which invertibility can generally fail will be explored in Section 3.1.

We will find many of our physical examples to be self-adjoint. We will prove later that if an operator is self-adjoint, then its spectrum lies entirely on the real line; this makes it much more convenient to model, bound and calculate a wide range of results.

Definition. ***(Adjoint, symmetric, and self-adjoint operators)*** *(Adapted from [3])*

- *Let T be an operator on a Hilbert space $\mathcal{H}$. Let $\mathrm{Dom}(T^*)$ be the set of all $v \in \mathcal{H}$ such that the functional*

$$u \mapsto (Tu, v) \quad on\ u \in \mathrm{Dom}(T)$$

  *is bounded. We then define an **adjoint** operator T^* to be the operator such that $(Tu, v) = (u, T^*v)$ for all $u \in \mathrm{Dom}(T), v \in \mathrm{Dom}(T^*)$*
- *An operator T is symmetric if for all $u, v \in \mathrm{Dom}(T)$, (Tu, v) = (u, Tv).*
- *An operator T is self-adjoint if it is symmetric and $Dom(T^*) = Dom(T)$, i.e. it is equal to its adjoint operator.*

Note that a bounded operator is symmetric if and only if it is self-adjoint; bounded operators are defined on the whole of $\mathcal{H}$, so these issues with the domain do not occur.

As we have alluded to, there is no universal algorithm for the calculation of operator spectra in the way that there is the QR algorithm [4] for matrices. To devise a formula for the spectrum of even a specific subset of a class of operators is a mathematical feat, and varieties of operators important to fields such as quantum physics ([5]), hydrodynamics ([6]), and crystallography ([7]) still withhold the structure of their spectra from decades-long attempts at discovery. To this end, we must employ numerical methods. We shall find that even approximating spectra computationally is not so easy.

1.2 Approximating spectra; Ritz-Galerkin methods

It would be a reasonable hypothesis that we can approximate spectra by reducing the infinite-dimensional problem to a finite-dimensional one, where we are on much firmer ground when it comes to finding eigenvalues.

Definition. ***(Compressions & truncations)*** *(Adapted from [8]) Let T be an operator on a Hilbert space $\mathcal{H}$, $\mathcal{L} \subseteq Dom(T)$ a closed linear subspace, and $P_{\mathcal{L}}$ the orthogonal projection of $\mathcal{H}$ onto $\mathcal{L}$.*

- *The **compression** of the operator T, which we will often denote $T_{\mathcal{L}}$, is defined*

$$T_{\mathcal{L}} := P_{\mathcal{L}} T\big|_{\mathcal{L}}$$

where $\big|_{\mathcal{L}}$ denotes domain restriction to $\mathcal{L}$.

- *If $\{\phi_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for $\mathcal{H}$, the k'th **truncation** of T is the compression of T to $Span\{\phi_1, \phi_2, ..., \phi_k\}$. If there is no confusion, we will denote the k'th truncation T_k.*

We will now derive the Ritz-Galerkin method by example; we will motivate and explain the method via constructing an approximation method for a one-dimensional Sturm-Liouville operator, analogously to the Galerkin method for solving general symmetric second-order PDEs given in [4] and [9].

The one-dimensional Sturm-Liouville operator is

$$Lu = -\frac{d}{dx}(P(x)\frac{du}{dx}) + Q(x)u$$

which gives rise to the Sturm-Liouville eigenvalue problem with general boundary conditions[1]:

$$\begin{cases} Lu = \lambda u \\ a_1 u(a) = a_2 u'(a),\ b_1 u(b) = b_2 u'(b) \end{cases} \tag{S-L}$$

on a suitable subspace of $L^2[a,b]$, where P and Q are scalar-valued functions. Q is often called the 'potential' of the operator. These operators occur frequently in partial differential equations; for example, both the heat and wave equations are governed by Sturm-Liouville operators.

To start, we will weaken the problem - currently, u is required to be twice-differentiable. We multiply both sides by a 'test' function v in some suitable space, and integrate both sides to obtain

$$-[Pu'v]_b^a + \int_a^b P(x)u'(x)v'(x) + Q(x)u(x)v(x)dx = \lambda \int_a^b u(x)v(x)dx \tag{1}$$

which we tidy up into the form $\mathcal{A}[u,v] = \lambda(u,v)$ where $\mathcal{A}$ is the bilinear form defined as the left side of (1). and $(\cdot,\cdot)$ is the natural scalar product on L^2. The boundary term nicely encodes the boundary conditions of the original problem into the equation.[2] We call a function u such that $\mathcal{A}[u,v] = \lambda(u,v)$ for all test functions v a 'weak solution' of problem (S-L). Note that to create this weak form, we do not even need u to be differentiable once; it just needs to be integrable by parts with a test function. This provides our suitable subspace of $L^2[a,b]$ (a Sobolev space) which is of course much larger than the space of twice-differentiable functions. We will not discuss the theory of Sobolev spaces in detail, but it is easy to see by direct calculation that if u is a weak solution and also twice-differentiable, then it is a strong solution (as we can reverse the operation done in (1) to recover the initial problem) [10].

The Ritz-Galerkin method then involves approximating this weak formulation over some finite-dimensional subspace. Take the subspace $S = \text{Span}\{\phi_1, \phi_2, ..., \phi_k\}$ and consider the weak problem (1) restricted to S, that is, we want a function $\tilde{u} \in S$ such that $\mathcal{A}[\tilde{u}, \tilde{v}] = \lambda(\tilde{u}, \tilde{v})$ for all $\tilde{v}$ in S. As any $\tilde{v}$ is a combination of basis functions, $\tilde{u}$ is a solution if and only if $\mathcal{A}[\tilde{u}, \phi_j] = \lambda(\tilde{u}, \phi_j)$ for all $j \in \{1, ..., k\}$. Then as $\tilde{u} \in S$,

$$\tilde{u} = \sum_{i=1}^{k} u_i \phi_i$$

[1] See that the general boundary conditions cover the 'regular' set of homogeneous boundary conditions; Dirichlet if $a_2 = 0$, Neumann if $a_1 = 0$, Robin otherwise.

[2] For most boundary conditions, one can substitute u' via the boundary condition; for Dirichlet boundary conditions, we will require v to be zero at the boundaries.

for some values u_i; if we substitute this into the problem we obtain a system of linear equations:

$$\mathcal{A}[\tilde{u}, \phi_j] = \sum_{i=1}^{k} \mathcal{A}[\phi_i, \phi_j] u_i$$
$$\lambda(\tilde{u}, \phi_j) = \lambda u_j$$
$$\Rightarrow \sum_{i=1}^{k} \mathcal{A}[\phi_i, \phi_j] u_i = \lambda u_j$$

making use of the orthonormality of ϕ_i and ϕ_j. This is equivalently a matrix eigenvalue problem $Mu = \lambda u$ where $M_{i,j} = \mathcal{A}[\phi_i, \phi_j]$.

Notice that if u is in the domain of L, we have $\mathcal{A}[u, v] = (Lu, v)$. For an operator T defined on the whole space (e.g. a bounded operator), we can approximate the spectrum by the eigenvalues of a matrix $M_{i,j} = (T\phi_i, \phi_j)$ [11]. We will call these matrices Ritz matrices; in general, a wide variety of boundary value problems can be weakened and discretised to be represented by these methods. The following example shows that this method can be effective:

Example 1. *Consider the following Schrödinger operator, which describes the movement of a hydrogen atom's electron [9]:*

$$Hu = -u'' + (-\frac{1}{x} + \frac{2}{x^2})u$$

The spectrum of this operator, fascinatingly, corresponds to the energy levels of certain 'stable' states of the atom in quantum mechanics. The analysis of this operator for larger atoms is an open problem; for example, the spectrum for the equivalent equation in the helium atom (let alone any larger atom!) has only been calculated numerically [8]. The hydrogen atom's spectrum can be found concretely, and in the space $L^2(0, \infty)$ it has eigenvalues $\lambda_k = -\frac{1}{(2k+4)^2}$ for $k \in \mathbb{N}_0$.

Let us see how well a Ritz-Galerkin method can approximate this known spectrum. We can see the results of an approximation in Figure 1 done via the `specpol` software. We choose the weighted Laguerre polynomial basis $\phi_n = \exp(-x/2)L_n$, where L_n is the n'th Laguerre polynomial; this is a complete orthonormal set on the half-line [12]. In this case, it is a very accurate approximation; even at the lowest matrix size of 50, the 6 smallest eigenvalues are accurate to 3 significant figures (and despite possible error induced via other numerical parts of the computation, e.g. inaccuracy in the calculation of basis functions or in the quadrature used)

Eigenvalue	Exact	Approximated
λ_0	-0.0625	
λ_1	$-0.02\dot{7}$	
λ_2	-0.015625	
λ_3	-0.01	
λ_4	$-0.0069\dot{4}$	
λ_5	-0.00510204...	

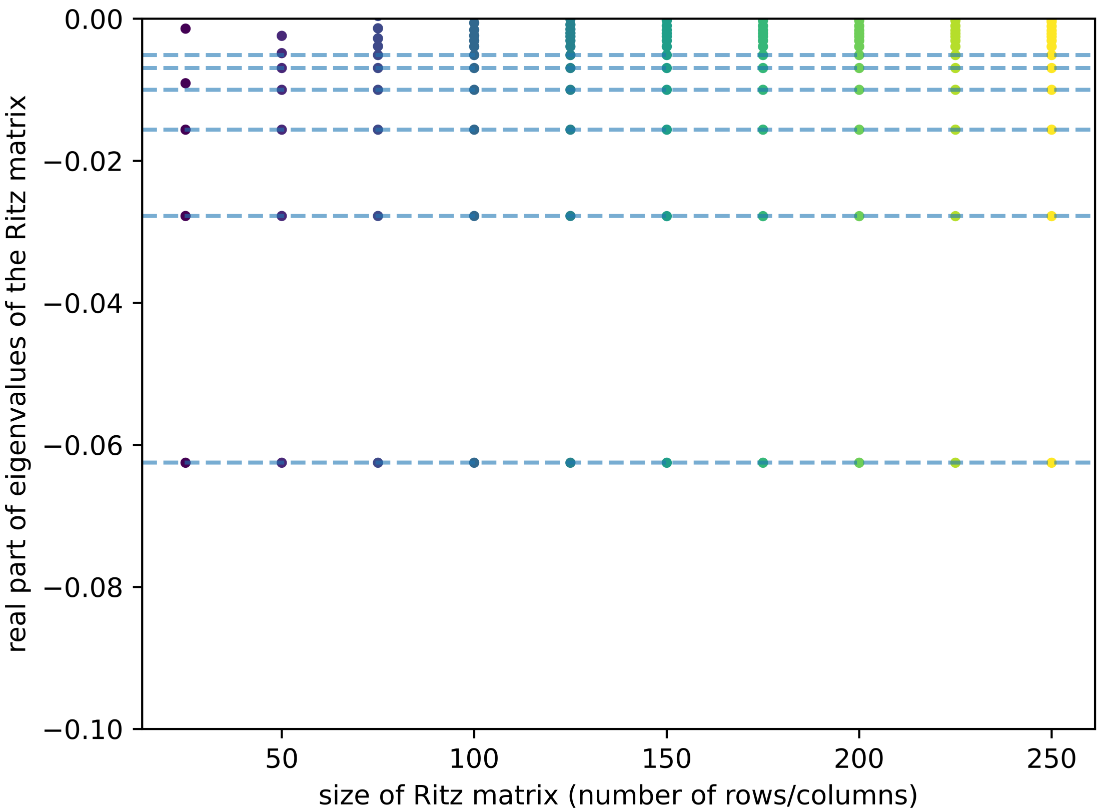

Figure 1: A Ritz-Galerkin approximation for the Schrödinger operator given above, cropped to ignore the positive spectrum (which is known to be the whole positive half-axis). The dotted lines correspond to where the first 6 eigenvalues should be according to the formula. Note how as the size of the Ritz matrix increases, the higher eigenvalues (closer to the origin, as they are a sequence converging to 0) 'fill in'.

