To infinity... and beyond!

A short course on infinite-dimensional spectral computations.

Matthew J. Colbrook

FSMP Fellow, Data Science Center, École Normale Supérieure

Junior Research Fellow, Trinity College, University of Cambridge

This course is based on [Col20a] and some further papers that can be found at
`http://www.damtp.cam.ac.uk/user/mjc249/Publications.html`.
Please email `m.colbrook@damtp.cam.ac.uk` with any comments, corrections, or questions.

"The infinite! No other question has ever moved so profoundly the spirit of man; no other idea has so fruitfully stimulated his intellect; yet no other concept stands in greater need of clarification."

— David Hilbert (1925)

Contents

Notation

We use the following notation and further notation will be introduced whenever appropriate.

$\mathcal{H}$	separable Hilbert space
$\mathcal{B}(\mathcal{H})$	set of bounded linear operators on $\mathcal{H}$
$B_r(x)$	closed ball (in a metric space) of radius r centred at x
$D_r(x)$	open ball (in a metric space) of radius r centred at x
$\mathrm{cl}(S)$	closure of a set S in a topological space
$d_{\mathrm{H}}(\mathcal{S},\mathcal{T})$	Hausdorff distance between compact sets $\mathcal{S}$ and $\mathcal{T}$
$\mathrm{Re}(z)$	real part of complex number z
$\mathrm{Im}(z)$	imaginary part of complex number z
$\overline{z}$	conjugate of complex number z
$\sigma_{\mathrm{inf}}(C)$	smallest singular value of rectangular matrix C, extended to operators in (3.2.1)
A^*	adjoint of operator A (when defined on a Hilbert space)
$\mathcal{D}(A)$	domain of operator A
$R(z,A)$	resolvent operator of operator A defined as $(A-zI)^{-1}$ for $z\notin\mathrm{Sp}(A)$
$\mathrm{Sp}(A)$	spectrum of operator A defined as $\{z\in\mathbb{C}: R(z,A)$ does not exist as a bounded operator$\}$
$\mathrm{Sp}_\epsilon(A)$	pseudospectrum of operator A defined as $\mathrm{cl}(\{z\in\mathbb{C}:\|(A-zI)^{-1}\|>1/\epsilon\})$ for $\epsilon>0$
$\mathrm{Sp}_d(A)$	discrete spectrum of operator A (evals. of finite multiplicity isolated from rest of $\mathrm{Sp}(A)$)
$\mathrm{Sp}_{\mathrm{ess}}(A)$	essential spectrum of operator A which we define as $\{z\in\mathbb{C}: A-zI$ is not Fredholm$\}$
$r_{\mathrm{ess}}(A)$	essential numerical radius of operator A defined as $\sup\{\lvert z\rvert : z\in\mathrm{Sp}_{\mathrm{ess}}(A)\}$
$W(A)$	numerical range of operator A defined as $\{\langle A\xi,\xi\rangle : \|\xi\|=1\}$
$W_e(A)$	essential numerical range of operator A defined as $\bigcap_{K\text{ compact}}\mathrm{cl}(W(A+K))$

If $A\in\mathcal{B}(\mathcal{H})$, then the pseudospectrum can equivalently be defined as

$$\mathrm{Sp}_\epsilon(A)=\{z\in\mathbb{C}:\|R(z,A)\|^{-1}\leq\epsilon\},$$

where we use the convention that $\|S^{-1}\|=\infty$ and $\|S^{-1}\|^{-1}=0$ if S^{-1} does not exist. We also remind the reader that the Hausdorff distance between $\mathcal{S}$ and $\mathcal{T}$ is

$$d_{\mathrm{H}}(\mathcal{S},\mathcal{T})=\max\left\{\sup_{\lambda\in\mathcal{S}}\mathrm{dist}(\lambda,\mathcal{T}),\sup_{\lambda\in\mathcal{T}}\mathrm{dist}(\lambda,\mathcal{S})\right\},$$

where $\mathrm{dist}(\lambda,\mathcal{T})=\inf_{\rho\in\mathcal{T}}\lvert\rho-\lambda\rvert$. Finally, when considering decision problems, we will use the discrete metric on $\{0,1\}$, with 1 interpreted as 'yes' and 0 interpreted as 'no'.

Chapter 1

Introduction

Given a suitable linear operator A on some Hilbert space $\mathcal{H}$, the spectrum of A is defined by

$$\mathrm{Sp}(A) := \{z \in \mathbb{C} : (A - zI)^{-1} \text{ does not exist as a bounded operator}\}$$

This set includes the familiar notion of eigenvalues, but in general is much richer! For example we might have continuous spectra. It is hard to overestimate the importance of computing spectra of infinite-dimensional operators in applied mathematics, quantum chemistry/mechanics, matter physics, statistical mechanics, optics and many other fields. Amongst its uses, the spectrum allows scientists to conduct stability, vibrational and asymptotic analysis, compute the energy levels of physical systems, diagonalise or decompose operators for analysis, perform data-driven analysis of systems, and compute solutions to PDEs. The problem of computing spectra is one of the most studied areas of computational mathematics over the last half-century, investigated by mathematicians and physicists alike since the 1950s. However, the many applications and theoretical studies of spectra depend on computations which are infamously difficult.

Computational approaches to obtain spectral information date back to leading mathematicians and physicists such as Anderson [And58], Goldstine [GMvN59], Kato [Kat49], Murray [GMvN59], Schrödinger [Sch40], Schwinger [Sch60b, Sch60a] and von Neumann [GMvN59]. For example, Schwinger introduced finite-dimensional approximations to quantum systems in infinite-dimensional spaces that allow for spectral computations, ideas which were already present in the work of Weyl [Wey50]. In [DVV94], Digernes, Varadarajan, and Varadhan proved convergence of spectra of Schwinger's finite-dimensional discretisation matrices for Schrödinger operators with continuous potentials bounded below and diverging at infinity (the resolvents of which are compact). We will solve this problem in a much more general setting in Chapter 3.

From an operator point of view, the computational spectral problem goes back as far as Szegő's work [Sze20] on finite section approximations. Since then, it has been studied intensely by both mathematicians [Aro51, Kat49, DLT85, Böt94, Böt96, LS96, BS99, BCN01, Zwo99, BBIN10, BIN11, Zwo13] and physicists [Sch40, And58, BC71, Hof76, Lie05, DS06b]. For instance, the seminal work of Fefferman and Seco [FS90, FS92, FS93, FS94b, FS94c, FS95, FS96b, FS96a, FS94a] on proving the Dirac–Schwinger conjecture is a striking example of computations used in order to obtain complete information about the asymptotic behaviour of the ground state of a family of Schrödinger operators.

The corresponding literature is vast (see [Col20a] for further discussion). However, whilst the above results undoubtedly represent triumphs for computational mathematics and theoretical physics, they only partially solve the problem and only hold for specific cases.

1.1 The goal

A reliable algorithm computing the spectrum should converge locally on compact subsets of $\mathbb{C}$. In other words, it should converge to the full spectrum and have no limiting points that are not in the spectrum. Moreover, we wish to have a guarantee that any point in the output is close to the spectrum, up to a chosen error tolerance. A key question is: do such algorithms exist? Despite more than 90 years of quantum theory, the answer to this question has been unknown, even for the case of general Schrödinger operators and even when also excluding the additional property of error control. Arveson, who helped develop the combination of spectral computations and C^*-algebra techniques[1] [Arv93a, Arv93b, Arv94a, Arv94b], summarises this open question for the problem of computing spectra of general self-adjoint operators,[2]

> *"Most operators that arise in practice are not presented in a representation in which they are diagonalized, and it is often very hard to locate even a single point in the spectrum... Thus, one often has to settle for numerical approximations [to the spectrum], and this raises the question of how to implement the methods of finite dimensional numerical linear algebra to compute the spectra of infinite dimensional operators. Unfortunately, there is a dearth of literature on this basic problem and, so far as we have been able to tell, there are no proven techniques."*
>
> — W. Arveson, UC Berkeley [Arv94b]

It is precisely the computational spectral problem, encapsulated in Arveson's question and dating back to the work of Schwinger in the 1960s [Sch60b, Sch60a], that this course addresses. The boundaries of what computers can achieve in computational spectral theory and mathematical physics remain largely unknown, leaving many open questions that have been unsolved for decades. Our goal is to solve some of these long-standing problems. Determining these computational boundaries means two things:

- Developing new algorithms that can handle problems previously out of reach,
- Providing mathematical proofs that the new algorithms are optimal.

In this course, we will do both for a range of infinite-dimensional spectral problems.

1.2 A motivating example

The spectrum of a general operator on a separable Hilbert space cannot be computed in finitely many operations. This holds even in the finite-dimensional case (which is mathematically equivalent to polynomial root-finding), and, in general, finite-dimensional spectral problems are solved numerically via iterative methods.[3] We must, therefore, give a precise meaning to a 'computational spectral problem'. For instance,

[1] This combination can be traced back to the work of Böttcher and Silbermann [BS83].

[2] There is, of course, a rich literature on using finite-dimensional algorithms to compute the spectrum of infinite-dimensional operators. Arveson is referring to the existence of a procedure that converges in general, using, for example, matrix elements of the operator with respect to an orthonormal basis.

[3] Computing the eigenvalues and eigenvectors of finite-dimensional matrices dates back to Wilkinson [Wil65] with guaranteed convergence for self-adjoint matrices via Wilkinson shifts, see [Par98].

suppose our operator is bounded and acts on $l^2(\mathbb{N})$. We can represent A by an infinite matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots \\ a_{21} & a_{22} & a_{23} & \cdots \\ a_{31} & a_{32} & a_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{1.2.1}$$

with respect to the canonical basis. Consider the case that an 'algorithm' can access matrix elements of A, which is natural for many Hamiltonian operators in physics. The algorithm uses a finite number of matrix elements, though it can adaptively choose which ones to use, and produces an output $\Gamma_n(A) \subset \mathbb{C}$. For example, if each a_{ij} is rational (or a rational approximation of a complex number), we could consider the output being produced by a Turing machine [Tur36]. If we allow real number arithmetic, then we could consider a Blum–Shub–Smale (BSS) [BCSS98] machine. At the very least, we should enforce consistency[4] in how the algorithm reads information and produces an output (see Definition 2.1.1 in Chapter 2). The algorithm is written with a subscript n because it is usual in numerical analysis to have a sequence of approximations (or even a sequence of different algorithms) that converge as $n \to \infty$. For example, in finite dimensions, n could correspond to the number of iterations of the famous QR algorithm, which converges under favourable conditions (see [CH19] for the infinite-dimensional version). The question is: do algorithms exist that converge in infinite dimensions? Surprisingly, the answer to this question is 'no' for many important problems, regardless of one's model of computation.

1.2.1 A 'three limit' algorithm

In [Han11] it was shown that, without any structural assumptions, it is possible to build an algorithm depending on *three parameters*, so that for general bounded operators acting on the canonical Hilbert space $l^2(\mathbb{N})$ the following holds with respect to the Hausdorff metric

$$\lim_{n_3\to\infty} \lim_{n_2\to\infty} \lim_{n_1\to\infty} \Gamma_{n_3,n_2,n_1}(A) = \mathrm{Sp}(A).$$

In other words, the process uses three successive limits. The algorithm roughly works as follows:

- For given $n_1, n_2 \in \mathbb{N}$, define the function

  $$\gamma_{n_2,n_1}(z) = \min\left\{\sigma_{\mathrm{inf}}(P_{n_1}(A - zI)P_{n_2}), \sigma_{\mathrm{inf}}(P_{n_1}(A^* - \overline{z}I)P_{n_2})\right\}.$$

  [DRAW PICTURE ON BOARD]

- One can prove that as $n_1 \to \infty$,

  $$\gamma_{n_2,n_1}(z) \uparrow \gamma_{n_2}(z) := \min\left\{\sigma_{\mathrm{inf}}((A - zI)P_{n_2}), \sigma_{\mathrm{inf}}((A^* - \overline{z}I)P_{n_2})\right\}.$$

  Similarly, as $n_2 \to \infty$,

  $$\gamma_{n_2}(z) \downarrow \gamma(z) := \min\left\{\sigma_{\mathrm{inf}}(A - zI), \sigma_{\mathrm{inf}}(A^* - \overline{z}I)\right\} = \left\|(A - zI)^{-1}\right\|^{-1} =: \left\|R(z, A)\right\|^{-1},$$

  with locally uniform convergence (uniform on compact subsets of $\mathbb{C}$).

  Exercise: Prove these statements and that $\mathrm{Sp}(A) = \{z \in \mathbb{C} : \gamma(z) = 0\}$.

[4] Our discussion can also be extended to the case of random algorithms, though we do not discuss this topic in this course.