Indeed, comparing this derivation to the earlier definitions of truncations, we can see that we are calculating the eigenvalues of truncations T_n of the operator T. The natural question which follows is to ask about the convergence of these methods as n becomes large. The answer is that this convergence is not perfect - in fact, for Schrödinger operators in particular it has been shown that a 'perfect' truncation algorithm is impossible [13]. In our case, the spectrum of the operator will be approximated, but in many cases there will also be a lot of other spurious values in the spectrum. This brings us to our main subject; the study of these spurious eigenvalues, known as spectral pollution.

1.3 Spectral pollution

Definition. ***(Spectral pollution)*** *(Adapted from [8]) Let* $(T_n)_{n\in\mathbb{N}}$ *be an increasing sequence of truncations of an operator* T*. A value* $\lambda \in \mathbb{C}$ *is said to be a point of* ***spectral pollution*** *if there is a sequence* $\lambda_n \in \mathrm{Spec}(T_n)$ *such that* $\lambda_n \to \lambda$ *but* $\lambda \notin \mathrm{Spec}(T)$*.*

Points of spectral pollution are, intuitively, artefacts of the approximation which will never converge to a point in the actual spectrum. We will see that they exist, that they are relatively common, and that they *get worse* as the approximation goes to higher iterations. Unless we already know what the spectrum of the operator is, it can be incredibly hard for us to decide whether a point is actually in the spectrum or whether it is spurious. In applications of spectral theory, this difference can be beyond a simple 'noisy data' nuisance - rather, a confounding problem.

Example 2.

We see in Figure **??** that this approximation does not work so well. It looks like it successfully covers the relevant parts of the spectrum, but there are a lot of additional eigenvalues which do not exist!

There are, of course, other methods for approximating operator spectra, such as the popular finite difference or 'shooting' methods [4], or specialised methods such as Prüfer and Pruess methods for Sturm-Liouville problems [9], which are not subject to pollution. So why do we care about truncation methods? The central motivation for these methods is that they make *almost no* assumptions about the operator itself or the location of its spectrum. Even for the operators covered by these specialised methods, eigenvalues in certain parts of the spectrum cannot be accessed without significant tweaking [14], not to mention that many methods only apply to problems in one dimension. If the spectral pollution for a sequence of truncations can be discarded, detected or otherwise dealt with, it would provide a unified and powerful approach to numerical spectral theory for almost any operator.

II | Spectral pollution in a multiplication operator

2.1 The spectrum of a multiplication operator

Definition. ***(Multiplication operator)*** *Let $\mathcal{H}$ be a function space. For a given $a \in \mathcal{H}$, the multiplication operator M_a on $\mathcal{H}$ is defined by the action $M_a f(x) = a(x)f(x)$; that is, it acts by pointwise multiplication with a. We call a the 'symbol' of the multiplication operator.*

We can see immediately that the adjoint of M_a is $M_{\overline{a}}$, and thus that M_a is self-adjoint iff a is real-valued.

The spectrum of a multiplcation operator is easy to calculate. We first need to define the 'essential range' of a function. Intuitively, this is similar to the standard range of a function, but ignoring values taken by the function on a set of measure zero - two functions which are equal almost everywhere will have the same essential range.

Definition. ***(Essential range and essential supremum)*** *The essential range of a real-valued function f is the set:*

$$\{k \in \mathbb{R} : \forall \varepsilon > 0, \mu\{x : |f(x) - k| < \varepsilon\} > 0\}$$

where μ is the Lebesgue measure.

The essential supremum of f, denoted $\text{esssup} f$*, is the supremum of the essential range of f. We define* $\text{essinf} f$ *mutatis mutandis for the infimum.*

The normed vector space L^∞ is the space of all bounded measurable functions, and its norm is

$$\|f\|_\infty = \text{esssup}(|f|).$$

Lemma 2.1. *(Adapted from [1]) A multiplication operator on L^2 is bounded if and only if its symbol is in L^∞.*

Proof. Let M_a be a multiplication operator on a Hilbert space $\mathcal{H}$ with symbol $a \in L^\infty$. We see that $|f(z)| \leq \|a\|_\infty$ for almost all z, so for any $g \in \mathcal{H}$, $|ag| = |a||g| \leq \|a\|_\infty |g|$ pointwise almost everywhere; thus

$$\|ag\|_2 = \sqrt{\int |ag|^2} \leq \sqrt{\int \|a\|_\infty^2 |g|^2} = \|a\|_\infty \sqrt{\int |g|^2} = \|a\|_\infty \|g\|_2.. \qquad (\star)$$

This implies $\|M_a\| = \sup_{g \in \mathcal{H}} \|ag\|_2 \leq \|a\|_\infty$, so M_a is bounded.

Conversely, assume M_a is bounded, $a \neq 0$ and let $c < \|a\|_\infty$ (of course, we do not assume $\|a\|_\infty$ is finite). Then we know that the set $\{z : |a(x) > c|\}$ must have positive measure, and by σ-finiteness we have a subset $E \subseteq \{z : |a(x) > c|\}$, and its characteristic function $\mathbf{1}_E$ is in L^2. Then $|a\mathbf{1}_E| \geq c\mathbf{1}_E$, and by a similar calculation to $(\star)$ and taking the supremum, $\|M_a \mathbf{1}_E\|_2 \geq c\|\mathbf{1}_E\|_2$, so $\|M_a\| \geq c$. Taking the supremum over all c gives $\|M_a\| \geq \|a\|_\infty$, so $\|a\|_\infty$ is finite and so $a \in L^\infty$. □

Note from this lemma we also get that $\|M_a\| = \|a\|_\infty$.

Theorem 2.2. *The spectrum of the multiplication operator M_a is the essential range of its symbol a.*

Proof. (Adapted from [15]) Let $(M_a - \lambda)f = g$. Then $f(z) = \frac{g}{a(z)-\lambda}$, and we see that $(M_a - \lambda)$ is invertible if and only if the operator $M_{(a-\lambda)^{-1}}$ is bounded, which is the case if and only if $(a - \lambda)^{-1} \in L^\infty$ by Lemma 2.1. This is the case exactly when $(a - \lambda) \geq \epsilon$ almost everywhere - by definition, this means that $(M_a - \lambda)$ is invertible if and only if λ is not in the essential range of a. □

We will now take advantage of the ability to create simply-defined operators with easy-to-calculate spectra to observe the existence of spectral pollution.

2.1.1 Estimating spectra computationally

Example 3. *Let M_f be the multiplication operator on $L^2(0,1)$ with symbol*

$$f : x \mapsto \begin{cases} x & x < 1/2 \\ x + 1/2 & \text{otherwise.} \end{cases}$$

By Theorem 2.2, the spectrum of M_f is the set $[0, 1/2] \cup [1, 3/2]$. If our Ritz approximation works as expected, we should see the approximation create two dense clusters of eigenvalues which approach these intervals.

Figure 2 shows a plot of the approximate spectrum for various Ritz matrix sizes. This approximation was done for the sequence of truncations T_{25n} for $n \in \{2, 3, ..., 20\}$ over the orthonormal basis $\phi_n(x) = \exp(2i\pi nx)$ for $n \in \mathbb{Z}$; each truncation was over the space $\text{span}\{\exp(2i\pi kx), k \in \mathbb{Z}, |k| < n/2\}$.

We do successfully and rather quickly get eigenvalues corresponding to our actual spectrum, but we also get a lot of other eigenvalues; some of them are converging to points in the spectrum, but as the approximation improves, the pollution doesn't fully dissipate.

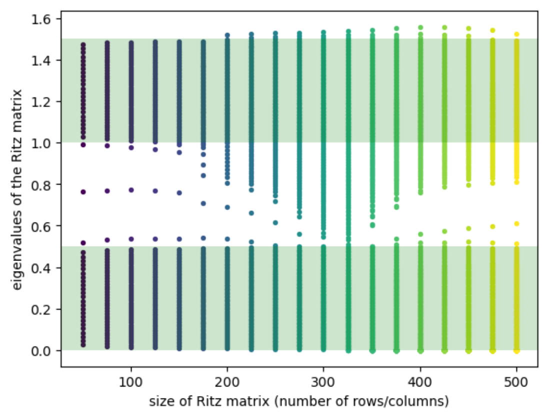

Figure 2: The approximate spectrum of the multiplication operator M_f for Ritz matrices of increasing size. The shaded green areas correspond to the actual spectrum of the operator.

Some of these extra values seem to eventually converge into the correct part of the spectrum, but others stay around; in particular, for all of these approximations we have an eigenvalue of roughly 0.95 which does not exist in the operator's spectrum.

Of course, this is not particularly rigorous - we do not yet know anything about the asymptotic behaviour of the eigenvalues of these approximations (it may well converge far beyond our biggest estimate here of a 500×500 matrix) - but this example is certainly motivating. If one had no knowledge of the actual spectrum of M_f, they would be forgiven for considering this to be strong empirical evidence that the spectrum contains the range $[0.8, 1.0]$ or at least some subset of it.

We have also raised a variety of other questions about the nature of this pollution - why are we only getting it in the gap between the intervals, and not far outside? Is it particular to our choice of sequence for our orthogonal projections (in particular, is there some choice which avoids pollution entirely)?

We will first discuss the nature of spectral pollution for a multiplication operator, taking advantage of an underlying structure to its Ritz matrices which make it possible to concretely identify the existence and location of spectral pollution. Following chapters will then devise technology that allows us to answer these questions in greater generality.

2.1.2 The structure of approximations

If we take a closer look at the structure of our approximating Ritz matrices we uncover deeper structure, which will lead to our next topic.

Example 4. *Let M_f be a multiplication operator on $L^2(0,1)$. Now we create the Ritz matrix, $A_{j,k} = (M_f\phi_j, \phi_k)$, choosing the orthonormal basis $\phi_n(x) = \exp(2\pi inx)$.*

We now note that there is a structure to our matrix:

$$\begin{aligned}(M_f\phi_j, \phi_k) &= \int_0^1 f(x)\exp(2\pi ijx)\exp(-2\pi ikx)dx \\ &= \int_0^1 f(x)\exp(2\pi i(j-k)x) \\ &= c_{j-k}\end{aligned}$$

where c_n is the n'th Fourier coefficient. Thus our Ritz matrix depends only on the value of $j-k$; in particular, it is constant along each diagonal. This is a special type of matrix known as a Toeplitz matrix.