- Define

  $$\Gamma_{n_3,n_2,n_1}(A) = \left\{z \in G_{n_2} : \gamma_{n_2,n_1}(z) \leq \frac{1}{n_3}\right\},$$

  where $G_n = B_n(0) \cap \frac{1}{n}(\mathbb{Z} + i\mathbb{Z})$. Then

  $$\lim_{n_1\to\infty} \Gamma_{n_3,n_2,n_1}(A) = \Gamma_{n_3,n_2}(A) := \left\{z \in G_{n_2} : \gamma_{n_2}(z) \leq \frac{1}{n_3}\right\}.$$

  And

  $$\lim_{n_2\to\infty} \Gamma_{n_3,n_2}(A) = \Gamma_{n_3}(A) := \left\{z \in \mathbb{C} : \gamma(z) \leq \frac{1}{n_3}\right\}.$$

  This set is called the (n_3^{-1}-)pseudospectrum of A. The final limit then shrinks this set to the spectrum.

  Exercise: Prove that $\lim_{n_3\to\infty} \Gamma_{n_3}(A) = \mathrm{Sp}(A)$.

  [DRAW PICTURE ON BOARD]

Question: Can we do away with the three limits?

Answer: No! Three successive limits turns out to be sharp if we consider the whole class of bounded operators. This means it is impossible to compute spectra of completely general operators using two limits (i.e., for all operators, without further information, even though standard algorithms can converge for different classes of operators) in any model of computation. This is most easily proven by embedding certain combinatorial problems of descriptive set theory within this problem - see Chapter 2.

This result gives rise to the solvability complexity index (SCI). Informally, the SCI is *the number of successive limits* needed to solve a computational problem, a measure of its difficulty. We will make this precise in Chapter 2. The SCI covers many areas in computational mathematics, extending beyond the spectral problem. It also has roots in the work of Smale [Sma81, Sma97], and his programme on the foundations of computational mathematics and scientific computing, though it is quite distinct. The notions of Turing computability [Tur36] and computability in the Blum–Shub–Smale (BSS) [BCSS98] sense become special cases, and impossibility results that are proven in the SCI hierarchy hold in all models of computation.

1.2.2 A 'one limit' algorithm with error control

The fact that *general* spectral problems require three limits poses a severe problem in applications: how can we guarantee that the outputs of numerical simulations converge and are sound? Fortunately, there is another class in the SCI hierarchy: Σ_1. This is the class of problems which require only one limit and for which there exists a convergent algorithm whose output is guaranteed to be included in the ϵ-neighbourhood of the spectrum, for an arbitrarily small ϵ. In other words, given an output, we know that it is sound, but we do not know if we have approximated all of the spectrum yet (though we must eventually converge to all of the spectrum).

Under very general assumptions,[5] there exists an algorithm $\Gamma_n(A)$ such that

$$\lim_{n\to\infty} d_{\mathrm{H}}(\Gamma_n(A), \mathrm{Sp}(A)) = 0,$$

with d_{H} the usual Hausdorff metric on non-empty compact subsets of $\mathbb{C}$. We also obtain error control, in the sense that the algorithm computes an error bound $E_n(A; z)$ such that

$$\mathrm{dist}(z, \mathrm{Sp}(A)) \leq E_n(A; z) \quad \forall z \in \Gamma_n(A) \quad \text{and} \quad \lim_{n\to\infty} \sup_{z\in\Gamma_n(A)} E_n(A; z) = 0. \tag{1.2.2}$$

[5] The assumptions hold in the majority of applications. See §3.1.1 and §3.1.2 for the precise details.

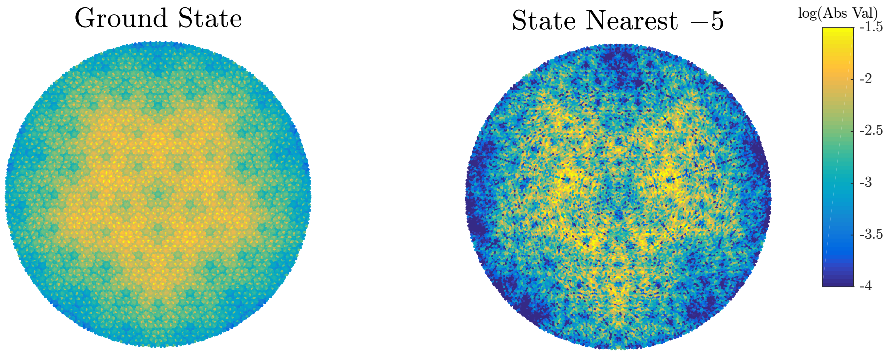


Figure 1.1: The ground ‘state’ for the Penrose Laplacian from [CRH19] and an approximate state corresponding to energy nearest -5. The algorithm allows us to choose which states to compute without direct diagonalisation. It should be emphasised that we are not necessarily approximating eigenvectors since the spectrum may not consist solely of eigenvalues.

This notion of error control, denoted by Σ_1, is discussed in detail in §2.2, along with its dual notion Π_1. The constructed algorithm is parallelisable and can also be extended to compute quantities such as approximate states (see Figure 1.1). The results hold when considering infinite matrix representations of operators, and also for partial differential operators when sampling the coefficients.

However, stricter error control, in the sense of computing E_n with

$$d_{\mathrm{H}}(\Gamma_n(A), \mathrm{Sp}(A)) \leq E_n(A) \tag{1.2.3}$$

is in general impossible (we denote this stricter sense of error control by Δ_1) in any model of computation. As a very simple example, consider the class of all bounded diagonal operators $A \in \mathcal{B}(l^2(\mathbb{N}))$ of the form

$$A = \begin{pmatrix} a_1 & & & \\ & a_2 & & \\ & & a_3 & \\ & & & \ddots \end{pmatrix}, \qquad a_j \in \mathbb{C}. \tag{1.2.4}$$

Since an algorithm can only deal with a finite amount of information at any one time (i.e., finitely many of the a_i), it is clear that the problem of computing the spectrum $\mathrm{Sp}(A)$ cannot be done with error control in the sense of (1.2.3). However, one can simply choose an algorithm Γ_n to collect $\{a_j\}_{j=1}^n$ and then one trivially has that $\Gamma_n(A) \rightarrow \mathrm{Sp}(A)$ as $n \rightarrow \infty$. We also clearly have the extra feature that

$$\Gamma_n(A) \subset \mathrm{Sp}(A), \quad n \in \mathbb{N}.$$

In particular, we have convergence from below, and this is much stronger than just convergence, since $\Gamma_n(A)$ always produces a correct output. Such a type of convergence is incredibly important, since it gives a guarantee of reliability. We extend this type of convergence (up to an arbitrarily small user-chosen error tolerance given by the E_n in (1.2.2)) to a vast number of spectral problems. In some sense, given the above simple example, we show that the computational spectral problem is not harder than computing the spectrum of a diagonal operator.

An example from physics

Suppose that A is sparse, meaning that it has only finitely many non-zero entries in each column, and suppose also that $A^* = A$ (self-adjoint). As an example, we consider Schrödinger operators on quasicrystals. Quasicrystals are non-repeating (aperiodic) structures with a long-range, self-similar nature. More generally, systems with long-range order and short-range disorder are abundant in nature. Currently, aperiodic systems are not nearly as well understood as their periodic cousins. We might ask, then: what are the physics of aperiodic systems? Understanding spectral properties is key to answering these types of questions. However, the aperiodic nature of quasicrystals, which makes them so interesting to study in the first place, also makes it a considerable challenge to approximate spectra associated with these systems!

We consider a Penrose tile, a canonical model of a quasicrystal in 2D, and generated the lattice shown in Figure 1.2a by considering a lattice 'site' to exist at each vertex (the black dots) and tunnelling bonds along the edges of the tiles. The model taken is that of a charged single-particle, which can exist on the set of sites and can tunnel between the sites along the bonds. We then apply a perpendicular magnetic field, which modifies the tunnelling strengths to enforce the usual circular motion of a free charged particle in a magnetic field. The operator in this scenario is a Hamiltonian A which, in matrix form, is given by

$$(A\psi)_j = -\sum_{\langle j,k\rangle} e^{i\alpha_{kj}}\psi_k,$$

with summation over sites connected by an edge. Here α_{kj} is a phase factor that is given in terms of the strength of the magnetic field and ψ denotes the wave function.

The most common approach to computing spectra is to truncate the operator. Physically, in our example, this corresponds to truncating the tile and studying the interactions of a finite number of sites within the truncation (Figure 1.2b). Mathematically, this corresponds to studying a finite section of the operator and computing spectra of the corresponding finite-dimensional system (eigenvalues of finite square matrices shown as a red box in Figure 1.2). In this model, the dimension of this finite-dimensional system is precisely the number of sites included in the truncation. Figure 1.3a shows the output of this approach, where the approximation of the spectrum is plotted for different magnetic field strengths. We have labelled portions of this picture as 'spectral pollution'. This approach does not approximate the correct solution and does not provide any form of error bounds.

Instead, we can compute spectra as follows, by reducing the number of limits in the above algorithm:

- Since A is sparse, we have access to $f : \mathbb{N} \to \mathbb{N}$ such that $(I - P_{f(n)})AP_n = 0$.

  Exercise: Prove that $\gamma_{n_2}(z) = \gamma_{n_2,f(n_2)}(z)$.
- Since A is self-adjoint, we can avoid the final shrinking step.

  Exercise: Prove that $\gamma(z) = \mathrm{dist}(z, \mathrm{Sp}(A))$.
- We will see later how to compute $\gamma_{n_2}(z)$ and use a local optimisation routine to compute $\mathrm{Sp}(A)$!

Physically, the *rectangular* truncation $P_{f(n)}AP_n$ corresponds to including the interactions of the finite truncation with the rest of the tile (Figure 1.2c). We can think of this as a tool for studying the full infinite-dimensional operator directly, even on a finite computer. Leveraging this idea, we can now approximate spectra in such a way that (i) our approximations approach the correct solution as our truncation size increases, and (ii) such that we can explicitly bound the error of any computed approximation. The practitioner can now provide a desired error bound, which our algorithm will then adaptively realise. Figure 1.3b

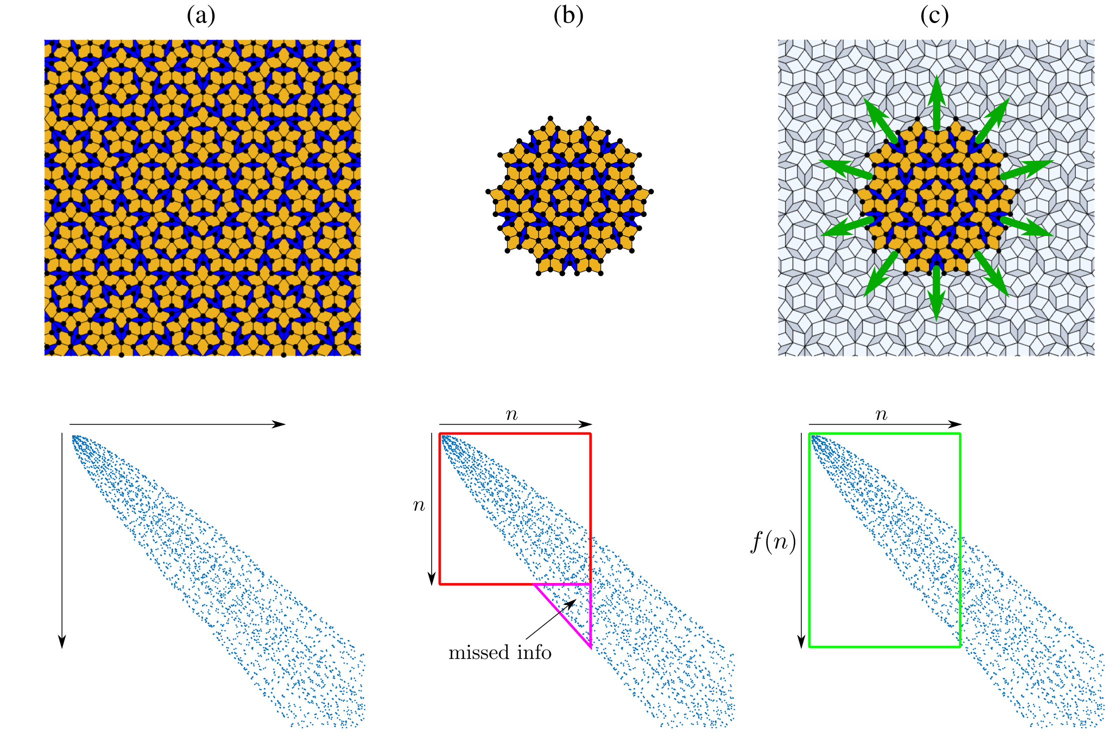


Figure 1.2: Top: (a) Infinite aperiodic Penrose tile. (b) Finite truncation of tile to n sites. (c) Finite truncation with interactions shown as green arrows (proposed method). Bottom: The corresponding sparsity patterns (non-zero entries of the infinite matrix of the operator A). The boxes show the different types of truncations of the operator. In (c), $f(n)$ is chosen to include all of the interactions of the first n sites.