As a result, the approximation of our operator M_f is equal to the approximation of a infinite matrix $(T_f)_{j,k} = c_{j-k}, j,k \in \mathbb{N}_0$ by its truncations $(T_{f,n})_{j,k} = c_{j-k}$ for $j,k \leq n$. And nowhere in this derivation did we use particular properties of f; we can repeat this reasoning with any function capable of being represented by a Fourier series. Let us systematise what we have seen.

2.2 Toeplitz operators

2.2.1 Toeplitz operators, Toeplitz matrices, and their equivalence

The approximation of spectra provides a natural gateway to the study of Toeplitz operators, a type of operator with an elegant cluster of representations that will provide intuitive and concrete insight into spectral pollution. A full account of Toeplitz theory is the subject of whole monographs (such as [16]) and is an entire subfield in itself; here we will stick to exploring their spectra, and how these relate to the spectra of their multiplication operator neighbours.

Definition. ***(Toeplitz matrix)*** *A matrix A (finite or infinite) is Toeplitz if it is constant along its diagonals; that is, $A_{i,j} = A_{i+1,j+1}$ for any i,j (where $i,j < N \in \mathbb{N}$ if A is finite, or $i,j \in \mathbb{Z}_+$ when A is infinite).*

Note that an infinite Toeplitz matrix induces an operator on $\ell^2(\mathbb{Z}_+)$. We see from our example that if $f \in L^\infty$ is represented by the Fourier series

$$f(z) = \sum_{k=-\infty}^{\infty} c_k e^{in\theta},$$

then it induces a Toeplitz matrix T_f with $(T_f)_{j,k} = c_{j-k}$. A natural question is to then ask whether a Toeplitz matrix induces a function, and the answer is affirmative. Firstly, we must define a relevant setting for our matrices; indeed, an important part of our example was that $M_f \exp(2\pi ijx)$ was well-defined. The following definitions are adapted from [1].

Definition. ***(Hardy space)*** *Let ζ be the monomial function on $L^2(\mathbb{T})$, $\zeta(z) = z$, where $\mathbb{T}$ is the unit circle. Then the Hardy space H^2 is the span of all non-negative exponents of ζ; span$\{1, \zeta, \zeta^2 ...\}$.*

As we are on the unit circle, z is more familiarly $e^{i\theta}$ for some θ; then the basis of Hardy space becomes $\{e^{in\theta}\}_{n\in\mathbb{Z}_+}$. Then we can identify any element $f \in H^2$ as any square-integrable function defined on the unit circle with the Fourier series

$$f(e^{i\theta}) \sim \sum_{k=0}^{\infty} c_k e^{ik\theta},$$

that is, with all negative Fourier coefficients equal to zero. This can be identified with the operator on $\ell^2(\mathbb{Z}+)$ via the isomorphism $\sum_{k=0}^{\infty} c_k e^{ik\theta} \mapsto (c_k)_{k\in\mathbb{Z}_+}$ [16].

A generalised definition can be made for H^p via any $L^p(\mathbb{T})$; we will almost entirely use H^2 (with the exception of needing H^1 later on), which is often defined as 'the' Hardy space.

Definition. ***(Toeplitz operator)*** *Let $\phi \in L^\infty(\mathbb{T})$ be bounded and measurable on the unit circle. The Toeplitz operator T_ϕ is the compression of M_ϕ to the Hardy space: $T_\phi = P_{H^2} M_\phi\big|_{H^2}$. We call ϕ the 'symbol' of T_ϕ.*

One may notice what appears like a clash of notation between the induced Toeplitz matrix that was just discussed with the Toeplitz operator. This is not so; there is an elegant relation between Toeplitz operators and Toeplitz matrices.

Theorem 2.3. *Let A be a bounded operator on H^2 such that $(A\zeta^j, \zeta^k) = a_{j-k}$ for some sequence $(a_n)_{n\in\mathbb{Z}}$. Then there is some function $\phi \in L^\infty$ such that $A = T_\phi$ and a_n are the Fourier coefficients of ϕ.*

Proof. □

Many properties of Toeplitz operators are hard to see via infinite matrices. Being able to represent them as both Fourier series and equally as the compressions of multiplication operators puts us on the firmer ground of functional analysis, rather than asymptotic linear algebra. From this, we are now in the position to exactly calculate the spectrum of a Toeplitz operator with real-valued symbol.

2.2.2 The spectrum of a Toeplitz operator with real symbol

To begin, we require a pair of properties regarding functions in $L^1(\mathbb{T})$'s Hardy space, H^1.

Lemma 2.4. ***(Properties of H^1 functions)*** *Let H^1 be the space of all functions $f \in L^1(\mathbb{T})$ where f has the Fourier series*

$$f(e^{i\theta}) \sim \sum_{n=0}^{\infty} a_n e^{in\theta}$$

i.e. has no negative Fourier coefficients. Then the following properties hold:

1. *If $f, g \in H^2$, then $fg \in H^1$;*
2. *If $f \in H^1$ is real-valued, then f is constant.*

Proof. (1.) Let f, g be in H^2. Then in particular, $f, g \in L^2$, and so their product fg is in L^1 (this can be seen directly by the Hölder inequality). Now if f has Fourier coefficients a_k and g has Fourier coefficients b_k, the k'th Fourier coefficient of fg is given by $\sum_{n\in\mathbb{Z}} a_n b_{k-n}$. For negative k, we now have that either $n < 0$ so $a_n = 0$, or $n \geq 0$ so $b_{k-n} = 0$ as $k - n < 0$; this means that the Fourier series of fg satisfies the Hardy space property, but we must justify that the Fourier series of a product is equal to the product of the Fourier series - indeed, they converge to f in L^2 as

(2.) For any real-valued function, we have the relation $\overline{c_{-n}} = c_n$ for the Fourier coefficients of f:

$$\overline{c_{-n}} = \overline{\int f(x)e^{-inx}dx} = \int \overline{f(x)}e^{inx}dx = \int f(x)e^{inx}dx = c_n.$$

But if f is in Hardy space, $c_{-n} = 0$ for all $n \in \mathbb{N}$, and so $c_n = \overline{c_{-n}} = 0$ for all $n \in \mathbb{N}$. Thus the only coefficient remaining is c_0, and a function with a constant Fourier series is a constant function. (This proof is valid for any $H^p, p \in [1, \infty)$.) □

Lemma 2.5. *Let $\phi \in L^\infty$ be real-valued; then the Toeplitz operator T_ϕ is self-adjoint.*

Proof. Let $u, v \in H^2$. Then

$$\begin{aligned}(T_\phi u, v) &= (P_{H^2} M_\phi u, v) \\ &= (M_\phi u, v) && \textit{as } P_{H^2} \textit{ is self-adjoint and } P_{H^2} v = v \\ &= (u, M_\phi v) = (P_{H^2} u, M_\phi v) = (u, T_\phi v),\end{aligned}$$

using that M_ϕ is self-adjoint if ϕ is real-valued. □

We are now in the position to calculate the form of the spectrum for any Toeplitz operator. This spectrum was first found by Toeplitz himself under a stronger set of regularity assumptions, and then weakened by Wiener via his Tauberian theorem [17]. Our proof will consist of a series of claims about the resolvent set of T_ϕ, following a proof outlined in an exercise of Arveson ([1], Chapter 4.6, exercises 2-5) which is based on a proof by Hartman and Wintner.

Theorem 2.6. ***(Hartman-Wintner)*** *Let $\phi \in L^\infty$ be real-valued. Then* $\operatorname{Spec}(T_\phi) = [m, M]$*, where m and M are the infimum and supremum of the essential range of ϕ (Definition 2.1) respectively.*

Proof. Note we already know that as ϕ is real-valued, T_ϕ is self-adjoint, and so its spectrum is on the real line. Furthermore, we will assume that ϕ is non-constant, as if ϕ is constant then T_ϕ is some constant multiple of the identity operator and its spectrum is simply the set containing that constant, and the result holds.

I. Let $\lambda \in \mathbb{R}$ be such that $T_\phi - \lambda$ is invertible. Then for some $k \in \mathbb{C}$ there is a non-zero function $f \in H^2$ such that $(T_\phi - \lambda)f(z) = k$ for any z. By the definition of invertibility this is true for $f = (T_\phi - \lambda)^{-1}\kappa$, where κ is the constant function in H^2; $\kappa : z \mapsto k$.

II. Now we claim that $(\phi - \lambda)|f|^2$ is constant almost everywhere. To do this, we first show that it is in H^1. Indeed, we have $(\phi - \lambda)|f|^2 = ((\phi - \lambda)\overline{f})f$. Then by the previous part,

$$(\phi - \lambda)\overline{f} = \overline{(\phi - \lambda)f} = \overline{\kappa}$$

where $\overline{\kappa}$ maps z to $\overline{k}$. Then it is still a constant function so is still in H^2, and by Lemma 2.4.1, $((\phi - \lambda)\overline{f})f$ is in H^1 as a product of two H^2 functions. Now we use Lemma 2.4.2: because ϕ and λ are real-valued, $(\phi - \lambda)|f|^2$ is also real-valued, so must be constant. Let this constant be called c.

III. Our next claim is that $\phi - \lambda$ crosses the x-axis almost nowhere. This is almost immediate once we invoke a theorem of F. and M. Riesz[3], which states

> *Let f be a non-zero function in H^2. Then the set $\{z \in \mathbb{T} : f(z) = 0\}$ has Lebesgue measure zero.*

This means that $(\phi(z) - \lambda) = \frac{c}{|f(z)|^2}$ is well-defined on L^∞. Then because $|f|^2 > 0$ a.e., $(\phi - \lambda)$ is positive almost everywhere if $c > 0$, and negative almost everywhere if $c < 0$. Note that if $c = 0$, then $(\phi(z) - \lambda) = 0$, so ϕ is the constant function with value λ and the result holds by our remark at the beginning of this proof.

IV. Finally, we show $\operatorname{Spec}(T_\phi) = [m, M]$. By our previous claim, if $T_\phi - \lambda$ is invertible, then either:

- $\phi(z) - \lambda < 0$ a.e., so $\phi(z) < \lambda$ a.e., so $\lambda > M$, or
- $\phi(z) - \lambda > 0$ a.e., so by the same argument $\lambda < m$.

Thus $T_\phi - \lambda$ is invertible only outside of the set $[m, M]$, as required.

□

2.2.3 Multiplication operators, Toeplitz operators, and spectral pollution

Let us now bring the discussion back to spectral pollution, and to our discovery in Example 4. The Ritz matrices corresponding to the multiplication operator M_f is *identical* to the Ritz matrices corresponding to the Toeplitz operator T_f (which are just truncations of the corresponding Toeplitz matrix), but these operators have different spectra. No wonder we are seeing extra eigenvalues in the gap; the approximation 'cannot tell' the difference between an operator which has the essential range as its spectrum from one which has the same maximum and minimum but with all gaps filled in. Heuristically, one may even expect that

[3] Which can be proven as a corollary of a famous theorem of Buerling; see [1], chapter 4.5.

with a large enough approximation, the *entire gap* could fill up with spurious eigenvalues[4].

More rigorously, it is possible to show that for any point in a gap of the multiplication operator's spectrum, some subsequence of truncations will have an approximate eigenvalue converging to that point.