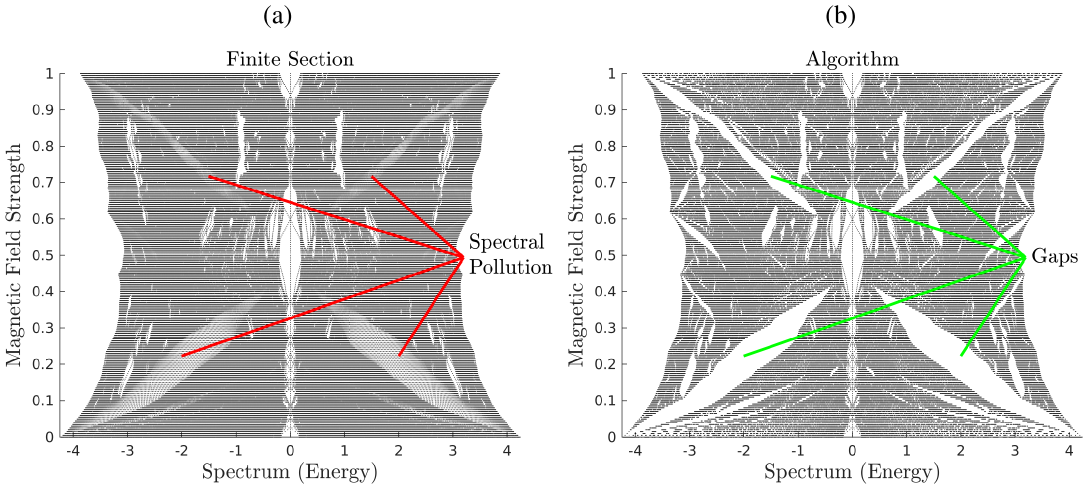


Figure 1.3: Computation of spectra using (a) finite section and (b) the proposed method.

shows the output of this approach for our example. We now (i) have the correct gaps in the spectrum, (ii) approximate the correct spectrum, and, for this example, (iii) have a guaranteed error bound of 0.01. With this technique in hand, we can reliably probe the bulk physical properties of such aperiodic systems. Indeed, this technique is already allowing for the discovery and investigation of new physics in quasicrystalline systems, including their transport and topological properties [JCN$^+$21].

Chapter 2

The Solvability Complexity Index

This chapter discusses the Solvability Complexity Index (SCI) hierarchy. We use this to show that the algorithms in this course realise the boundary of what computers can achieve. All of the results concerning the hierarchy itself are placed in one chapter. For further discussion on the hierarchy, the reader is advised to consult [Col20a, BACH^{+}20]. For extensions to randomised algorithms, see [CAH22a].

Disclaimer: This is not a course on logic or descriptive set theory. This chapter is quite dense but is largely self-contained. However, once completed, we will have the tools to tackle infinite-dimensional spectral computations.

2.1 The Basic SCI Hierarchy

First, we define a computational problem. The four basic objects of a computational problem are:

- Ω: some set, called the primary set,
- Λ: a set of complex-valued functions on Ω, called the evaluation set,
- $\mathcal{M}$: a metric space,
- $\Xi : \Omega \to \mathcal{M}$: the problem function.

Ω is the class of objects that give rise to our computational problem. The problem function $\Xi : \Omega \to \mathcal{M}$ is the map we wish to compute. The set Λ is the collection of functions that provide us with the information we are allowed access to. The collection $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ is referred to as a *computational problem*.

For example, we could have $\Omega = \mathcal{B}(l^2(\mathbb{N}))$ and Ξ the problem function that takes $A \in \Omega$ and maps it to its spectrum $\mathrm{Sp}(A)$. Since the spectrum is a non-empty compact subset of $\mathbb{C}$ (in this case), we can let $\mathcal{M}$ be the set of non-empty compact subsets of $\mathbb{C}$ equipped with the Hausdorff metric. In this case, Λ could correspond to the evaluation of matrix entries of a given $A \in \Omega$.

Occasionally we will consider a function Ξ such that for $A \in \Omega$ we have that $\Xi(A) \subset \mathcal{M}$. In this case, we still require that algorithms produce a single-valued output. However, we replace the metric in order to define convergence. In particular, $\Gamma_n(A) \to \Xi(A)$ as $n \to \infty$ means $\inf_{y \in \Xi(A)} d_{\mathcal{M}}(\Gamma_n(A), y) \to 0$.

Definition 2.1.1 (General Algorithm)**.** *Given a computational problem $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$, a general algorithm is a mapping $\Gamma : \Omega \to \mathcal{M}$ such that for each $A \in \Omega$*

(i) there exists a (non-empty) finite subset of evaluations $\Lambda_\Gamma(A) \subset \Lambda$,

(ii) the action of Γ on A only depends on $\{A_f\}_{f\in\Lambda_\Gamma(A)}$ where $A_f := f(A)$,

(iii) for every $B \in \Omega$ such that $B_f = A_f$ for every $f \in \Lambda_\Gamma(A)$, it holds that $\Lambda_\Gamma(B) = \Lambda_\Gamma(A)$.

The three properties of a general algorithm are the most basic natural properties we would expect any deterministic computational device to obey. The first condition says that the algorithm can only take a finite amount of information, though it is allowed adaptively to choose, depending on the input, the finite amount of information it reads. The second condition ensures that the algorithm's output only depends on its input, or rather the information that it has accessed. The final condition is very important and ensures that the algorithm produces outputs and accesses information in a consistent manner. In other words, if it sees the same information for two different inputs, then it cannot behave differently for those inputs.

Note that the definition of a general algorithm allows a stronger form of computation than the definition of a Turing machine [Tur36] (digital computer) or a Blum–Shub–Smale (BSS) machine [BCSS98] (analog computer). A general algorithm has no restrictions on the operations allowed. Whilst complete generality seem to be at odds with practical computation, we use this model for two primary reasons:

(i) *Strongest lower bounds (and complementary strongest upper bounds):* Since Definition 2.1.1 is completely general, the lower bounds hold in any model of computation, such as a Turing machine or a Blum–Shub–Smale machine. This is not an issue for practical computation since the algorithms in this course can be made to work using only arithmetic operations over the rationals. Hence, we obtain the strongest possible lower bounds and the strongest possible upper bounds.

(ii) *Focus on information:* Using the concept of a general algorithm considerably simplifies the proofs of lower bounds. The proven lower bounds are due to the problem at hand being inherently noncomputable. It is not a question of the type of operations allowed being too restrictive, but rather that the information about each input available to the algorithm is insufficient to solve the problem.

With a definition of a general algorithm, we can define the concept of towers of algorithms.

Definition 2.1.2 (Tower of algorithms)**.** *Given a computational problem $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$, a tower of algorithms of height k for $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ is a collection of sequences of functions*

$$\Gamma_{n_k} : \Omega \to \mathcal{M}, \quad \Gamma_{n_k,n_{k-1}} : \Omega \to \mathcal{M}, \ \dots\ , \Gamma_{n_k,\dots,n_1} : \Omega \to \mathcal{M},$$

where $n_k, \dots, n_1 \in \mathbb{N}$ and the functions $\Gamma_{n_k,\dots,n_1}$ at the lowest level in the tower are general algorithms in the sense of Definition 2.1.1. Moreover, for every $A \in \Omega$,

$$\begin{aligned}
\Xi(A) &= \lim_{n_k\to\infty} \Gamma_{n_k}(A),\\
\Gamma_{n_k}(A) &= \lim_{n_{k-1}\to\infty} \Gamma_{n_k,n_{k-1}}(A),\\
&\ \ \vdots\\
\Gamma_{n_k,\dots,n_2}(A) &= \lim_{n_1\to\infty} \Gamma_{n_k,\dots,n_1}(A),
\end{aligned}$$

with convergence in the metric space $\mathcal{M}$.

Throughout this course, a general tower will refer to the very general definition in Definition 2.1.2 specifying that there are no further restrictions. This will be denoted by $\alpha = G$. When we specify the type of tower, we specify requirements on the functions $\Gamma_{n_k,\dots,n_1}$ in the hierarchy, in particular, what kind of operations may be allowed. A tower of algorithms for a computational problem is the toolbox allowed.

Definition 2.1.3 (Arithmetic tower)**.** *Given a computational problem $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$, an arithmetic tower of algorithms of height k for $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ is a tower of algorithms where the lowest functions $\Gamma = \Gamma_{n_k,\dots,n_1} : \Omega \to \mathcal{M}$ satisfy the following: For each $A \in \Omega$ the action of Γ on A consists of only performing finitely many arithmetic operations and comparisons on $\{A_f\}_{f\in\Lambda_\Gamma(A)}$, where we remind the reader that $A_f = f(A)$. For arithmetic towers we let $\alpha = A$.*

Definition 2.1.4 (Solvability Complexity Index)**.** *A computational problem $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ is said to have Solvability Complexity Index $\mathrm{SCI}(\Xi, \Omega, \mathcal{M}, \Lambda)_\alpha = k$, with respect to a tower of algorithms of type α, if k is the smallest integer for which there exists a tower of algorithms of type α of height k. If no such tower exists then $\mathrm{SCI}(\Xi, \Omega, \mathcal{M}, \Lambda)_\alpha = \infty$. If there exists a tower $\{\Gamma_n\}_{n\in\mathbb{N}}$ of type α and height one such that $\Xi = \Gamma_{n_1}$ for some $n_1 < \infty$, then we define $\mathrm{SCI}(\Xi, \Omega, \mathcal{M}, \Lambda)_\alpha = 0$.*

With the definition of the SCI, we can define the SCI hierarchy. Without any extra structure on the metric space $\mathcal{M}$, the Δ_k^α classes are the finest refinement we can obtain in terms of the SCI. However, as described below, when more structure is allowed, the hierarchy becomes much richer.

Definition 2.1.5 (The Solvability Complexity Index hierarchy)**.** *Consider a collection $\mathcal{C}$ of computational problems and let $\mathcal{T}$ be the collection of all towers of algorithms of type α for the computational problems in $\mathcal{C}$. Define*

$$\Delta_0^\alpha := \{\{\Xi, \Omega\} \in \mathcal{C} \mid \mathrm{SCI}(\Xi, \Omega)_\alpha = 0\}$$
$$\Delta_{m+1}^\alpha := \{\{\Xi, \Omega\} \in \mathcal{C} \mid \mathrm{SCI}(\Xi, \Omega)_\alpha \leq m\}, \qquad m \in \mathbb{N},$$

as well as

$$\Delta_1^\alpha := \{\{\Xi, \Omega\} \in \mathcal{C} \mid \exists\, \{\Gamma_n\}_{n\in\mathbb{N}} \in \mathcal{T} \text{ s.t. } \forall A \in \Omega\ d(\Gamma_n(A), \Xi(A)) \leq 2^{-n}\}.$$

2.2 Error Control Extensions of the SCI Hierarchy

When there is extra structure on the metric space $\mathcal{M}$, say $\mathcal{M} = \mathbb{R}$ or $\mathcal{M} = \{0, 1\}$ with the standard metrics (or more generally, a totally ordered set), one may be able to define convergence of functions from above or below. This is an extra form of structure that allows for a type of error control. Such error control is important, for example, in computer-assisted proofs, and of course, crucial in scientific computing.

Definition 2.2.1 (The SCI Hierarchy for a Totally Ordered Set)**.** *Given the set-up in Definition 2.1.5 and suppose in addition that $\mathcal{M}$ is a totally ordered set. Define*

$$\Sigma_0^\alpha = \Pi_0^\alpha = \Delta_0^\alpha,$$
$$\Sigma_1^\alpha = \{\{\Xi, \Omega\} \in \Delta_2 \mid \exists\, \{\Gamma_n\} \in \mathcal{T} \text{ s.t. } \Gamma_n(A) \nearrow \Xi(A)\ \forall A \in \Omega\},$$
$$\Pi_1^\alpha = \{\{\Xi, \Omega\} \in \Delta_2 \mid \exists\, \{\Gamma_n\} \in \mathcal{T} \text{ s.t. } \Gamma_n(A) \searrow \Xi(A)\ \forall A \in \Omega\},$$

where $\nearrow$ and $\searrow$ denotes convergence from below and above respectively, as well as, for $m \in \mathbb{N}$,

$$\Sigma_{m+1}^\alpha = \{\{\Xi, \Omega\} \in \Delta_{m+2} \mid \exists\, \{\Gamma_{n_{m+1},\dots,n_1}\} \in \mathcal{T} \text{ s.t. } \Gamma_{n_{m+1}}(A) \nearrow \Xi(A)\ \forall A \in \Omega\},$$
$$\Pi_{m+1}^\alpha = \{\{\Xi, \Omega\} \in \Delta_{m+2} \mid \exists\, \{\Gamma_{n_{m+1},\dots,n_1}\} \in \mathcal{T} \text{ s.t. } \Gamma_{n_{m+1}}(A) \searrow \Xi(A)\ \forall A \in \Omega\}.$$

If the metric space $\mathcal{M} = \{0,1\}$, it is clearly a totally ordered set and hence, from Definition 2.2.1, we obtain the SCI hierarchy for arbitrary decision problems. We want to generalise the above notions of error control to scenarios suitable for spectral computations. In the case where $\mathcal{M}$ is the collection of non-empty compact subsets of another metric space $\mathcal{M}'$, it is custom to equip $\mathcal{M}$ with the Hausdorff metric

$$d_{\mathrm{H}}(X,Y) = \max\left\{\sup_{x\in X}\inf_{y\in Y} d(x,y), \sup_{y\in Y}\inf_{x\in X} d(x,y)\right\}.$$

In the case where $\mathcal{M}$ is the collection of non-empty closed subsets of $\mathcal{M}'$, we use the Attouch–Wets metric

$$d_{\mathrm{AW}}(C_1,C_2) = \sum_{n=1}^{\infty} 2^{-n} \min\left\{1, \sup\nolimits_{d_{\mathcal{M}'}(x_0,x)\leq n} \left|\mathrm{dist}(x,C_1) - \mathrm{dist}(x,C_2)\right|\right\},$$

where C_1 and C_2 are non-empty closed subsets of $\mathbb{C}$, $x_0 \in \mathcal{M}'$ is some fixed element of $\mathcal{M}'$ and where $d(x,C)$ is the usual distance between the point x and a set C. Note that $d_{\mathrm{AW}}(C_1,C_2) \in [0,1]$. In the case that $\mathcal{M}' = \mathbb{C}$ with the usual metric, we take $x_0 = 0$ without loss of generality. One should view the Attouch–Wets metric as a generalisation of the familiar Hausdorff metric on compact subsets. In other words, we seek local uniform convergence. In fact, both metrics can be viewed in terms of metrics on spaces of continuous functions [Bee93].

The following provides the generalisation and we remark on the intuition behind this definition below.

Definition 2.2.2 (The SCI Hierarchy (Attouch–Wets/Hausdorff metric))**.** *Given the set-up in Definition 2.1.5 and suppose in addition that $(\mathcal{M}, d)$ is the Attouch–Wets or the Hausdorff metric induced by another metric space $\mathcal{M}'$. Define for $m \in \mathbb{N}$*

$$\begin{aligned}
\Sigma_0^{\alpha} &= \Pi_0^{\alpha} = \Delta_0^{\alpha},\\
\Sigma_1^{\alpha} &= \{\{\Xi,\Omega\} \in \Delta_2 \mid \exists\, \{\Gamma_n\} \in \mathcal{T},\ \{X_n(A)\} \subset \mathcal{M} \textit{ s.t. } \Gamma_n(A) \underset{\mathcal{M}'}{\subset} X_n(A),\\
&\qquad \lim_{n\to\infty} \Gamma_n(A) = \Xi(A),\ \ d(X_n(A), \Xi(A)) \leq 2^{-n}\ \ \forall A \in \Omega\},\\
\Pi_1^{\alpha} &= \{\{\Xi,\Omega\} \in \Delta_2 \mid \exists\, \{\Gamma_n\} \in \mathcal{T},\ \{X_n(A)\} \subset \mathcal{M} \textit{ s.t. } \Xi(A) \underset{\mathcal{M}'}{\subset} X_n(A),\\
&\qquad \lim_{n\to\infty} \Gamma_n(A) = \Xi(A),\ \ d(X_n(A), \Gamma_n(A)) \leq 2^{-n}\ \ \forall A \in \Omega\},
\end{aligned}$$

where $\subset_{\mathcal{M}'}$ means inclusion in the metric space $\mathcal{M}'$. Moreover,

$$\begin{aligned}
\Sigma_{m+1}^{\alpha} &= \{\{\Xi,\Omega\} \in \Delta_{m+2} \mid \exists\, \{\Gamma_{n_{m+1},\ldots,n_1}\} \in \mathcal{T},\ \{X_{n_{m+1}}(A)\} \subset \mathcal{M} \textit{ s.t. } \Gamma_{n_{m+1}}(A) \underset{\mathcal{M}'}{\subset} X_{n_{m+1}}(A),\\
&\qquad \lim_{n_{m+1}\to\infty} \Gamma_{n_{m+1}}(A) = \Xi(A),\ \ d(X_{n_{m+1}}(A), \Xi(A)) \leq 2^{-n_{m+1}}\ \ \forall A \in \Omega\},\\
\Pi_{m+1}^{\alpha} &= \{\{\Xi,\Omega\} \in \Delta_{m+2} \mid \exists\, \{\Gamma_{n_{m+1},\ldots,n_1}\} \in \mathcal{T},\ \{X_{n_{m+1}}(A)\} \subset \mathcal{M} \textit{ s.t. } \Xi(A) \underset{\mathcal{M}'}{\subset} X_{n_{m+1}}(A),\\
&\qquad \lim_{n_{m+1}\to\infty} \Gamma_{n_{m+1}}(A) = \Xi(A),\ \ d(X_{n_{m+1}}(A), \Gamma_{n_{m+1}}(A)) \leq 2^{-n_{m+1}}\ \ \forall A \in \Omega\}.
\end{aligned}$$

Intuitively, this captures convergence from below or above respectively, up to a small error parameter 2^{-n}. Note that to build a Σ_1 algorithm in the Hausdorff case, it is enough (by taking subsequences of n) to construct $\Gamma_n(A)$ such that $\Gamma_n(A) \subset \Xi(A) + B_{E_n(A)}(0)$ with some computable $E_n(A)$ that converges to zero. A visual demonstration of these classes for the Hausdorff metric is shown in Figure 2.1. The SCI hierarchy gives rise to the following structure:

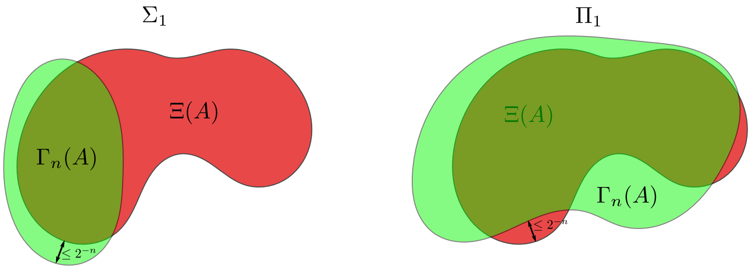


Figure 2.1: Meaning of Σ_1 and Π_1 convergence for problem function Ξ computed in the Hausdorff metric. The red area represents $\Xi(A)$, whereas the green areas represent the output of the algorithm $\Gamma_n(A)$. Σ_1 convergence means convergence as $n \to \infty$ but each output point in $\Gamma_n(A)$ is at most distance 2^{-n} from $\Xi(A)$. Similarly, in the case of Π_1, we have convergence as $n \to \infty$ but any point in $\Xi(A)$ is at most distance 2^{-n} from $\Gamma_n(A)$. The same notion holds for Σ_1 and Π_1 in the Attouch–Wets topology, but now when restricting to arbitrary compact balls (see Lemma 3.2.2).

$$\begin{tikzcd}
\Pi_0^\alpha \arrow[d, phantom, "\parallel"] & & \Pi_1^\alpha \arrow[dr, phantom, "\subsetneq" sloped] & & \Pi_2^\alpha \arrow[dr, phantom, "\subsetneq" sloped] & & {} \\
\Delta_0^\alpha \arrow[r, phantom, "\subsetneq"] \arrow[d, phantom, "\parallel"] & \Delta_1^\alpha \arrow[r, phantom, "\subsetneq"] \arrow[ur, phantom, "\subsetneq" sloped] \arrow[dr, phantom, "\subsetneq" sloped] & \Sigma_1^\alpha \cup \Pi_1^\alpha \arrow[r, phantom, "\subsetneq"] & \Delta_2^\alpha \arrow[r, phantom, "\subsetneq"] \arrow[ur, phantom, "\subsetneq" sloped] \arrow[dr, phantom, "\subsetneq" sloped] & \Sigma_2^\alpha \cup \Pi_2^\alpha \arrow[r, phantom, "\subsetneq"] & \Delta_3^\alpha \arrow[r, phantom, "\subsetneq"] \arrow[ur, phantom, "\subsetneq" sloped] \arrow[dr, phantom, "\subsetneq" sloped] & \cdots \\
\Sigma_0^\alpha & & \Sigma_1^\alpha \arrow[ur, phantom, "\subsetneq" sloped] & & \Sigma_2^\alpha \arrow[ur, phantom, "\subsetneq" sloped] & & {}
\end{tikzcd}$$

Note, it is precisely the classes Σ_1^α and Π_1^α that are crucial in computer-assisted proofs.

To say a bit more about the structure, we need the following definition (which holds for standard spaces such as $\{0, 1\}$ or $\mathbb{R}$ with the usual metric).

Definition 2.2.3. *Given a totally ordered metric space $(\mathcal{M}, d)$, we say that the metric is order respecting if for any $a, b, c \in \mathcal{M}$ with $a \leq b \leq c$ we have $d(a, b) \leq d(a, c)$.*

The following proposition gives some insight into the extended SCI hierarchy as defined above, and shows that the results of later chapters are sharp.

Proposition 2.2.4 (Properties of the SCI hierarchy II)**.** *Given the above set-up, let $(\mathcal{M}, d)$ be either the Hausdorff or Attouch–Wets metric or a totally ordered metric space with order respecting metric. Let $k = 1, 2$ or 3, then we have the following.*

(i) $\Delta_k^G = \Sigma_k^G \cap \Pi_k^G$. In particular, if for a problem $\Xi : \Omega \to \mathcal{M}$ we have $\Delta_k^G \not\ni \{\Xi, \Omega\} \in X_k^\alpha$, where $X = \Sigma$ or Π and α denotes any type of tower, then $\{\Xi, \Omega\} \notin Y_k^\alpha$, where $Y = \Pi$ or Σ respectively.

(ii) Suppose for a computational problem $\Xi : \Omega \to \mathcal{M}$ we have a corresponding convergent Σ_k^A tower $\Gamma^1_{n_k,\ldots,n_1}$ and a corresponding convergent Π_k^A tower $\Gamma^2_{n_k,\ldots,n_1}$. Suppose also that we can compute for every $A \in \Omega$ the distance $d(\Gamma^1_{n_k,\ldots,n_1}(A), \Gamma^2_{n_k,\ldots,n_1}(A))$ to arbitrary precision using finitely many arithmetic operations and comparisons. Then $\{\Xi, \Omega\} \in \Delta_k^A$.

Exercise (hard): Prove Proposition 2.2.4.

Throughout this course, we will prove results of the form $\Delta_k^G \not\ni \{\Xi, \Omega\} \in X_k^\alpha$. Part (i) says that this is an optimal classification in the SCI hierarchy if $k \leq 3$. It is an open problem whether part (i) of the proposition extends to larger k (the proof for $k = 3$ is already very technical).

2.3 A Link with Descriptive Set Theory

Next, we shall link the SCI hierarchy in a particular specific case to the Baire hierarchy (on a suitable topological space). As well as being interesting in its own right, this link provides canonical problems high up in the SCI hierarchy. In particular, the results proven here hold for towers of general algorithms, without restrictions such as arithmetic operations or notions of recursivity. This fact will be used extensively in the proofs of lower bounds for spectral problems that have SCI > 2, where we typically reduce the problems discussed in this section to the given spectral problem.

It is beyond the scope of this course to provide an extensive discussion of descriptive set theory, but we refer the reader to [KL87, Mos09] for excellent introductions that cover the main ideas.[1] It should be stressed that such a link to existing hierarchies only exists in special cases (when Ω and $\mathcal{M}$ are particularly well-behaved). Even when such a link exists, the induced topology on Ω is often too complicated, unnatural or strong to be useful from a computational viewpoint. We also take the view that for problems of scientific interest, the mappings Λ and metric space $\mathcal{M}$ are often given to us apriori from the corresponding applications and may not be compatible with topological viewpoints of computation.

2.3.1 Some results from descriptive set theory

We briefly state the definition of the Borel hierarchy as well as some well-known theorems from descriptive set theory. Let X be a metric space and define

$$\Sigma_1^0(X) = \{U \subset X : U \text{ is open}\}, \quad \Pi_1^0(X) = \sim\Sigma_1^0(X) = \{F \subset X : F \text{ is closed}\},$$

where for a class $\mathcal{U}$, $\sim\mathcal{U}$ denotes the class of complements (in X) of elements of $\mathcal{U}$. Inductively define

$$\begin{aligned}&\Sigma_\xi^0(X) = \{\cup_{n\in\mathbb{N}} A_n : A_n \in \Pi_{\xi_n}^0, \xi_n < \xi\}, \text{ if } \xi > 1,\\ &\Pi_\xi^0(X) = \sim\Sigma_\xi^0(X), \quad \Delta_\xi^0(X) = \Sigma_\xi^0(X) \cap \Pi_\xi^0(X).\end{aligned}$$

The full Borel hierarchy extends to all $\xi < \omega_1$ (ω_1 being the first uncountable ordinal) by transfinite induction but we do not need this here.