Theorem 2.7. ***(Schmidt-Spitzer [17])*** *Let T_n be the $n \times n$ matrix created by taking the first n rows and columns of a Toeplitz operator T_f. Consider the set*

$$B = \{\lambda \in \mathbb{C} : \lambda = \lim_{m \to \infty} \lambda_{i_m}, \lambda_{i_m} \in \mathrm{Spec}(T_{i_m}), i_m \to \infty\}.$$

where i_m is a subsequence of $\mathbb{N}$. Then if f is real-valued, $B = \mathrm{Spec}(T_f)$.

Proof. If f is real-valued, then the corresponding Toeplitz matrix is Hermitian; its Fourier coefficients have the property $\overline{c_{-n}} = c_n$ (see the proof of Lemma 2.4.2), so if A is the corresponding Toeplitz matrix, $A_{i,j} = c_{i-j} = \overline{c_{j-i}} = \overline{A_{j,i}}$. Label the eigenvalues of T_n as $\lambda_{n_1}, \ldots \lambda_{n_{n+1}}$.

We now invoke a theorem of Szegö [18] regarding the distribution of eigenvalues: on any real interval $[a, b]$,

$$\lim_{n \to \infty} \frac{1}{n+1} \sum_{i=1}^{n+1} \mathbf{1}_{[a,b]}(\lambda_{n_i}) = \frac{1}{2\pi} \mu\{\theta : a \leq f(e^{i\theta}) \leq b\} \tag{2}$$

where μ is the Lebesgue measure, and $\mathbf{1}$ the characteristic function of $[a, b]$ (i.e. $\mathbf{1}_{[a,b]}(\lambda_{n_i})$ is 1 if $\lambda_{n_i} \in [a, b]$ and zero otherwise). Note that this makes sense, as T_n is Hermitian and therefore every eigenvalue is real-valued. Note also that by Theorem 2.6 and the definition of the unit circle $\mathbb{T}$, the spectrum of T_f is $\{f(e^{i\theta}) : \theta \in [0, 2\pi]\}$. The claim of equation (2) is therefore "as the size of the Toeplitz matrix increases, the proportion of eigenvalues of T_n in $[a, b]$ converges to the proportion of the spectrum in $[a, b]$". Of course, this proportion will never be 0 for any interval $[a, b] \subseteq [\mathrm{essinf} f, \mathrm{esssup} f]$ of positive measure, so we can guarantee that a sequence of truncations has an eigenvalue which converges in $[a, b]$. The result follows. □

This is a striking result - for any point in the spectral gap $\mathrm{Spec}(T_f) \backslash \mathrm{Spec}(M_f)$ (or indeed in $\mathrm{Spec}(M_f)$, but they aren't really polluting anything), we can choose a sequence of truncations such that there is guaranteed to be spectral pollution at that point!

[4] We also have the inverse question; if we are approximating the Toeplitz operator T_f, why does it take so much longer to approximate the spectrum in the interval $\mathrm{Spec}(T_f) \setminus \mathrm{Spec}(M_f)$?

III | Bounding spectral pollution

3.1 The essential spectrum of an operator

We begin by producing a result which allows us to constructively classify points in the spectrum of an operator.

Theorem 3.1. ***(Approximate eigenvalue theorem)*** *Consider an operator T on a Hilbert space $\mathcal{H}$, and $\lambda \in \mathbb{C}$. λ is in the spectrum of T if there exists a sequence u_n in $\mathcal{H}$ with the following properties:*

- $\|u_n\| = 1 \quad \forall n \in \mathbb{N}$, *and*
- $\lim_{n\to\infty} \|(T-\lambda)u_n\| \to 0$.

Proof. Assume for contradiction that the resolvent $(T-\lambda)^{-1}$ exists. Then:

$$\begin{aligned} 0 \le \lim_{n\to\infty} \|u_n\| &= \lim_{n\to\infty} \|(T-\lambda)^{-1}(T-\lambda)u_n\| \\ &\le \|(T-\lambda)^{-1}\| \lim_{n\to\infty} \|(T-\lambda)u_n\| && \textit{(as } (T-\lambda)^{-1} \textit{ is bounded)} \\ &= 0 \end{aligned}$$

and so $\|u_n\| \to 0$. But $\|u_n\| = 1$ for every n, so it cannot converge to zero! Thus this bounded inverse does not exist. □

We will call the subset of the spectrum created by this theorem the **approximate point spectrum**, denoted Spec_{ap}:

$$\mathrm{Spec}_{ap}(T) := \{\lambda \in \mathbb{C} : \exists u_n \text{ s.t. } \|u_n\| = 1 \quad \forall n \in \mathbb{N}, \text{ and } \lim_{n\to\infty} \|(T-\lambda)u_n\| \to 0\}.$$

Note that any eigenvalue is in the approximate point spectrum - if φ is a normalised eigenvector corresponding to λ, then the constant sequence $u_n = \varphi$ satisfies this property. Indeed, for a specific (and very common) type of operator, we can see that this criterion covers the entire spectrum.

Lemma 3.2. *For any operator T,* $\mathrm{Spec}(T) = \overline{\mathrm{Spec}_{ap}(T^*)} \cup \mathrm{Spec}_{ap}(T)$.

To prove this, consider what makes a point λ capable of being in $\mathrm{Spec}(T) \setminus \mathrm{Spec}_{ap}(T)$. $(T-\lambda)$ must be injective (else $(T-\lambda)u = (T-\lambda)v$ for some $u \neq v$ and so $u - v$ is an eigenfunction) and $\mathrm{Ran}(T-\lambda)$ must not be dense in $\mathcal{H}$ (else it is surjective and thus bijective, or can be extended to an operator which is surjective and thus bijective).

We can prove that for any operator T, $\mathrm{Ran}(T)^\perp = \mathrm{Ker}(T^*)$:

$$\begin{aligned} \text{for } u \in \mathrm{Ran}(T), v \in \mathrm{Ker}(T^*), (u,v) &= (Tw, v) \text{ for some } w \in \mathcal{H} \\ &= (w, T^*v) = (w, 0) = 0 \\ &\Rightarrow \mathrm{Ker}(T^*) \subseteq \mathrm{Ran}(T)^\perp; \\ x \in \mathrm{Ran}(T)^\perp &\Rightarrow (Aw, x) = 0, \\ &\Rightarrow (w, A^*x) = 0 \ \forall w \in \mathcal{H} \\ &\Rightarrow A^*x = 0 \\ &\Rightarrow \mathrm{Ran}(T)^\perp \subseteq \mathrm{Ker}(T^*). \end{aligned}$$

Then as there is some element η not in the closure of $\mathrm{Ran}(T-\lambda)$, by the Projection Theorem $\eta \in \mathrm{Ker}(T^* - \overline{\lambda})$; hence $(T^* - \overline{\lambda})\eta = 0$ and so η is an eigenvector for T^* with eigenvalue $\overline{\lambda}$.

Theorem 3.3. *[19] Let T be a normal operator (i.e. it commutes with its adjoint; $TT^* = T^*T$). Then*

$$\mathrm{Spec}(T) = \mathrm{Spec}_{ap}(T).$$

Proof. With this lemma proven, the result follows almost immediately: for a normal operator we have

$$\|Tu\|^2 = (Tu, Tu) = (T^*Tu, u) = (TT^*u, u) = (T^*u, T^*u) = \|T^*u\|^2$$

and so $\|(T-\lambda)u_n\| = \|(T^* - \overline{\lambda})u_n\|$, which means that $\mathrm{Spec}_{ap}(T) = \overline{\mathrm{Spec}_{ap}(T^*)}$. $\square$

This is fortunate, as most operators relevant to physical examples are normal (if not self-adjoint); we can give a constructive definition for the entire spectrum of a normal operator!

We now specialise to a subset of the approximate point spectrum, known as the essential spectrum. The essential spectrum has several definitions, the most popular usually denoted $\mathrm{Spec}_{e,i}$ for $i \in \{1, 2, 3, 4, 5\}$ in order of size. For most well-behaved operators the definitions are equivalent. This particular definition is known as Weyl's criterion, $\mathrm{Spec}_{e,2}$. [2]

Definition. ***(Essential spectrum)*** *The essential spectrum of an operator T on a Hilbert space $\mathcal{H}$ is defined as the set of all $\lambda \in \mathbb{C}$ such that a* ***Weyl sequence*** *u_n exists for T and λ , i.e. a sequence with the properties:*

- $\|u_n\| = 1 \quad \forall n \in \mathbb{N}$*;*
- $u_n \rightharpoonup 0$ *(where $\rightharpoonup$ denotes weak convergence: $u_n \rightharpoonup u \Leftrightarrow (u_n, g) \to (u, g) \quad \forall g \in \mathcal{H}$);*
- $\lim_{n\to\infty} \|(T-\lambda)u_n\| \to 0$.

It may not be immediately obvious from the definition that the essential spectrum is in the spectrum at all. In fact, we do not even require $u_n \rightharpoonup 0$ for this to be the case:

We can loosen the definition of weak convergence to just require convergence in a dense subspace of $\mathcal{H}$:

Lemma 3.4. *A bounded sequence u_n, $\|u_n\| \le C$, is weakly convergent to u in $\mathcal{H}$ if and only if it is weakly convergent to u in L where L is a dense subspace of $\mathcal{H}$.*

Proof. Weak convergence in $\mathcal{H}$ implying the same in L is obvious by the definition. Conversely, take $g \in \mathcal{H}$. For any $\varepsilon > 0$, we have $\|g - \varphi\| < \varepsilon$ for $\varphi \in L$; furthermore by the weak convergence of u_n in L we have $N \in \mathbb{N}$ such that $(u_n - u, \varphi) < \varepsilon$ for $n \ge N$. Then:

$$(u_n - u, g) = (u_n - u, g - \varphi + \varphi) = (u_n - u, g - \varphi) + (u_n - u, \varphi) < \|u_n - u\|\|g - \varphi\| + \varepsilon < \varepsilon(C + 1) \to 0.$$

$\square$

We will now construct a Weyl sequence for the essential spectrum of the 'Laplacian' or 'free Schrödinger' operator $T = -\Delta$ on $L^2(\mathbb{R})$, where on $\mathbb{R}^1$, Δ is the operator $\Delta f(x) = \frac{d^2}{dx^2} f(x)$

Example 5. *The essential spectrum of the operator $T = -\Delta$ is the closed half-axis $[0, +\infty)$.*

Proof. First, note that for the exponential function we have

$$T \exp(i\omega x) = \omega^2 \exp(i\omega x). \tag{3}$$

This gives much of the intuition for this proof; the function $\exp_{i\omega} : x \mapsto \exp(i\omega x)$ is not an eigenvector as it is not in $L^2(\mathbb{R})$, but it satisifes the eigenvalue equation for T and so any number $\lambda = \omega^2$ - and thus any $\lambda \in [0, +\infty)$ - is 'almost' an eigenvalue for T.