Definition 2.3.1 ([KL87]). *Given a class of subsets, $\mathcal{U}$, of a metric space X and given another metric space Y, we say that the function $f : X \to Y$ is $\mathcal{U}$-measurable if $f^{-1}(U) \in \mathcal{U}$ for every open set $U \subset Y$.*

Given metric spaces X and Y, the Baire hierarchy is defined as follows. A function $f : X \to Y$ is of Baire class 1, written $f \in \mathcal{B}_1$, if it is $\Sigma_2^0(X)$-measurable. For $1 < \xi < \omega_1$, a function $f : X \to Y$ is of Baire class ξ, written $f \in \mathcal{B}_\xi$, if it is the pointwise limit of a sequence of functions f_n in $\mathcal{B}_{\xi_n}$ with $\xi_n < \xi$. The following theorem is well-known (see for example [KL87] section 24) and provides a useful link between the Borel and Baire hierarchies.

Theorem 2.3.2 (Lebesgue, Hausdorff, Banach). *Let X, Y be metric spaces with Y separable and $1 \leq \xi < \omega_1$. Then $f \in \mathcal{B}_\xi$ if and only if it is $\Sigma_{\xi+1}^0(X)$ measurable. Furthermore, if X is zero-dimensional (Hausdorff with a basis of clopen (closed and open) sets) and $f \in \mathcal{B}_1$, then f is the pointwise limit of a sequence of continuous functions.*

[1] The reader wishing to assimilate the bare minimum quickly will find Chapter 2 of [KL87] sufficient for this section.

The assumption that X is zero-dimensional in the last statement is important. Without any assumptions, the final statement of the theorem is false, as is easily seen by considering $X = \mathbb{R}$. Examples of zero-dimensional spaces include products of the discrete space $\{0, 1\}$ or the Cantor space. Any such space is necessarily totally disconnected, meaning that the connected components in the space are the one-point sets (the converse is true for locally compact Hausdorff spaces). Our primary interest will be the cases when Y is equal to $\{0, 1\}$ or $[0, 1]$, both with their natural topologies.

2.3.2 Linking the SCI hierarchy to the Baire hierarchy in a special case

The following definition will be used as a sufficient criterion for a topology to exist on Ω such that Δ_1 problems are precisely the continuous functions from Ω to $\mathcal{M}$.

Definition 2.3.3. *Given the triple $\{\Omega, \mathcal{M}, \Lambda\}$, a class of algorithms $\mathcal{A}$ is closed under search with respect to $\{\Omega, \mathcal{M}, \Lambda\}$ if whenever*

1. *$\mathcal{I}$ is an index set,*
2. *$\{n_i\}_{i\in\mathcal{I}}$ a family of natural numbers,*
3. *$\{\Gamma_{i,l} : \Omega \to \mathcal{M}\}_{i\in\mathcal{I},l\leq n_i} \subset \mathcal{A}$,*
4. *$\{U_{i,l}\}_{i\in\mathcal{I},l\leq n_i}$ family of basic open sets in $\mathcal{M}$ with $\cup_{i\in\mathcal{I}} \cap_{l\leq n_i} \Gamma_{i,l}^{-1}(U_{i,l}) = \Omega$, where $\Gamma_{i,l}^{-1}(U_{i,l}) = \{x \in \Omega : \Gamma_{i,l}(x) \in U_{i,l}\}$,*
5. *$\{c_i\}_{i\in\mathcal{I}}$ a family of points in some arbitrary dense subset of $\mathcal{M}$,*

then there is some $\Gamma \in \mathcal{A}$ such that for every $x \in \Omega$ there exists some $i \in \mathcal{I}$ with $\Gamma(x) = c_i$ and for all $l \leq n_i$ we have $\Gamma_{i,l}(x) \in U_{i,l}$.

Proposition 2.3.4. *Suppose that $\mathcal{A}$ is closed under search with respect to $\{\Omega, \mathcal{M}, \Lambda\}$, then there exists a topology $\mathcal{T}$ on Ω such that $\Delta_1^{\mathcal{A}}$ is precisely the set of continuous functions from $(\Omega, \mathcal{T})$ to $\mathcal{M}$.*

Proof. Let $\mathcal{T}$ be the topology generated by $\{\Gamma^{-1}(B) : \Gamma \in \mathcal{A}, B \subset \mathcal{M} \text{ basic open}\}$. Now, clearly any $\Gamma \in \mathcal{A}$ is continuous with respect to this topology. The fact that uniform limits of continuous functions into metric spaces are also continuous shows that any function in $\Delta_1^{\mathcal{A}}$ is continuous with respect to $\mathcal{T}$.

For the other direction, suppose that $f : (\Omega, \mathcal{T}) \to \mathcal{M}$ is continuous. Choose $\{c_i\}_{i\in\mathcal{I}} \subset \mathcal{M}$ such that $\mathcal{M} \subset \cup_{i\in\mathcal{I}} D(c_i, 2^{-n})$. Continuity of f implies that $f^{-1}(D(c_i, 2^{-n}))$ are open. This implies that there is an index set $\mathcal{J}$, natural numbers $\{n_{i,j}\}_{j\in\mathcal{J}}$, a family $\{\Gamma_{i,j,l}\}_{i\in\mathcal{I},j\in\mathcal{J},l\leq n_{i,j}}$ (in $\mathcal{A}$) and a family of basic open sets $\{U_{i,j,l}\}_{i\in\mathcal{I},j\in\mathcal{J},l\leq n_{i,j}}$ with the property that

$$f^{-1}(D(c_i, 2^{-n})) = \bigcup_{j\in\mathcal{J}} \bigcap_{l\leq n_{i,j}} \Gamma_{i,j,l}^{-1}(U_{i,j,l}).$$

It follows that

$$\bigcup_{i\in\mathcal{I},j\in\mathcal{J}} \bigcap_{l\leq n_{i,j}} \Gamma_{i,j,l}^{-1}(U_{i,j,l}) = \Omega.$$

Since $\mathcal{A}$ is closed under search, there exists $f_n \in \mathcal{A}$ such that for every $x \in \Omega$ there exists some $i \in \mathcal{I}$ and $j \in \mathcal{J}$ with $f_n(x) = c_i$ and for all $l \leq n_{i,j}$

$$x \in \Gamma_{i,j,l}^{-1}(U_{i,j,l}).$$

But this implies that $d(f_n(x), f(x)) < 2^{-n}$. Since n was arbitrary, we have $f \in \Delta_1^{\mathcal{A}}$. □

The generated topology can be very perverse and not every class of algorithms is closed under search. However, we do have the following useful theorem when Ω (and Λ) is a particularly simple discrete space, which shows that the SCI corresponds to the Baire hierarchy index.

Theorem 2.3.5. *Suppose that $\Omega = \{0,1\}^{\mathbb{N}} = \{\{a_i\}_{i\in\mathbb{N}} : a_i \in \{0,1\}\}$ with the set of evaluation functions Λ equal to the set of pointwise evaluations $\{\lambda_j(a) := a_j : j \in \mathbb{N}\}$ and let $\mathcal{M}$ be an arbitrary separable metric space with at least two separated points. Endow Ω with the product topology, $\tilde{\mathcal{T}}$, induced by the discrete topology on $\{0,1\}$ and consider the Baire hierarchy, $\{\mathcal{B}_\xi((\Omega,\tilde{\mathcal{T}}),\mathcal{M}) = \mathcal{B}_\xi\}_{\xi<\omega_1}$, of functions $f : \Omega \to \mathcal{M}$. Then for any problem function $\Xi : \Omega \to \mathcal{M}$ and $m \in \mathbb{N}$,*

$$\{\Xi,\Omega,\Lambda\} \in \Delta^G_{m+1} \Leftrightarrow \Xi \in \mathcal{B}_m.$$

In other words, the SCI corresponds to the Baire hierarchy index.

Remark 2.3.6. *The proof will make clear that we can replace Ω by $\{0,1\}^{\mathbb{N}\times\mathbb{N}}$ or any other such product space (induced by discrete topology) of the form A^B with A, B countable, with Λ the corresponding component-wise evaluations, as long as $\mathcal{M}$ has at least $|A|$ jointly separated points and is separable.*

Proof. First we show that general algorithms are closed under search and that the topology $\mathcal{T}$ in Proposition 2.3.4 is equal to the product topology $\tilde{\mathcal{T}}$. Without loss of generality we can assume that $\mathcal{I}$ is well-ordered by $\prec$. Given $x \in \Omega$, let $k \in \mathbb{N}$ be minimal such that there exists $i \in \mathcal{I}$ with $x \in \cap_{l\leq n_i}\Gamma_{i,l}^{-1}(U_{i,l})$ and $\Lambda_{\Gamma_{i,l}}(x) \subset \{\lambda_j : j \leq k\}$ for $l \leq n_i$. Let i_0 be the $\prec$-least index such that this holds for k and define $\Gamma(x) = c_{i_0}$. The well-ordering of $\mathcal{I}$ implies that Γ is a general algorithm and it clearly satisfies the requirements in the definition of closed under search. Note that this part of the proof only uses countability of Λ.

To equate the topologies, suppose that $\Gamma \in \Delta^G_0$ is a general algorithm. For each $a \in \Omega$, $\Lambda_\Gamma(a)$ is finite and we can assume without loss of generality that it is equal to $\{\lambda_j : j \leq I(a)\}$ for some finite $I(a)$. In particular, there exists an open set U_a such that any $b \in U_a$ has $\lambda_j(b) = \lambda_j(a)$ for $j \leq I(a)$ and hence $\Gamma(b) = \Gamma(a)$. Then for any open set $B \subset \mathcal{M}$

$$\Gamma^{-1}(B) = \bigcup_{a\in\Gamma^{-1}(B)} U_a$$

is open. Hence each Γ is continuous with respect to the product topology on Ω. It follows that $\mathcal{T} \subset \tilde{\mathcal{T}}$. To prove the converse, we must show that each projection map λ_j is continuous with respect to $\mathcal{T}$. Let x_1, x_2 be separated points in $\mathcal{M}$ and consider $f : \{0,1\} \to \mathcal{M}$ with $f(0) = x_1$ and $f(1) = x_2$. Then the composition $f \circ \lambda_j$ is a general algorithm and hence continuous with respect to $\mathcal{T}$. But this implies that λ_j is continuous. It follows from Proposition 2.3.4 that $\{\Xi,\Omega,\Lambda\} \in \Delta^G_1$ if and only if Ξ is continuous.

Now the space $(\Omega,\mathcal{T})$ is zero-dimensional and $\mathcal{M}$ is separable, hence by Theorem 2.3.2, any element of $\mathcal{B}_1$ is a limit of continuous functions. The converse holds in greater generality. It follows that $\Xi \in \mathcal{B}_m$ if and only if there are $f_{n_m,\dots,n_1} \in \Delta^G_1$ with

$$\Xi(a) = \lim_{n_m\to\infty} \dots \lim_{n_1\to\infty} f_{n_m,\dots,n_1}(a). \tag{2.3.1}$$

If this holds then there exists general algorithms $\Gamma_{n_m,\dots,n_1}$ such that for all $a \in \Omega$,

$$d(\Gamma_{n_m,\dots,n_1}(a), f_{n_m,\dots,n_1}(a)) \leq 2^{-n_1}$$

and hence

$$\lim_{n_m\to\infty} \dots \lim_{n_1\to\infty} \Gamma_{n_m,\dots,n_1}(a) = \Xi(a)$$

so that $\{\Xi, \Omega, \Lambda\} \in \Delta^G_{m+1}$. Conversely if $\{\Xi, \Omega, \Lambda\} \in \Delta^G_{m+1}$ with tower of algorithms $\Gamma_{n_m,\dots,n_1}$, then since each general algorithm is continuous, (2.3.1) holds with $f_{n_m,\dots,n_1}(a) = \Gamma_{n_m,\dots,n_1}$. □

2.3.3 Combinatorial problems high up in the SCI hierarchy

We can now combine the results of the previous two subsections and obtain combinatorial array problems high up in the SCI hierarchy. Let $k \in \mathbb{N}_{\geq 2}$ and let Ω_k denote the collection of all infinite arrays $\{a_{m_1,\dots,m_k}\}_{m_1,\dots,m_k\in\mathbb{N}}$ with entries $a_{m_1,\dots,m_k} \in \{0,1\}$. As usual Λ_k is the set of component-wise evaluations/projections. Consider the formulas

$$P(a, m_1, \dots, m_{k-2}) = \begin{cases} 1, & \text{if } \exists i\ \forall j\ \exists n > j \text{ s.t. } a_{m_1,\dots,m_{k-2},n,i} = 1 \\ 0, & \text{otherwise} \end{cases},$$

$$Q(a, m_1, \dots, m_{k-2}) = \begin{cases} 1, & \text{if } \forall^\infty i \forall j\ \exists n > j \text{ s.t. } a_{m_1,\dots,m_{k-2},n,i} = 1 \\ 0, & \text{otherwise} \end{cases},$$

where $\forall^\infty$ means 'for all but a finite number of'. In words, P decides whether the corresponding matrix has a column with infinitely many 1's, whereas Q decides whether the matrix has only finitely many columns with only finitely many 1's. For $R = P, Q$ consider the problem function for $a \in \Omega_k$

$$\Xi_{k,R}(a) = \begin{cases} \exists m_1\ \forall m_2 \dots \forall m_{k-2} R(a, m_1, \dots, m_{k-2}), & \text{if } k \text{ is even} \\ \forall m_1\ \exists m_2 \dots \forall m_{k-2} R(a, m_1, \dots, m_{k-2}), & \text{otherwise} \end{cases},$$

that is, so that all quantifier types alternate.