We take advantage of this by choosing some smooth bump function $\rho \in C_c^\infty(\mathbb{R})$ with $\|\rho\|_2 = 1$. We then define $\rho_n = \frac{1}{\sqrt{n}}\rho(x/n)$. ρ_n has some nice properties: by a substitution of variables and direct calculation we have $\|\rho_n\|_2 = \|\rho\|_2$, and furthermore any k'th derivative $\rho_n^{(k)}$ of ρ_n converges to 0 in L^2. Indeed:

$$\|\rho_n^{(k)}\|_2 = \frac{1}{n^k}\|\frac{1}{\sqrt{n}}\rho^{(k)}(x/n)\|_2 = \frac{\|\rho^{(k)}\|_2}{n^k} \to 0 \tag{4}$$

where one can see $\|\frac{1}{\sqrt{n}}\rho^{(k)}(x/n)\|_2 = \|\rho^{(k)}\|_2$ by the same calculation as $\|\rho_n\|_2 = \|\rho\|_2$.

Now, let our candidate Weyl sequence be $u_n : x \mapsto \rho_n(x)\exp(i\omega x)$, which truncates $\exp(i\omega x)$ to $\mathrm{supp}\rho_n$; this means u_n is in $L^2(\mathbb{R})$. $\|u_n\| = \|\rho_n\|_2 = \|\rho\|_2 = 1$ by direct calculation, and $u_n \rightharpoonup 0$: we can bound u_n

by $\frac{1}{\sqrt{n}}M\mathbf{1}_{(\mathrm{supp}u_n)}$, where $\mathbf{1}_A$ is the characteristic function of the set A and M is the maximum value of ρ. Then by Lemma 3.4, we can simply show weak convergence for any $\varphi \in C_0^\infty$, which is dense in L^2:

$$\begin{aligned}
(u_n, \varphi) &= \int_{\mathbb{R}} u_n \varphi \\
&\leq \int_{\mathbb{R}} \frac{1}{\sqrt{n}} M \mathbf{1}_{(\mathrm{supp}u_n)} \varphi \\
&\leq \int_{\mathrm{supp}\varphi} \frac{1}{\sqrt{n}} M \varphi \\
&= \frac{M}{\sqrt{n}} \int_{\mathrm{supp}\varphi} \varphi \to 0, && \text{as the integral of } \varphi \text{ is finite and independent of } n.
\end{aligned}$$

Finally, we show that $\lim_{n\to\infty} \|(T-\lambda)u_n\|_2 \to 0$ for $\lambda = \omega^2$:

$$\begin{aligned}
\|(T-\lambda)u_n\|_2 &= \|(T(\exp_{i\omega}\rho_n) - \omega^2(\exp_{i\omega}\rho_n)\|_2 \\
&= \|(T(\exp_{i\omega}\rho_n) - T(\exp_{i\omega})\rho_n\|_2 && \textit{(by equation (3))} \\
&= \|\exp_{i\omega} T\rho_n - 2\omega \exp_{i\omega} \frac{d}{dx}\rho_n\|_2 && \textit{(by the product rule)} \\
&= \|T\rho_n - 2\omega \frac{d}{dx}\rho_n\|_2 && \textit{(see } \|\exp_{i\omega}\phi\|_2 = \|\phi\|_2 \textit{ for any } \phi \in L^2\textit{)} \\
&\leq \| - \frac{d^2}{dx^2}\rho_n\|_2 + 2\omega\|\frac{d}{dx}\rho_n\|_2 \to 0,
\end{aligned}$$

converging by equation (4). Thus u_n forms a Weyl sequence for T and $\lambda \in [0, +\infty)$, as required. □

We can use a similar idea for another example to find the essential spectrum of the multiplication operator:

Example 6. *Consider $f \in C(0,1) \cap L^\infty(0,1)$. The essential spectrum of the operator M_f on $L^2(0,1)$, where $M_f u(x) = f(x)u(x)$, is the range of f.*

Proof. Similar to before, our initial idea comes from an 'almost-eigenvector'. In this case, if λ is in the range of f with $f(x_0) = \lambda$, we see that $M_f\delta_{x_0} = \lambda\delta_{x_0}$, where δ_{x_0} is the Dirac delta centred at x_0. Again, δ_{x_0} is not an eigenfunction of M_f as it is not in the correct domain - this time, it isn't even strictly a function (it is a distribution).

Now consider a Friedrichs mollifier ρ. This is a function in $C_0^\infty(\mathbb{R})$ with the property that $\sqrt{n}\rho(ny) \to \delta_0$ as $n \to \infty$; we renormalise it such that $\|\rho\|_2 = 1$, and take the sequence

$$u_n : x \mapsto \begin{cases} \sqrt{n}\rho(n(x-x_0)) & x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

thus u_n "converges to δ_{x_0}" in the sense of distributions. Note that $\|u_n\|_2 = \|\rho\|_2 = 1$ for all n, and this sequence converges weakly to 0:

$$\begin{aligned}
|(u_n, g)| &= \int_{\mathrm{supp}u_n} \sqrt{n}\rho(n(x-x_0))g(x) && \textit{(for any } g \in L^2(0,1)\textit{)} \\
&\leq \|u_n\|_2 \sqrt{\int_{\mathrm{supp}u_n} |g(x)|^2} && \textit{(by Hölder's inequality)} \\
&= \sqrt{\int_{\mathrm{supp}u_n} |g(x)|^2} \to 0, && \text{as } \mathrm{supp}(u_n) \text{ decreases to } 0.
\end{aligned}$$

Then we see $\|(M_f - \lambda)u_n\|_2$ converges to zero by similar reasoning:

$$\begin{aligned}
\|(M_f - \lambda)u_n\|_2^2 &= \int_{\mathrm{supp}\rho_n} |(f(x) - f(x_0)\sqrt{n}\rho(n(x-x_0))|^2 && \textit{(using that } \lambda = f(x_0)\textit{)} \\
&= \|(f(x) - f(x_0))^2\|_{L^\infty(\mathrm{supp}\rho_n)} \|\rho_n^2\|_1 && \textit{(by Hölder's inequality)} \\
&= \sup_{x \in \mathrm{supp}\rho_n} \|(f(x) - f(x_0))^2\| \to 0 && \textit{(note } \|\rho_n^2\|_{L^1} = \|\rho_n\|_2 = 1\textit{)}
\end{aligned}$$

converging to zero as $\mathrm{supp}\rho_n$ shrinks around x_0 by the continuity of f. □

Compare this example to Theorem 2.2, and see that the *entire spectrum* of a multiplication operator is essential spectrum; the operator has no eigenvalues.

One interesting property of the essential spectrum that is not easily visible from our earlier definition is that it is invariant under compact perturbations. This is not the case for eigenvalues!

Definition. ***(Compact operators and rank)*** *An operator T on a normed vector space X is* compact *if for every bounded sequence $(x_n)_{n\in\mathbb{N}}$ in X, the sequence $(Tx_n)_{n\in\mathbb{N}}$ has a convergent subsequence.*

The rank of an operator T, denoted $\mathrm{rank}T$*, is the dimension of its range.*

Note in particular that if a bounded operator T has finite rank, then T is compact; as its image is finite-dimensional and bounded, the Bolzano-Weierstrass theorem holds for $(Tx_n)_{n\in\mathbb{N}}$.

Note also that any compact operator is necessarily bounded, as otherwise we could choose a bounded sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ such that $\|Tx_n\| \to \infty$, and then it would not be possible for $(Tx_n)_{n\to\infty}$ to have a bounded subsequence.

Theorem 3.5. *Let λ be in the essential spectrum of an operator T on a Hilbert space $\mathcal{H}$. Then*

$$\lambda \in \bigcap_{K \in \mathcal{K}(\mathcal{H})} \mathrm{Spec}_e(T+K) \tag{5}$$

where $\mathcal{K}(\mathcal{H})$ is the space of all compact linear operators on $\mathcal{H}$.

Proof. First, let $\lambda \in \mathrm{Spec}_e(T)$ have the Weyl sequence x_n for the operator T. Then

$$\|(T+K-\lambda)x_n\| = \|(T-\lambda)x_n + Kx_n\| \le \|(T-\lambda)x_n\| + \|Kx_n\|$$

.

And as K is compact, Kx_n has a convergent subsequence Kx_{n_k}, where $x_{n_k} \rightharpoonup 0$ because $x_n \rightharpoonup 0$. Then Kx_{n_k} also weakly converges to 0 by two applications of the Riesz representation theorem and the boundedness of K: for any $\phi \in \mathcal{H}$,

$$\begin{aligned}
(Kx_{n_k}, \phi) &= f(Kx_{n_k}) && \textit{for some bounded linear functional } f \in \mathcal{H}^* \\
&= (f \circ K)(x_{n_k}) && \textit{which is also a bounded linear functional in } \mathcal{H}^* \\
&= (x_{n_k}, \psi) && \textit{for some } \psi \in \mathcal{H} \\
&\to 0 && \textit{by weak convergence of } x_{n_k} \textit{ to 0}
\end{aligned}$$

and as it is (strongly) covergent to some value K, it must also be weakly convergent to the same value; thus $Kx_{n_k} \to 0$.

Thus $\|(T+K-\lambda)x_{n_k}\| \le \|(T-\lambda)x_{n_k}\| + \|Kx_{n_k}\| \to 0$, so x_{n_k} is a Weyl sequence for λ and $T+K$ for any compact operator K, so $\lambda \in \bigcap_{K\in\mathcal{K}(\mathcal{H})} \mathrm{Spec}_e(T+K)$. □

Remark. *As mentioned before, eigenvalues do not have this property: let λ be an eigenvalue of T with eigenvector u, and P the orthogonal projection onto the space* $\mathrm{span}\{u\}$ *(so* $\mathrm{rank}P = 1$*). Then:*

$$(T+P)u = Tu + u = \lambda u + u = (\lambda+1)u$$

so λ is not an eigenvalue of $(T+P)$. This will become a very useful property in discussing a method of detecting spectral pollution known as dissipative barrier methods*, which we will explore in section 4.2.*

3.2 Rayleigh quotients and numerical range

The gateway to bounding the spectrum (and pollution) of an operator T on a Hilbert space lies in a functional known as the **Rayleigh quotient**, $R_T : \mathrm{Dom}(T) \to \mathbb{C}$, defined:

$$R_T : u \mapsto \frac{(Tu,u)}{(u,u)}$$

or equivalently (by linearityc) $R_T : u \mapsto (Tu,u)$ on the domain $\{u \in \mathrm{Dom}(T) : \|u\| = 1\}$.

For operators where we want to weaken the domain (e.g. a differential operator) it is suitable to replace (Tu,u) with the relevant bilinear form $\mathcal{A}[u,u]$.

Definition. ***(Numerical range of an operator)*** *Let T be an operator on a Hilbert space. The numerical range $\mathrm{W}(T)$ is defined $\mathrm{W}(T) := \mathrm{Ran}(R_T)$.*

The numerical range has a variety of interesting properties which make them useful for roughly approximating the location of spectra.

Proposition 3.6. *The numerical range $\mathrm{W}(T)$ of an operator T has the following properties:*

1. $\mathrm{W}(T) \in \mathbb{R}$ *if T is self-adjoint;*
2. $\mathrm{W}(T_{\mathcal{L}}) \subseteq \mathrm{W}(T)$*, where $T_{\mathcal{L}}$ is the compression of T to the closed subspace $\mathcal{L}$;*
3. *(Toeplitz-Hausdorff theorem)* $\mathrm{W}(T)$ *is a convex set;*
4. $\mathrm{Spec}_{ap}(T) \subseteq \overline{\mathrm{W}(T)}$*, where $\overline{\mathrm{W}(T)}$ is the closure of the numerical range of T.*

It is important to state the usefulness of these properties with regards to spectral pollution. Not only does $\mathrm{W}(T)$ bound the spectrum of T, it bounds the spectrum of $T_{\mathcal{L}}$ - effectively, bounding the region in which spectral pollution can occur to a convex set around $\mathrm{Spec}(T)$. We will also use this fourth property to derive a theorem on how well the Galerkin method approximates the spectrum outside of $\mathrm{conv}(\mathrm{Spec}_{ess})$.

Proof. (1.) If T is self-adjoint, $(Tu,u) = (u,Tu)$ for all u; by conjugate symmetry of scalar products, $(u,Tu) = \overline{(Tu,u)}$; we combine these to find $(Tu,u) = \overline{(Tu,u)}$ and the result follows.

(2.) $\mathrm{W}(T_{\mathcal{L}}) = \{(PTPu,u) : u \in \mathcal{L}, \|u\| = 1\}$. We then use the self-adjointness of P to see

$$(PTPu,u) = (TPu,Pu)$$

Then as $u \in \mathcal{L}$, $\|Pu\| = \|u\| = 1$, so $(T(Pu),(Pu)) \in \mathrm{W}(T)$, and the result follows.

(3. [20]) Take $\lambda = (Tx,x), \mu = (Ty,y) \in \mathrm{W}(T)$. Define the line segment between them as $\nu = t\lambda + (1-t)\mu$ for $t \in [0,1]$, and $T_{\mathcal{L}}$ the compression of T to the subspace $\mathcal{L} = \mathrm{span}\{x,y\}$. Then we note that $(T_{\mathcal{L}}x,x) = (Tx,x)$ and $(T_{\mathcal{L}}y,y) = (Ty,y)$, so λ, μ are in $\mathrm{W}(T_{\mathcal{L}})$. $T_{\mathcal{L}}$ is two-dimensional, so is a 2×2 matrix; it can be proven by direct calculation (see [20]) that the numerical range of a 2×2 matrix is an ellipse (with foci at either eigenvalue of the matrix!) and so ν is in $\mathrm{W}(T_{\mathcal{L}})$. Then by property 2, $\mathrm{W}(T_{\mathcal{L}}) \subseteq \mathrm{W}(T)$, so ν is also in $\mathrm{W}(T)$, as required.

(4.) $\overline{\mathrm{W}(T)}$ is the set of all points η such that there is a sequence of unit vectors u_n where

$$\lim_{n\to\infty} (Tu_n, u_n) = \eta.$$

Now let $\lambda \in \mathrm{Spec}_{ap}(T)$. We can combine the approximate eigenvalue theorem (Theorem 3.1) with the Cauchy-Schwarz inequality:

$$|((T-\lambda)u_n, u_n)| \le \|(T-\lambda)u_n\| \to 0, \text{ and so}$$

$$\begin{aligned} |((T-\lambda)u_n, u_n)| &= |((Tu_n,u_n)) - (\lambda u_n, u_n)| \\ &= |((Tu_n,u_n)) - \lambda\|u_n\|^2| \\ &= |((Tu_n,u_n)) - \lambda| \to 0; \\ &\Rightarrow (Tu_n,u_n) \to \lambda. \end{aligned}$$

So $\lambda \in \overline{\mathrm{W}(T)}$. ∎

Corollary 3.7. *If T is a normal operator (i.e. $TT^* = T^*T$), then its entire spectrum is in the numerical range.*

Proof. Combine Theorem 3.3 with Theorem 3.6.4. ☐

Corollary 3.8. *In particular, if T is self-adjoint,* $\mathrm{Spec}(T) \subseteq \mathbb{R}$. *Furthermore, if T is bounded, then we have*

$$\inf(\mathrm{Spec}(T)) = \inf(\mathrm{W}(T)), \textit{ and}$$
$$\sup(\mathrm{Spec}(T)) = \sup(\mathrm{W}(T)).$$

Proof. We can see immediately that any self-adjoint operator is normal ($T = T^* \Rightarrow T^*T = TT = TT^*$); thus the entire spectrum of the self-adjoint operator is in the numerical range, and by Proposition 3.6.1 we have $\mathrm{Spec}(T) \subseteq \overline{\mathrm{W}(T)} \subseteq \mathbb{R}$.

Now, $\inf(\mathrm{Spec}(T)) \geq \inf(\mathrm{W}(T))$. Let $\inf(\mathrm{W}(T)) = w_0$; then for any unit vector u, $((T - w_0)u, u) = (Tu, u) - w_0 \geq 0$, and so $u, v \mapsto ((T - w_0)u, v)$ defines a positive-semidefinite Hermitian form, for which the Cauchy-Schwarz inequality holds[5]. We then find the following bound for unit vectors u, v:

$$|\tau[u, v]|^2 \leq \tau[u, u]\tau[v, v] = ((T - w_0)u, u)((T - w_0)v, v) \leq \|T - w_0\|((T - w_0)u, u).$$

Now let us take a minimising sequence u_n, $\|u_n\| = 1$ such that $(Tu_n, u_n) \to w_0$. Then we have

$$\|(T - w_0)u_n\|^2 = |\tau[u_n, (T - w_0)u_n]|^2 \leq \|T - w_0\|((T - w_0)u_n, u_n) \leq \|T - w_0\||(Tu_n, u_n) - w_0| \to 0,$$

and therefore $w_0 \in \mathrm{Spec}(T)$ by Theorem 3.1. The proof for the supremum w_1 is analogous with some sign changes. ☐

This corollary extends to self-adjoint **semibounded** operators, which are operators such that their Rayleigh quotient is bounded above or below by some constant c - the result holds for the supremum or infimum for operators which are semibounded above or below respectively. We omit this more general proof as we would require a significant tangent to acquire the prerequisite results: see e.g. ([21], Corollary 1.11). Moreover, note that a bounded operator is semibounded both above and below.

Corollary 3.9. *Let T be a self-adjoint, semibounded operator. Then spectral pollution does not occur outside of* $\mathrm{convSpec}(T)$.

Proof. Combine Corollary 3.8 with Proposition 3.6.2. ☐

Is there a better bound? From heuristic evidence we may expect one - if we look at Example 1, we can see that there is no pollution even within the negative semiaxis, even though the region plotted is well within $\mathrm{convSpec}(T)$. The numerical range has been refined in a variety of ways, one of which is particularly profitable when it comes to bounding spectral pollution. This shall be our next topic.

3.3 Essential numerical range

A similar notion to that of the numerical range is the essential numerical range, $\mathrm{W}_e(T)$. This set lowers its aim to simply estimating the essential spectrum, but in the process manages to do so much more accurately for some operators.

Definition. ***(Essential numerical range)*** *(adapted from [22]) The essential numerical range of an operator T is given by*[6]

$$\mathrm{W}_e(T) := \{\lim_{n\to\infty}(Tu_n, u_n) : (u_n)_{n\in\mathbb{N}} \textit{ in } \mathrm{Dom}(T), \|u_n\| = 1, u_n \rightharpoonup 0.\}$$

Note the parallels with our definition of the essential spectrum, $\mathrm{Spec}_e(T)$. Indeed, these parallels are reflected in the properties of $\mathrm{W}_e(T)$:

[5] This can easily be verified by looking at a standard proof for the inequality.

[6] Much like the essential spectrum, there are multiple definitions of the essential numerical range. However (at least for bounded operators) there is much more equivalence between the definitions than we have for essential spectrum! [22] We choose the definition with the most natural relation to our choice of definition for essential spectrum.

Proposition 3.10. *The essential numerical range* $\mathrm{W}_e(T)$ *of an operator* T *has the following properties:*

1. $\mathrm{W}_e(T)$ *is convex;*
2. $\mathrm{W}_e(T) \subseteq \overline{\mathrm{W}(T)}$*;*
3. $\mathrm{conv}(\mathrm{Spec}_e(T)) \subseteq \mathrm{W}_e(T)$*, with equality if* T *is self-adjoint and bounded.*

Proof. (1. [23]) We prove this by applying the Toeplitz-Hausdorff theorem (Proposition 3.6.3) to a sequence. We take $\lambda = \lim_{n\to\infty}(Tx_n, x_n), \mu = \lim_{n\to\infty}(Ty_n, y_n) \in \mathrm{W}_e(T)$ and define a sequence

$$\nu_n = t(Tx_n, x_n) + (1-t)(Ty_n, y_n) \quad t \in [0,1],$$

which obviously converges to $\nu = t\lambda + (1-t)\mu$ as $n \to \infty$. We then create a sequence of compressions T_n, where each compression is to $\mathrm{span}\{x_n, y_n\}$. By the Toeplitz-Hausdorff theorem, we know that ν_n is in $\mathrm{Spec}(T_n)$ and get a sequence $\nu_n = (Tz_n, z_n)$ converging to ν; the elements z_n are unit vectors in $\mathrm{span}\{x_n, y_n\}$, and they weakly converge to 0 because x_n and y_n both do:

$$(z_n, g) = (\alpha x_n + \beta y_n, g) = \alpha(x_n, g) + \beta(y_n, g) \to 0 \quad \forall g \in \mathcal{H}.$$

This means that $\nu = \lim_{n\to\infty}(Tz_n, z_n)$ is in $\mathrm{Spec}_e(T)$ as required.

(2.) This can be seen directly from looking at the two definitions. By definition, we have

$$\overline{\mathrm{W}(T)} = \{lim_{n\to\infty}(Tu_n, u_n) : (u_n)_{n\in\mathbb{N}} \text{ in } \mathrm{Dom}(T), \|u_n\| = 1\},$$

and $\mathrm{W}_e(T)$ is the subset of this with the extra condition that $u_n \rightharpoonup 0$.

(3.) The inclusion $\mathrm{Spec}_e(T) \subseteq \mathrm{W}_e(T)$ comes from an analogous argument to that of Proposition 3.6.4; then $\mathrm{conv}(\mathrm{Spec}_e(T)) \subseteq \mathrm{W}_e(T)$ by this inclusion and that $\mathrm{W}_e(T)$ is a convex set. ☐

It remains to show that $\mathrm{conv}(\mathrm{Spec}_e(T)) = \mathrm{W}_e(T)$ when T is self-adjoint and bounded. This will be proven after the following theorem, which describes another similarity with essential spectra; invariance under compact perturbation.

Theorem 3.11. *A value* λ *is in the essential numerical range of an operator* T *on a Hilbert space* $\mathcal{H}$ *if and only if*

$$\lambda \in \bigcap_{K\in\mathcal{K}(\mathcal{H})} \overline{\mathrm{W}(T+K)}$$

where $\mathcal{K}(\mathcal{H})$ *is the set of all compact linear operators on* $\mathcal{H}$*.*

Proof. Let $\lambda = \lim_{n\to\infty}(Tx_n, x_n)$ be in $\mathrm{W}_e(T)$. Then $((T+K)x_n, x_n) = (Tx_n, x_n) + (Kx_n, x_n)$ which converges to λ if $(Kx_n, x_n) \to 0$, which is true by the weak convergence of x_n (using that K is bounded as it is compact). Thus for any compact operator K, $\lambda \in \mathrm{W}_e(T+K) \subseteq \overline{\mathrm{W}(T+K)}$.