Theorem 2.3.7. *Let $\mathcal{M}$ be either $\{0,1\}$ with the discrete metric or $[0,1]$ with the usual metric and consider the above problems $\{\Xi_k, \Omega_k, \mathcal{M}, \Lambda_k\}$. For $k \in \mathbb{N}_{\geq 2}$ and $R = P, Q$,*

$$\Delta^G_{k+1} \not\ni \{\Xi_{k,R}, \Omega_k, \mathcal{M}, \Lambda_k\} \in \Delta^A_{k+2}.$$

In other words, we can solve the problem via a height $k+1$ arithmetic tower but it is impossible to do so with a height k general tower.

Proof. We will deal with the case of $R = P$ since the case of $R = Q$ is completely analogous. It is easy to see that $\{\Xi_{k,P}, \Omega_k, \mathcal{M}, \Lambda_k\} \in \Delta^A_{k+2}$. First consider the case $k = 2$ and set

$$\Gamma_{n_3,n_2,n_1}(a) = \max_{j\leq n_3} \chi_{(n_2,\infty)}\left(\sum_{i=1}^{n_1} a_{i,j}\right).$$

This is the decision problem that decides whether there exists a column with index at most n_3 such that there are at least n_2 1's in the first n_1 rows. This is clearly an arithmetic tower and it is straightforward to show that this converges to $\Xi_{2,P}$ in $\mathcal{M}$ (in either of the $\{0,1\}$ and $[0,1]$ cases). For $k > 2$ we simply alternate taking products (which corresponds to minima in this case) and maxima. Explicitly, we set

$$\Gamma_{n_{k+1},\dots,n_1}(a) = \begin{cases} \max\limits_{m_1\leq n_{k+1}} \prod\limits_{m_2=1}^{n_k} \dots \prod\limits_{m_{k-2}=1}^{n_4} \left\{\max\limits_{j\leq n_3} \chi_{(n_2,\infty)}\left(\sum\limits_{i=1}^{n_1} a_{m_1,\dots,m_{k-2},i,j}\right)\right\}, & \text{if } k \text{ is even} \\ \prod\limits_{m_1=1}^{n_{k+1}} \max\limits_{m_2\leq n_k} \dots \prod\limits_{m_{k-2}=1}^{n_4} \left\{\max\limits_{j\leq n_3} \chi_{(n_2,\infty)}\left(\sum\limits_{i=1}^{n_1} a_{m_1,\dots,m_{k-2},i,j}\right)\right\}, & \text{otherwise.} \end{cases}$$

Again, this is an arithmetic tower and it is straightforward to show that this converges to $\Xi_{k,P}$ in $\mathcal{M}$. It also holds that $\{\Xi_{k,P}, \Omega_k, \mathcal{M}, \Lambda_k\} \in \Sigma_{k+1}^{A}$ if k is even and $\{\Xi_{k,P}, \Omega_k, \mathcal{M}, \Lambda_k\} \in \Pi_{k+1}^{A}$ if k is odd (not to be confused with the notation for the Borel hierarchy).

Recall the topology $\mathcal{T}$ on Ω_k form Theorem 2.3.5. For the lower bound we note that P is Σ_3^0 complete (in the literature it is known as the problem 'S_3', see for example [KL87] section 23). This is terminology from the Wadge hierarchy, but in our case since $(\Omega_k, \mathcal{T})$ is zero-dimensional, a theorem of Wadge implies that this means that P is the indicator function of a set, also denoted by P, which lies in $\Sigma_3^0(\Omega_k)$ but not $\Pi_3^0(\Omega_k)$. It also follows that $\Xi_{k,P}$ is $\Sigma_{k+1}^0(\Omega_k)$ complete if k is even and $\Pi_{k+1}^0(\Omega_k)$ complete otherwise. Now suppose for a contradiction that $\{\Xi_{k,P}, \Omega_k, \mathcal{M}, \Lambda_k\} \in \Delta_{k+1}^{G}$. But then Theorem 2.3.5 implies that $\Xi_{k,P} \in \mathcal{B}_k(\Omega_k, \mathcal{M})$ and hence by Theorem 2.3.2, $\Xi_{k,P}$ is $\Sigma_{k+1}^0(\Omega_k)$ measurable. $\Xi_{k,P}$ is the indicator function of set, also denoted by $\Xi_{k,P}$, which is either $\Sigma_{k+1}^0(\Omega_k)$ or $\Pi_{k+1}^0(\Omega_k)$ complete depending on the parity of k. But 0 and 1 are separated in $\mathcal{M}$ and hence since $\Xi_{k,P}$ is $\Sigma_{k+1}^0(\Omega_k)$ measurable, $\Xi_{k,P}$ and its complement both lie in $\Sigma_{k+1}^0(\Omega_k)$. It follows that $\Xi_{k,P} \in \Sigma_{k+1}^0(\Omega_k) \cap \Pi_{k+1}^0(\Omega_k)$, contradicting the stated completeness. □

Throughout this course, we will make use of these theorem and analogous results for similar decision problems. In particular, we will use $\tilde{\Omega}$ to denote Ω_k and consider

$$\tilde{\Xi}_1 = \Xi_{2,P}, \quad \tilde{\Xi}_2 = \Xi_{2,Q}, \quad \tilde{\Xi}_3 = \Xi_{3,P}, \quad \tilde{\Xi}_4 = \Xi_{3,Q}.$$

We now have the framework and tools to study a range of infinite-dimensional spectral problems.

Chapter 3

Computing Spectra with Error Control

In this chapter, we consider the problem of computing the spectrum. This chapter is based on [CRH19, CHns]. The algorithms we develop compute spectra of a wide class of operators defined on separable Hilbert spaces. Moreover, the algorithms have the following desirable properties:

- They converge to the entire spectral set and avoid spectral pollution.
- They can be efficiently implemented.
- They are local and hence inherently parallelisable.
- They provide bounds on the error of the output, which converge to zero.
- In the self-adjoint (or normal) case, they provide 'approximate states'.

It has been a long-standing open problem to design such methods, even in the case of general one-dimensional discrete self-adjoint Schrödinger operators. Previous methods aimed at tackling the *general* problem either suffer from spectral pollution or do not converge to the full spectrum. Even in the cases where the finite section method converges, it only gives a Δ_2 algorithm (no error control). The problem of detecting spectral pollution is very difficult (see §7.3.2 for classification in the SCI hierarchy). The algorithms presented here are optimal in the sense of the SCI hierarchy described in Chapter 2, and can be used directly in many models in the physical sciences [JCN$^+$21, CHTW21].

The cases covered include unbounded operators on graphs and partial differential operators (PDOs), where we consider the determination of the spectrum from the coefficients of the PDO. In the case that the coefficients have locally bounded total variation on compact sets, we do this via point evaluations of the coefficients. The main idea, as outlined in §3.1.3, is to approximate the reciprocal of the resolvent norm, $\|R(z,A)\|^{-1}$, uniformly on compact subsets of $\mathbb{C}$, and use a local search routine.

3.1 Main Results

The spectrum (and pseudospectrum) of unbounded operators are closed but not necessarily bounded. When approximating the spectrum, we assume the operator has non-empty spectrum (for the SCI of testing if the spectrum intersected with a compact set is empty, see Theorem 3.1.6) and hence non-empty pseudospectra. Hence, we must introduce a metric on the set of non-empty closed subsets of $\mathbb{C}$, denoted by $\mathrm{Cl}(\mathbb{C})$.

Definition 3.1.1 (Attouch–Wets topology)**.** *The Attouch–Wets metric is defined by*

$$d_{\mathrm{AW}}(C_1, C_2) = \sum_{n=1}^{\infty} 2^{-n} \min\left\{1, \sup_{|x|\leq n} |\mathrm{dist}(x, C_1) - \mathrm{dist}(x, C_2)|\right\},$$

for $C_1, C_2 \in \mathrm{Cl}(\mathbb{C})$.

Throughout this section we take our metric space $(\mathcal{M}, d)$ to be $(\mathrm{Cl}(\mathbb{C}), d_{\mathrm{AW}})$. One should view this metric as a generalisation of the familiar Hausdorff metric on compact subsets defined in. We must be careful when defining the pseudospectrum, since the resolvent norm of an unbounded operator can be constant on open sets [Sha08]. The following definition agrees with the usual one for bounded operators.

Definition 3.1.2. *Let A be a closed and densely defined operator acting on a separable Hilbert space $\mathcal{H}$ and $\epsilon > 0$. We define the ($\epsilon-$)pseudospectrum of A by*

$$\mathrm{Sp}_\epsilon(A) = \mathrm{cl}\left(\left\{z \in \mathbb{C} : \|R(z, A)\|^{-1} < \epsilon\right\}\right),$$

the closure of the set of points with resolvent norm greater than $1/\epsilon$.

The pseudospectrum $\mathrm{Sp}_\epsilon(A)$ [KSTV15, TE05] is a generalisation of the spectrum (and measure of its stability), which is popular for non-Hermitian problems. The main results of this chapter, Theorems 3.1.4 and 3.1.9 below, also hold true when restricting the classes of operators to Schrödinger operators (on lattice systems in the discrete case and on $L^2(\mathbb{R}^d)$ or similar domains in the continuous case) and hence our results have direct implications within the computational boundaries in quantum mechanics [CRH19].

3.1.1 Spectra of unbounded operators on graphs

Consider a possibly unbounded operator A with domain $\mathcal{D}(A) \subset l^2(\mathbb{N})$ and non-empty spectrum, and

$$\Xi_1(A) = \mathrm{Sp}(A) \quad \text{and} \quad \Xi_2(A) = \mathrm{Sp}_\epsilon(A).$$

We have to define the domain Ω and evaluation functions Λ. Let $\mathcal{C}(l^2(\mathbb{N}))$ denote the set of closed, densely defined operators on $l^2(\mathbb{N})$, and consider the following assumptions.

(1) The subspace $\mathrm{span}\{e_n : n \in \mathbb{N}\}$ forms a core for both A and A^* ($\{e_j\}_{j\in\mathbb{N}}$ is the canonical basis).

(2) Given any $f : \mathbb{N} \to \mathbb{N}$ with $f(n) \geq n$ define

$$D_{f,n}(A) := \max\left\{\left\|(I - P_{f(n)})AP_n\right\|, \left\|(I - P_{f(n)})A^*P_n\right\|\right\}, \tag{3.1.1}$$

where P_n is the projection onto the span of $\{e_1, \ldots, e_n\}$. We say that an operator has bounded dispersion with respect to f if $\lim_{n\to\infty} D_{f,n}(A) = 0$. We will assume knowledge of a sequence $\{c_n\}_{n\in\mathbb{N}} \subset \mathbb{Q}$ that converges to zero with $D_{f,n}(A) \leq c_n$.

(3) We assume knowledge of a sequence $\{g_m\}$ of strictly increasing continuous functions $g_m : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ vanishing at 0 and with $\lim_{x\to\infty} g_m(x) = \infty$ such that

$$g_m(\mathrm{dist}(z, \mathrm{Sp}(A))) \leq \|R(z, A)\|^{-1}, \quad \forall z \in B_m(0). \tag{3.1.2}$$

In this case we say that A has resolvent bounded by $\{g_m\}$. Note that this implicitly assumes that the spectrum of A is non-empty (which always holds for bounded operators).