Conversely, ☐

We have seen that the essential numerical range estimates the bounds of the essential spectrum with quite astounding accuracy for some types of operator. But the essential numerical range far outdoes the regular numerical range on bounding spectral pollution; in fact, it provides an *exact* set on which it is possible for pollution to occur!

Theorem 3.12. *Let* T *be a bounded operator. All spectral pollution in the Ritz approximation of* $\mathrm{Spec}(T)$ *will be located inside of* $\mathrm{W}_e(T)$*; within this set, it can occur anywhere in* $\mathrm{W}_e(T) \setminus \mathrm{Spec}(T)$*.*

Proof. For some sequence of truncations T_i, let $\lambda_i \in T_i$ converge to λ. Then we note that as T_i is finite-dimensional, all of its spectral points are eigenvalues; for some eigenvector w_i, $T_i w_i = \lambda_i w_i$, and (with w_i normalised to $\|w_i\| = 1$ we have

$$(T_i w_i, w_i) = (\lambda_i w_i, w_i) = \lambda_i \|w_i\| = \lambda_i.$$

This means $\lambda_i \in \mathrm{W}(T_i) \subseteq \mathrm{W}(T)$, so $\lambda_i = (Tx_i, x_i)$ for some unit vector x_i, and

$$\lambda = \lim_{i\to\infty} \lambda_i = \lim_{i\to\infty} (Tx_i, x_i).$$

Then, for λ to be spectral pollution outside of the essential numerical range, we must have that $x_i \rightharpoonup x \neq 0$ (as otherwise $\lambda \in \mathrm{W}_e(T)$) and also $\|(T-\lambda)x_i\| \nrightarrow 0$ (as then λ is in the spectrum of T by the corollary to Theorem 3.1). $\square$

For a bounded operator, this is the main result of a paper by Pokrzywa [24]. The main theorem of the paper has the corollary that for $\lambda \notin \mathrm{W}_e(T)$, we have $\lambda \in \mathrm{Spec}(T)$ iff $\mathrm{dist}(\lambda, \mathrm{Spec}(T_n)) \to 0$; that is, outside of the essential numerical range, every point in the approximate spectrum $\mathrm{Spec}(T_n)$ converges to a point in the actual spectrum of T. This is followed by a lemma which claims that for any sequence $(\lambda_n)_{n\in\mathbb{N}}$ in the interior of $\mathrm{W}_e(T)$, there is a sequence of orthogonal projections such that $\lambda_{n-1} \in \mathrm{Spec}(T_n)$ - not only does all spectral pollution occur inside this range, but for *any point* in $\mathrm{W}_e(T) \setminus \mathrm{Spec}(T)$, spectral pollution occurs there in some approximation.

IV | DETECTING AND AVOIDING SPECTRAL POLLUTION

4.1 Properties of eigenfunctions

If there are only a small number of eigenvalues which we would like to learn more about (e.g. once we have whittled down some of the pollution via a dissipative barrier), it may be useful to then look at the eigenfunctions for each relevant value. Polluting eigenvalues have rather interesting-looking eigenfunctions (Figure **??**)

4.2 Dissipative barrier methods

Perhaps one of the simplest methods for separating eigenvalues from spectral pollution is the dissipative barrier method. This method leverages finite-rank perturbation of an operator.

Take an operator T on a Hilbert space, and let P be an orthonormal projection onto a finite-dimensional space such that for an eigenvector u, $\|Pu - u\|$ is sufficiently small. Let λ be the eigenvalue for u. Then for the operator $T + iP$,

$$(T + iP)u = Tu + iPu \approx \lambda u + iu = (\lambda + i)u.$$

i.e. $(\lambda + i)$ is approximately an eigenvalue for $T + iP$.

Then, the eigenvalues of the perturbed operator will have imaginary part of approximately 1, and the set of eigenvalues produced by the approximation can be filtered to discard points without imaginary part close to 1 as being pollution. This is particularly useful for self-adjoint operators, where the entire spectrum and the essential numerical range are a subset of the real line; in that case, the perturbed eigenvalues will be the only points on the spectrum with significant imaginary part. Even better, if the operator is also bounded, then all pollution is in the essential numerical range (Theorem **??**) which is invariant under compact perturbation (Theorem **??**) so the pollution will converge to the real axis.

Let us see this concretely with an example.

Example 7. *We return once again to our discontinuous multiplication operator M_f on $L^2(0,1)$ with symbol*

$$f : x \mapsto \begin{cases} x & x < 1/2 \\ x + 1/2 & \textit{otherwise.} \end{cases}$$

This time, we add a 'rank-one perturbation' to get the operator $\tilde{M}$ which has the action

$$\tilde{M}u = M_f u + (u, \varphi)\varphi,$$

where φ is a real-valued function on $L^2(0,1)$. Then $\tilde{M}$ has an eigenvector: we rearrange to get

$$\begin{aligned} f(x)u(x) + (u,\varphi)\varphi(x) &= \lambda u(x), \\ \textit{therefore } (\lambda - f(x))u(x) &= (u,\varphi)\varphi(x), \\ u(x) &= c\frac{\varphi(x)}{\lambda - f(x)} \end{aligned}$$

for some constant c. We then normalise this to $\frac{\varphi(x)}{\lambda - f(x)}$, which requires

$$\begin{aligned} f(x)u + (u,\varphi)\varphi &= \lambda u \\ (f(x) - \lambda)c\frac{\varphi(x)}{\lambda - f(x)} + \left(\frac{c\varphi}{\lambda - f}, \varphi\right)\varphi(x) &= 0 \\ -1 + \left(\frac{\varphi}{\lambda - f}, \varphi\right) &= 0 \end{aligned}$$

so we can normalise provided that $(\frac{\varphi}{\lambda - f}, \varphi) = \int_0^1 \frac{|\varphi|^2}{\lambda - m} = 1$.

Thus for any value λ we can choose $\varphi \in L^2(0,1)$ and scale it so that $\int_0^1 \frac{|\varphi|^2}{\lambda - m} = 1$ to get an operator $\tilde{M}$ with spectrum $\mathrm{Spec}(M) \cup \{\lambda \in \mathbb{C} : \int_0^1 \frac{|\varphi|^2}{\lambda - m} = 1\}$. In Figure 3 we can see the results of Ritz approximations with φ chosen such that the operator has an eigenvalue at 0.7.[7] In Figure 4 we see the same approximation but with the dissipative barrier iP where P is the projection onto $\mathrm{span}\{\phi_n, |n| \leq 25\}$. Note that the spectrum on the line $(x, y) : \mathrm{Imag}(x) = 1$ converges to what we'd expect the spectrum to be!

[7] In particular, φ was chosen to be the constant $(\log(3) - 3\log(2) + \log(5) + \log(7) - \log(10))^{-1}$; this satisfies the normalisation condition at $\lambda = 0.7$ but also at $\lambda \approx 4.4$; the eigenvalue at $\lambda \approx 4.4$ has been cropped out of the figure to improve illustration of the idea.

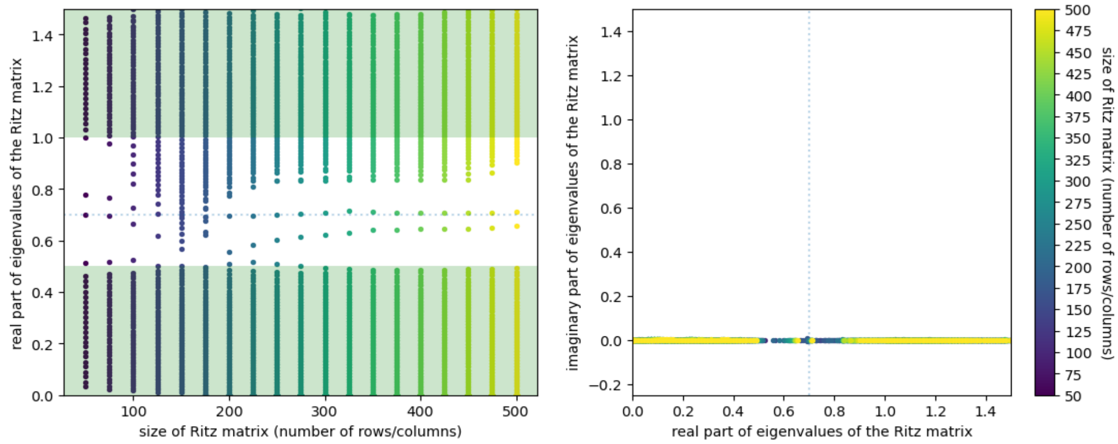

Figure 3: The real part of the approximate spectrum for $\tilde{M}$; on the left, the real parts of the approximate spectrum as the size of the Ritz matrix increases; on the left, the complex approximate spectrum where colour is used to donate the size of the approximation. The green shaded regions correspond to the essential spectrum of $\tilde{M}$, and the dotted lines are at $\mathrm{Re}(x) = 0.7$ to show where the added eigenvalue should be. (An intuitive way to view these figures is to see them as a three-dimensional plot, with the left figure 'top-down', and the right figure 'from the east')

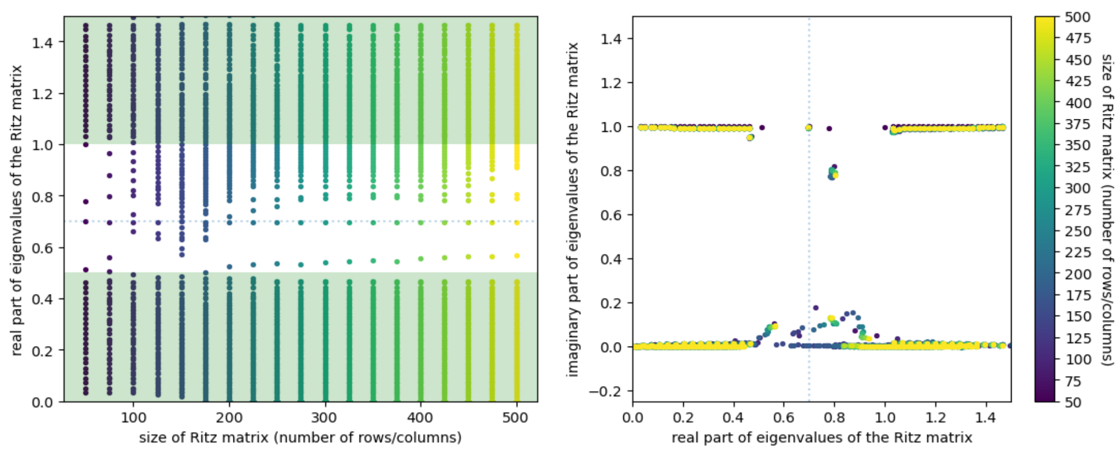

Figure 4: The real part of the approximate spectrum for $\tilde{M} + iP$; compare with Figure 3. See that the line at 1.0 on the imaginary axis converges to the actual spectrum of the operator, while the pollution remains below.

Remark. *One may also note that there are bands corresponding to the essential spectrum with imaginary part 1. The reason why dissipative barriers 'replicate' the essential spectrum is an open problem; it has recently been investigated specifically for Schrödinger operators [***stepanenkoTODO***] but in general remains unknown.*

As a second example, let us try to replicate some results from Aceto et al. (2006) [14]. In this paper, they use an algebraic method combined with a 'shooting technique' to find high-accuracy estimates of eigenvalues for a Sturm-Liouville operator; this highly-specialised algorithm is free from pollution but works only for a specific class of Sturm-Liouville operators. is a

Example 8. *In particular, take the following eigenvalue problem on* $L^2[0,\infty)$*:*

$$\begin{cases} -y'' + (\sin(x) - \frac{40}{1+x^2})y = \lambda y \\ y(0)\cos(\pi/8) + y'(0)\cos(\pi/8) = 0. \end{cases}$$

This operator has a 'band-gap' structure; it has intervals (bands) of essential spectrum, with eigenvalues dotted in the gaps between bands. In two of the spectral gaps $J_2 = (-0.34767, 0.59480)$ *and* $J_3 = (0.91806, 1.2932)$ *(denoted in line with the paper and rounded to 5sf) the algebraic method finds the following eigenvalues:*

J_2	J_3
0.33594	0.94963
0.53662	1.2447
0.58083	1.2919
0.59150	

Firstly, we will see whether we can reproduce this data. The algebraic method used first truncates the half-line to $[0, 70\pi]$*. We will do the same and perform a Ritz approximation on the truncated interval, applying a dissipative barrier; this is a success, as one can see in Figure 5.*

Now we will aim for better; whether we can reproduce these eigenvalues using a Ritz approximation on $[0,\infty)$*, with the orthonormal basis* $\{\phi_n\}_{n\in\mathbb{N}}, \phi_n = \exp(-x/2)L_n$*, where* L_n *is the n'th Laguerre polynomial (see [12] for a proof that this is indeed an orthonormal basis)*

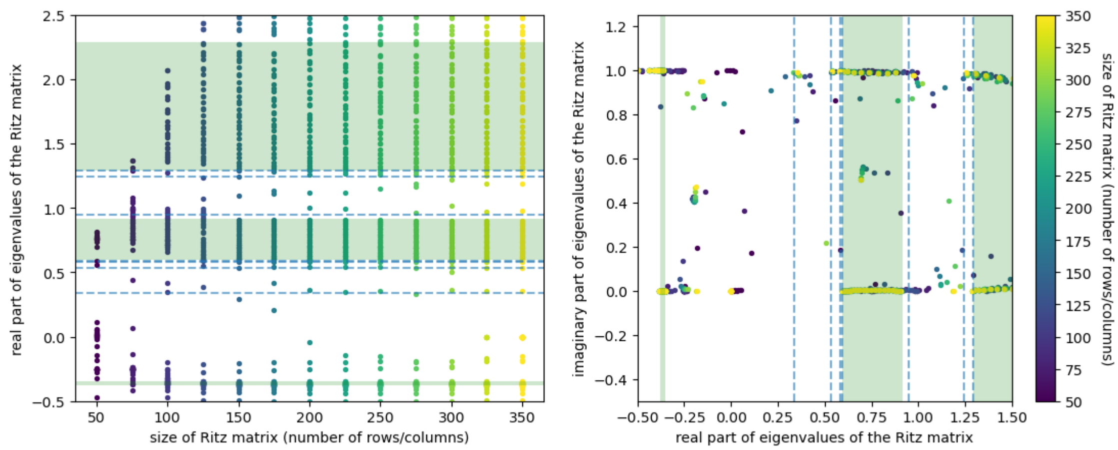

Figure 5: The results of truncating the domain and applying a dissipative barrier to the operator of Example 8. The green bands represent the essential spectrum of the operator, and the dotted blue lines indicate where the algebraic method found eigenvalues in two spectral gaps. Note that for larger approximations, the points in the spectral gaps with imaginary part 1.0 are very close to the algebraic method's approximation.

4.3 Supercell methods

Supercell methods are a type of truncation method which can avoid spectral pollution entirely for certain kinds of operators, via a crystallography-inspired approach. We will discuss supercell methods through two examples.

4.3.1 The Feinberg-Zee random hopping model

Feinberg and Zee (1999) [**feinberg1999nonhermitian**] models an unusual behaviour in superconductors via an interesting class of non-self-adjoint and random operators on $\ell^2(\mathbb{Z}$: matrices of the form

$$\begin{pmatrix} \ddots & \ddots & & & & \\ \ddots & 0 & 1 & & & \\ & c_{n-1} & 0 & 1 & & \\ & & c_n & 0 & 1 & \\ & & & c_{n+1} & 0 & \ddots \\ & & & & \ddots & \ddots \end{pmatrix}$$

4.3.2 The 'Ten Martini Problem'

References

[1] William Arveson. *A Short Course on Spectral Theory*. Vol. 209. Graduate Texts in Mathematics. New York, NY: Springer, 2002. ISBN: 978-1-4419-2943-3 978-0-387-21518-1. DOI: 10.1007/b97227. (Visited on 11/15/2023).

[2] David Edmunds and Des Evans. *Spectral Theory and Differential Operators*. Oxford University Press, May 2018. ISBN: 978-0-19-186113-0. DOI: 10.1093/oso/9780198812050.001.0001. (Visited on 11/23/2023).

[3] Brian C. Hall. *Quantum Theory for Mathematicians*. Vol. 267. Graduate Texts in Mathematics. New York, NY: Springer, 2013. ISBN: 978-1-4614-7115-8 978-1-4614-7116-5. DOI: 10.1007/978-1-4614-7116-5. (Visited on 01/25/2024).

[4] Endre Süli and David F. Mayers. *An Introduction to Numerical Analysis*. Cambridge University Press, Aug. 2003. ISBN: 978-0-521-00794-8.

[5] Mathieu Lewin and Eric Séré. "Spectral Pollution and How to Avoid It (With Applications to Dirac and Periodic Schr\"odinger Operators)". In: *Proceedings of the London Mathematical Society* 100.3 (May 2010), pp. 864–900. ISSN: 00246115. DOI: 10.1112/plms/pdp046. arXiv: 0812.2153 [math-ph]. (Visited on 11/29/2023).

[6] M. Lisa Manning, B. Bamieh, and J. M. Carlson. *Descriptor Approach for Eliminating Spurious Eigenvalues in Hydrodynamic Equations*. Dec. 2008. DOI: 10.48550/arXiv.0705.1542. arXiv: 0705.1542 [physics]. (Visited on 11/29/2023).

[7] Eric Cancès, Virginie Ehrlacher, and Yvon Maday. "Periodic Schrödinger Operators with Local Defects and Spectral Pollution". In: *SIAM Journal on Numerical Analysis* 50.6 (Jan. 2012), pp. 3016–3035. ISSN: 0036-1429. DOI: 10.1137/110855545. (Visited on 11/29/2023).

[8] E. Brian Davies. *Spectral Theory and Differential Operators*. Cambridge Studies in Advanced Mathematics. Cambridge: Cambridge University Press, 1995. ISBN: 978-0-521-58710-5. DOI: 10.1017/CBO9780511623721. (Visited on 11/15/2023).

[9] John Derwent Pryce. *Numerical Solution of Sturm-Liouville Problems*. Clarendon Press, 1993. ISBN: 978-0-19-853415-0.

[10] Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. New York, NY: Springer, 2011. ISBN: 978-0-387-70913-0 978-0-387-70914-7. DOI: 10.1007/978-0-387-70914-7. (Visited on 01/05/2024).

[11] E. B. Davies and M. Plum. *Spectral Pollution*. Feb. 2003. DOI: 10.48550/arXiv.math/0302145. arXiv: math/0302145. (Visited on 01/05/2024).

[12] Gábor Szegö. *Orthogonal Polynomials*. American Mathematical Society, 1975. ISBN: 978-0-8218-1023-1.

[13] Matthew J. Colbrook, Bogdan Roman, and Anders C. Hansen. "How to Compute Spectra with Error Control". In: *Physical Review Letters* 122.25 (June 2019), p. 250201. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.122.250201. (Visited on 02/02/2024).

[14] L Aceto, P Ghelardoni, and M Marletta. "Numerical Computation of Eigenvalues in Spectral Gaps of Sturm–Liouville Operators". In: *Journal of Computational and Applied Mathematics* (2006).

[15] Stephan Ramon Garcia, Javad Mashreghi, and William T. Ross. *Operator Theory by Example*. Oxford University Press, Jan. 2023. ISBN: 978-0-19-195456-6. DOI: 10.1093/oso/9780192863867.001.0001. (Visited on 11/16/2023).

[16] Albrecht Böttcher and Bernd Silbermann. *Analysis of Toeplitz Operators*. Springer Monographs in Mathematics. Berlin, Heidelberg: Springer, 1990. ISBN: 978-3-662-02654-0 978-3-662-02652-6. DOI: 10.1007/978-3-662-02652-6. (Visited on 11/15/2023).

[17] Palle Schmidt and Frank Spitzer. "The Toeplitz Matrices of an Arbitrary Laurent Polynomial". In: *Mathematica Scandinavica* 8.1 (1960), pp. 15–38. ISSN: 0025-5521. JSTOR: 24489115. (Visited on 11/15/2023).

[18] Ulf Grenander and Gabor Szegö. *Toeplitz Forms and Their Applications*. American Mathematical Society, 2001. ISBN: 978-0-8218-2844-1.

[19] Paul R. Halmos. *A Hilbert Space Problem Book*. Vol. 19. Graduate Texts in Mathematics. New York, NY: Springer, 1982. ISBN: 978-1-4684-9332-0 978-1-4684-9330-6. DOI: 10.1007/978-1-4684-9330-6. (Visited on 11/15/2023).

[20] Karl E. Gustafson and Duggirala K. M. Rao. *Numerical Range*. Ed. by S. Axler, F. W. Gehring, and P. R. Halmos. Universitext. New York, NY: Springer, 1997. ISBN: 978-0-387-94835-5 978-1-4613-8498-4. DOI: 10.1007/978-1-4613-8498-4. (Visited on 11/15/2023).

[21] Rupert L. Frank, Ari Laptev, and Timo Weidl. *Schrödinger Operators: Eigenvalues and Lieb–Thirring Inequalities*. Cambridge Studies in Advanced Mathematics. Cambridge: Cambridge University Press, 2022. ISBN: 978-1-00-921846-7. DOI: 10.1017/9781009218436. (Visited on 01/23/2024).

[22] P. A. Fillmore, J. G. Stampfli, and James P. Williams. "On the Essential Numerical Range, the Essential Spectrum, and a Problem of Halmos". In: *Acta Sci. Math.(Szeged)* 33.197 (1972), pp. 179–192.

[23] Sabine Bögli, Marco Marletta, and Christiane Tretter. "The Essential Numerical Range for Unbounded Linear Operators". In: *Journal of Functional Analysis* 279.1 (July 2020), p. 108509. ISSN: 00221236. DOI: 10.1016/j.jfa.2020.108509. (Visited on 11/15/2023).

[24] Andrzej Pokrzywa. "Method of Orthogonal Projections and Approximation of the Spectrum of a Bounded Operator". In: *Studia Mathematica* 65.1 (1979), pp. 21–29. ISSN: 0039-3223. (Visited on 11/15/2023).

Index