[DRAW PICTURE ON BOARD]

Bounded dispersion in (3.1.1) generalises the notion of a banded or sparse matrix to knowledge of off-diagonal decay. Given any operator with assumption (1), there exists an f such that $\lim_{n\to\infty} D_{f,n}(A) = 0$. The function f will be used to construct certain *rectangular* truncations of our operators (see §3.1.3), which is a key difference to previous methods that typically use *square* truncations.

To handle non-normal operators, we need to be able to control the resolvent as in (3.1.2). If A has $\mathrm{Sp}(A) \neq \emptyset$, then a simple compactness argument implies the existence of such a sequence of continuous functions. **Exercise:** Prove this! Suppose that A is bounded and we can take $g = g_m$, then we can view the function g as a measure of stability of the spectrum of A through the formula

$$\mathrm{Sp}_\epsilon(A) = \bigcup_{B\in\mathcal{B}(l^2(\mathbb{N})),\|B\|\leq\epsilon} \mathrm{Sp}(A+B).$$

Hence, the functions $\{g_m\}$ generalise the notion of condition number in the problem of computing $\mathrm{Sp}(A)$. Note that if our operator is normal, we can simply choose the functions $g_m(x) = g(x) = x$ through the identity $\mathrm{dist}(z, \mathrm{Sp}(A)) = \|R(z,A)\|^{-1}$. **Exercise:** Prove this! There are examples where such functions are known for non-normal operators, such as perturbations of self-adjoint operators [Gil03].

Defining Ω and Λ

Let f be as described in assumption (2) above, and $\hat{\Omega}$ be the class of all $A \in \mathcal{C}(l^2(\mathbb{N}))$ such that (1) and (2) hold and such that the spectrum is non-empty. Given a sequence as described in (3), let Ω_g be the class of all $A \in \hat{\Omega}$ such that (3.1.2) holds. We also let Ω_D denote the operators in $\hat{\Omega}$ that are diagonal.

Operators on graphs: For operators on graphs, consider any connected, undirected graph $\mathcal{G}$, such the set of vertices $V = V(\mathcal{G})$ is countably infinite. We consider operators on $l^2(V)$ that are closed, densely defined and of the form

$$A = \sum_{v,w\in V} \alpha(v,w)\left|v\right\rangle\left\langle w\right|, \tag{3.1.3}$$

for some $\alpha : V \times V \to \mathbb{C}$. We have also used the classical Dirac notation in (3.1.3) and identified any $v \in V$ by the element in $\psi_v \in l^2(V)$, such that $\psi_v(v) = 1$ and $\psi_v(w) = 0$ for $w \neq v$. When writing this, we assume that the linear span of such vectors forms a core of both A and its adjoint. We also assume that for any $v \in V$, the set of vertices w with $\alpha(v,w) \neq 0$ or $\alpha(w,v) \neq 0$ is finite. We then let $\Omega^{\mathcal{G}}$ be the class of all such A with non-empty spectrum and $\Omega_g^{\mathcal{G}}$ operators in $\Omega^{\mathcal{G}}$ of known $\{g_m\}$ such that (3.1.2) holds. We also assume that with respect to some given enumeration $\{e_1, e_2, ...\}$ of V, we have access to a function $S : \mathbb{N} \to \mathbb{N}$ such that if $m > S(n)$ then $\alpha(e_n, e_m) = \alpha(e_m, e_n) = 0$.

Remark 3.1.3 (Defining Λ). *For operators on $l^2(\mathbb{N})$, Λ contains the collection of matrix value evaluation functions, the functions describing the dispersion, and the family of the functions $\{g_m\}$ controlling the growth of the resolvent. For operators on $l^2(V)$, Λ contains the functions α, the function S and, in the case of $\Omega_g^{\mathcal{G}}$, the family g_m for $m \in \mathbb{N}$.*

Theorem 3.1.4. *Let Ξ_1 be the problem function $\mathrm{Sp}(\cdot)$ and Ξ_2 be the problem function $\mathrm{Sp}_\epsilon(\cdot)$ for $\epsilon > 0$, where these map into the metric space $(\mathrm{Cl}(\mathbb{C}), d_{\mathrm{AW}})$. Then*

$$\Delta_1^G \not\ni \{\Xi_1, \Omega_D\} \in \Sigma_1^A, \qquad \Delta_1^G \not\ni \{\Xi_1, \Omega_g\} \in \Sigma_1^A, \qquad \Delta_1^G \not\ni \{\Xi_1, \Omega_g^{\mathcal{G}}\} \in \Sigma_1^A,$$
$$\Delta_1^G \not\ni \{\Xi_2, \Omega_D\} \in \Sigma_1^A, \qquad \Delta_1^G \not\ni \{\Xi_2, \hat{\Omega}\} \in \Sigma_1^A, \qquad \Delta_1^G \not\ni \{\Xi_2, \Omega^{\mathcal{G}}\} \in \Sigma_1^A.$$

For Ξ_2, the constructed algorithm's output is always a subset of the true pseudospectrum.

Remark 3.1.5. *If any of the information given through the functions f or $\{g_m\}$ is missing, then the spectral problem does not lie in Δ_2^G (i.e., it cannot be computed in one limit, regardless of the model of computation). Hence the above conditions give a characterisation of when the spectral problem can be solved computationally in one limit. In other words, both types of information, the column decay structure and the conditioning of the spectrum, are needed.*

Finally, we consider two discrete problems which also include the case when the spectrum may be empty. Let K be a non-empty compact set in $\mathbb{C}$ and denote the collection of such subsets by $\mathcal{K}(\mathbb{C})$. Consider

$$\Xi_3 : (A, K) \to \text{Is } \mathrm{Sp}(A) \cap K = \emptyset?$$

$$\Xi_4 : (A, K) \to \text{Is } \mathrm{Sp}_\epsilon(A) \cap K = \emptyset?$$

The information we consider available to the algorithms in the $l^2(\mathbb{N})$ ($l^2(V(\mathcal{G}))$) case is given by the matrix elements of A (the functions α), the dispersion function f and dispersion bounds $\{c_n\}$ (the finite sets S_v), and a sequence of finite sets $K_n \subset \mathbb{Q} + i\mathbb{Q}$, with the property that $d_{\mathrm{H}}(K_n, K) \leq 2^{-(n+1)}$. For these problems, we take $(\mathcal{M}, d)$ to be $\{0, 1\}$ with the discrete metric (recall that 1 is interpreted as 'yes' and 0 as 'no'). Although the pseudospectrum is easier to compute as a whole, the following shows that this is not the case for testing on a given set. Note that these discrete problems are harder than computing the spectrum.

Theorem 3.1.6. *We have the following classifications for $j = 3, 4$:*

$$\Delta_2^G \not\ni \{\Xi_j, \hat{\Omega} \times \mathcal{K}(\mathbb{C})\} \in \Pi_2^A, \qquad \Delta_2^G \not\ni \{\Xi_j, \Omega_D \times \mathcal{K}(\mathbb{C})\} \in \Pi_2^A,$$

$$\Delta_2^G \not\ni \{\Xi_j, \Omega^{\mathcal{G}} \times \mathcal{K}(\mathbb{C})\} \in \Pi_2^A.$$

Furthermore, the proof will make clear that the lower bounds also hold when we restrict the allowed compact sets to any fixed compact subset of $\mathbb{R}$.

3.1.2 Spectra of partial differential operators

In this section, we provide classification results for general differential operators. Under very general assumptions, we obtain Σ_1^A classifications for the spectrum. Moreover, the computational problem can also be used for computer-assisted proofs. Finally, we establish how the problem makes a jump in the SCI hierarchy. In particular, with slightly weaker assumptions, the spectral problem $\notin \Sigma_1^G \cup \Pi_1^G$.

For $N \in \mathbb{N}$, consider the operator formally defined on $L^2(\mathbb{R}^d)$ by

$$Tu(x) = \sum_{k \in \mathbb{Z}^d_{\geq 0}, |k| \leq N} a_k(x) \partial^k u(x), \tag{3.1.4}$$

where we use multi-index notation with $|k| = \max\{|k_1|, ..., |k_d|\}$ and $\partial^k = \partial_{x_1}^{k_1} \partial_{x_2}^{k_2} ... \partial_{x_d}^{k_d}$. We will assume that the coefficients $a_k(x)$ are complex-valued measurable functions on $\mathbb{R}^d$. Suppose also that T can be defined on an appropriate domain $\mathcal{D}(T)$ such that T is closed and has a non-empty spectrum. Our aim is to compute the spectrum and pseudospectrum from the functions a_k.

Let Ω consist of all such T such that the following assumptions hold:

(1) The set $C_0^\infty(\mathbb{R}^d)$ of smooth, compactly supported functions forms a core of T and its adjoint T^*.

(2) The adjoint operator T^* can be initially defined on $C_0^\infty(\mathbb{R}^d)$ via

$$T^*u(x) = \sum_{k\in\mathbb{Z}^d_{\geq 0}, |k|\leq N} \widetilde{a}_k(x)\partial^k u(x),$$

where $\widetilde{a}_k(x)$ are complex-valued measurable functions on $\mathbb{R}^d$.

(3) For each $a_k(x)$ and $\widetilde{a}_k(x)$, there exists a positive constant A_k and an integer B_k such that

$$|a_k(x)|, |\widetilde{a}_k(x)| \leq A_k \left(1 + |x|^{2B_k}\right),$$

almost everywhere on $\mathbb{R}^d$, that is, we have at most polynomial growth.

(4) As in §3.1.1, we have access to functions $\{g_m\}$ (see (3.1.2) and the assumptions on $\{g_m\}$) such that

$$g_m(\mathrm{dist}(z, \mathrm{Sp}(T))) \leq \|R(z,T)\|^{-1}, \quad \forall z \in B_m(0).$$

(5) $\mathrm{Sp}(T)$ (and hence $\mathrm{Sp}_\epsilon(T)$) is non-empty.

Hence we consider the operator T defined as the closure of T acting on $C_0^\infty(\mathbb{R}^d)$. The initial domain $C_0^\infty(\mathbb{R}^d)$ is commonly encountered in applications, and it is straightforward to adapt our methods to other initial domains such as Schwartz space.

Remark 3.1.7 (The open problem of computing spectra of differential operators)**.** *There is no existing general theory or method guaranteeing convergence for PDOs (3.1.4), even when each a_k is a polynomial. The standard procedure is to discretise the differential operator via methods such as finite differences, truncate and then handle the finite matrix with standard algorithms designed for finite-dimensional problems. Such an approach does not always converge, and would at best give a Δ_2^A classification. Despite this, we prove below that one can achieve Σ_1 classification for a large class of operators.*

In the numerical applications, we will demonstrate this on anharmonic oscillators of the form

$$H = -\Delta + \sum_{j=1}^{d}(a_j x_j + b_j x_j^2) + \sum_{|\alpha|\leq M} c(\alpha)x^\alpha,$$

where $a_j, b_j, c(\alpha) \in \mathbb{R}$ (as well as more general Schrödinger operators). The multi-indices α are chosen such that $\sum_{|\alpha|\leq M} c(\alpha)x^\alpha$ is bounded from below. To the best of our knowledge, this algorithm is the first that computes the spectrum of such operators with error control in the sense of Σ_1^A. This has a wide number of applications and the problem has received a lot of attention [BO13, Wen96, BW73, FMT89].

Remark 3.1.8. *Throughout this section, the functions $\{g_m\}$ are not needed to compute the pseudospectrum.*

We consider the computation of the spectra/pseudospectra of operators $T \in \Omega$ from evaluations of the functions a_k and $\widetilde{a}_k$. For dimension d and $r > 0$ consider the space

$$\mathcal{A}_r = \{f \in M([-r,r]^d) : \|f\|_\infty + \mathrm{TV}_{[-r,r]^d}(f) < \infty\},$$

where $M([-r,r]^d)$ denotes the set of measurable functions on the hypercube $[-r,r]^d$ and $\mathrm{TV}_{[-r,r]^d}$ the total variation norm in the sense of Hardy and Krause (see [Nie92]). This space becomes a Banach algebra when equipped with the norm

$$\|f\|_{\mathcal{A}_r} = \|f\|_\infty + \sigma\mathrm{TV}_{[-r,r]^d}(f)$$

with $\sigma = 3^d + 1$ (see [BT89]). We will assume that each of the (appropriate restrictions of) a_k and $\widetilde{a}_k$ lie in $\mathcal{A}_r$ for all $r > 0$ and that we are given a sequence of positive numbers such that

$$\|a_k\|_{\mathcal{A}_n}, \|\widetilde{a}_k\|_{\mathcal{A}_n} \leq c_n, \quad c_n > 0, n \in \mathbb{N}, |k| \leq N. \tag{3.1.5}$$

The extra readable information is completely analogous to using bounded dispersion for matrix problems, and we shall see that it cannot be omitted if one wishes to gain error control in the sense of Σ_1. Let

$$\Omega_{\mathrm{TV}}^1 = \{T \in \Omega \mid \text{ such that (1) – (5) and (3.1.5) hold}\}.$$

In this case, Λ^1 contains functions that allow us to sample the functions $\{g_m\}_{m\in\mathbb{N}}$,$\{a_k, \widetilde{a}_k\}_{|k|\leq N}$ and the constants $\{A_k, B_k\}_{|k|\leq N}$, $\{c_n\}_{n\in\mathbb{N}}$. Consider the weaker assumption on Λ^1 that we can evaluate $b_n > 0$ (and not the A_k, B_k and the c_n) such that

$$\sup_{n\in\mathbb{N}} \frac{\max\{\|a_k\|_{\mathcal{A}_n}, \|\widetilde{a}_k\|_{\mathcal{A}_n} : |k| \leq N\}}{b_n} < \infty.$$

With a slight abuse of notation, we use Ω_{TV}^2 to denote the class of problems where we have this weaker requirement. We can now define the mappings

$$\Xi_j^1, \Xi_j^2 : \Omega_{\mathrm{TV}}^1, \Omega_{\mathrm{TV}}^2 \ni T \mapsto \begin{cases} \mathrm{Sp}(T) \in \mathcal{M}_{\mathrm{AW}}, & j = 1 \\ \mathrm{Sp}_\epsilon(T) \in \mathcal{M}_{\mathrm{AW}}, & j = 2. \end{cases}$$

Theorem 3.1.9. *Let $\Xi_j^1, \Xi_j^2, \Omega_{\mathrm{TV}}^1$ and Ω_{TV}^2 be as above. Then for $j = 1, 2$*

$$\Delta_1^G \not\ni \{\Xi_j^1, \Omega_{\mathrm{TV}}^1\} \in \Sigma_1^A,$$
$$\Sigma_1^G \cup \Pi_1^G \not\ni \{\Xi_j^2, \Omega_{\mathrm{TV}}^2\} \in \Delta_2^A.$$

3.1.3 Idea of the algorithms

To explain the idea of the algorithms, consider the case of computing the spectrum of a sparse self-adjoint $A \in \Omega_g$, such that the function f, which bounds the dispersion, also describes the sparsity structure in the sense that $A_{i,j} = 0$ if $j > f(i)$ or $i > f(j)$. Given z, we consider the rectangular matrix $P_{f(n)}(A - zI)P_n$. This was discussed in Section 1.2.2. In the case of finite range lattice models in condensed matter physics, which we can view as sparse matrices acting on $l^2(\mathbb{N})$, there is a nice physical interpretation. The rectangular truncation $P_{f(n)}AP_n$ contains all of the interactions of the first n sites without needing to apply boundary conditions. Using this, we approximate

$$E_n(z) \approx \sigma_{\inf}(P_{f(n)}(A - zI)|_{P_n(l^2(\mathbb{N}))}).$$

This corresponds to an estimate of the distance of z to the spectrum and physically corresponds to approximating the square root of the ground state energy of the folded Hamiltonian $P_n(A - zI)^*(A - zI)P_n$. We prove that our approximation converges uniformly to the resolvent norm $\|R(z,T)\|^{-1} = \mathrm{dist}(z, \mathrm{Sp}(A))$, on compact subsets of the complex plane. The convergence is also from above, meaning that we gain the rigorous error bound $\mathrm{dist}(z, \mathrm{Sp}(A)) \leq E_n(z)$. It is precisely the use of the rectangular truncation that leads to convergence from above, and, in general, taking a square truncation will not even converge. In the non-normal case, we use the functions $\{g_m\}$ to relate the approximation of $\|R(z,T)\|^{-1}$ to $\mathrm{dist}(z, \mathrm{Sp}(A))$.

Given a region $\mathcal{R} \subset \mathbb{C}$ of interest, the other ingredient of the algorithm is a search routine that seeks to approximate the spectrum locally on $\mathcal{R}$. We consider a grid of points $G_{\mathcal{R}}(n)$ of spacing $\delta(n) \to 0$ as

$n \rightarrow \infty$. The resolution $\delta(n)^{-1}$ (which can be viewed as a discretisation parameter) can be changed to allow one to vary the number of computed solutions. In our experiments, we chose $\delta(n)$ to ensure approximately n solutions for fair comparisons with other methods. The first step is to compute $E_n(\cdot)$ over $G_{\mathcal{R}}(n)$, which can be done in parallel. Given $z \in G_{\mathcal{R}}(n)$, we let I_z be the points in $G_{\mathcal{R}}(n)$ at distance most $E_n(z)$ away from z. We then let M_z be the minimisers of $E_n(\cdot)$ over the local set I_z. Since $E_n(\cdot)$ bounds the distance to the spectrum and converges to the true distance, M_z approximates the spectrum near the point z. This is a completely different approach to most previous methods, which typically seek to solve a finite-dimensional (linear and, in some cases, nonlinear) eigenvalue problem approximating the operator.

When dealing with PDOs, we construct an appropriate matrix representation of the operator with respect to a basis $\{\psi_n\}$ by sampling the coefficients. Our results rigorously indicate the sampling size and strategy needed, using the theory of quasi-Monte Carlo integration. We approximate inner products of the form

$$\langle (T - zI)\psi_m, (T - zI)\psi_n \rangle$$

directly, which allows us to compute a convergent upper bound of $\|R(z,T)\|^{-1}$. Once this is obtained, we can use a local search routine as before.

3.2 Proofs: Unbounded Operators on Graphs

We will now prove the theorems in §3.1.1. The following argument shows that it is sufficient to consider the $l^2(\mathbb{N})$ case. Given the graph $\mathcal{G}$ and enumeration $\{e_1, e_2, ...\}$ of the vertices, consider the induced isomorphism $l^2(V(\mathcal{G})) \cong l^2(\mathbb{N})$. This induces a corresponding operator on $l^2(\mathbb{N})$, where the functions α now become matrix values. For the lower bounds, we can consider diagonal operators in $\Omega^{\mathcal{G}}$ (that is, $\alpha(v,w) = 0$ if $v \neq w$) with the trivial choice of $S(n) = n$. Hence lower bounds for Ω_D translate to lower bounds for $\Omega^{\mathcal{G}}$ and $\Omega_g^{\mathcal{G}}$. For the upper bounds, the construction of algorithms for $l^2(\mathbb{N})$ will make clear that given the above isomorphism, we can compute a dispersion bounding function f for the induced operator on $l^2(\mathbb{N})$ simply by taking $f(n) = S(n)$. This has $D_{f,n}(A) = 0$. Note that any of the functions in Λ for the relevant class of operators on $l^2(\mathbb{N})$ can be computed via the above isomorphism using functions in Λ for the relevant class of operators on $l^2(V(\mathcal{G}))$. For instance, to evaluate matrix elements, we use $\alpha(e_i, e_j)$.

There is a useful characterisation of the Attouch–Wets topology. For any closed non-empty sets C and C_n, the convergence $d_{\mathrm{AW}}(C_n, C) \rightarrow 0$ holds if and only if $d_K(C_n, C) \rightarrow 0$ for any compact $K \subset \mathbb{C}$ where

$$d_K(C_1, C_2) = \max\left\{ \sup_{a \in C_1 \cap K} \mathrm{dist}(a, C_2), \sup_{b \in C_2 \cap K} \mathrm{dist}(b, C_1) \right\},$$

with the convention that the supremum over the empty set is 0. This occurs if and only if for any $\delta > 0$ and K, there exists N such that if $n > N$ then $C_n \cap K \subset C + B_\delta(0)$ and $C \cap K \subset C_n + B_\delta(0)$. Furthermore, it is enough to consider K of the form $B_m(0)$, the closed ball of radius m about the origin for $m \in \mathbb{N}$, for m large. Throughout this section we take our metric space $(\mathcal{M}, d)$ to be $(\mathrm{Cl}(\mathbb{C}), d_{\mathrm{AW}})$.

Remark 3.2.1 (A note on the empty set). *There is a slight subtlety regarding the empty set. It could be the case that the output of our algorithm is the empty set and hence $\Gamma_n(A)$ does not map to the required metric space. However, the proofs will make clear that for large n, $\Gamma_n(A)$ is non-empty and we gain convergence (this is also very rarely a problem in practice for $n \gtrsim 10$). By successively computing $\Gamma_n(A)$ and outputting $\Gamma_{m(n)}(A)$, where $m(n) \geq n$ is minimal with $\Gamma_{m(n)}(A) \neq \emptyset$, we see that this does not matter for the classification, but the algorithm in this case is adaptive.*

The following lemma is a useful criterion for determining Σ_1^A error control in the Attouch–Wets topology and will be used in the proofs without further comment.

Lemma 3.2.2. *Suppose that $\Xi : \Omega \to (\mathrm{Cl}(\mathbb{C}), d_{\mathrm{AW}})$ is a problem function and Γ_n is a sequence of arithmetic algorithms with each output a finite set such that*

$$\lim_{n\to\infty} d_{\mathrm{AW}}(\Gamma_n(A), \Xi(A)) = 0, \quad \forall A \in \Omega.$$

Suppose also that there is a function E_n provided by Γ_n (and defined over the output of Γ_n), such that

$$\lim_{n\to\infty} \sup_{z\in\Gamma_n(A)\cap B_m(0)} E_n(z) = 0$$

for all $m \in \mathbb{N}$ and such that

$$\mathrm{dist}(z, \Xi(A)) \leq E_n(z), \quad \forall z \in \Gamma_n(A).$$

Then:

1. *For each $m \in \mathbb{N}$ and given $\Gamma_n(A)$, we can compute in finitely many arithmetic operations and comparisons a sequence of non-negative numbers $a_n^m \to 0$ (as $n \to \infty$) such that*

$$\Gamma_n(A) \cap B_m(0) \subset \Xi(A) + B_{a_n^m}(0).$$

2. *Given $\Gamma_n(A)$, we can compute in finitely many arithmetic operations and comparisons a sequence of non-negative numbers $b_n \to 0$ such that*

$$\Gamma_n(A) \subset A_n$$

for some $A_n \in \mathrm{Cl}(\mathbb{C})$ with $d_{\mathrm{AW}}(A_n, \Xi(A)) \leq b_n$.

Hence we can convert Γ_n to a Σ_1^A tower using the sequence $\{b_n\}$ by taking subsequences if necessary.

Exercise: Prove Lemma 3.2.2.

To build our algorithms, we need to characterise the reciprocal of the resolvent norm in terms of the injection modulus. For $A \in \mathcal{C}(l^2(\mathbb{N}))$ define the injection modulus as

$$\sigma_{\mathrm{inf}}(A) = \inf\{\|Ax\| : x \in \mathcal{D}(A), \|x\| = 1\}, \tag{3.2.1}$$

and define the function

$$\gamma(z, A) = \min\{\sigma_{\mathrm{inf}}(A - zI), \sigma_{\mathrm{inf}}(A^* - \bar{z}I)\}.$$

Lemma 3.2.3. *For $A \in \mathcal{C}(l^2(\mathbb{N}))$, $\gamma(z, A) = 1/\|R(z, A)\|$, where $R(z, A)$ denotes the resolvent $(A - zI)^{-1}$ and we adopt the convention that $1/\|R(z, A)\| = 0$ if $z \in \mathrm{Sp}(A)$.*

Exercise: Prove Lemma 3.2.3.

Suppose we have a sequence of functions $\gamma_n(z, A)$ that converge uniformly to $\gamma(z, A)$ on compact subsets of $\mathbb{C}$. Define the grid

$$\texttt{Grid}(n) = \frac{1}{n}(\mathbb{Z} + i\mathbb{Z}) \cap B_n(0). \tag{3.2.2}$$

For a strictly increasing continuous function $g : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, with $g(0) = 0$ and $\lim_{x\to\infty} g(x) = \infty$, for $n \in \mathbb{N}$ and $y \in \mathbb{R}_{\geq 0}$ define

$$\texttt{CompInvg}(n, y, g) = \min\{k/n : k \in \mathbb{N}, g(k/n) > y\}. \tag{3.2.3}$$

Note that $\texttt{CompInvg}(n, y, g)$ can be computed from finitely many evaluations of the function g. We now build the algorithm converging to the spectrum step by step using the functions in (3.1.2). For each $z \in \texttt{Grid}(n)$, let

$$\Upsilon_{n,z} = B_{\texttt{CompInvg}(n,\gamma_n(z,A),g_{\lceil |z| \rceil})}(z) \cap \texttt{Grid}(n).$$

If $\gamma_n(z, A) > \left(|z|^2 + 1\right)^{-1}$ then set $M_z = \emptyset$, otherwise set

$$M_z = \{w \in \Upsilon_{n,z} : \gamma_n(w, A) = \min_{v \in \Upsilon_{n,z}} \gamma_n(v, A)\}.$$

Finally define $\Gamma_n(A) = \cup_{z \in \texttt{Grid}(n)} M_z$. It is clear that if $\gamma_n(z, A)$ can be computed in finitely many arithmetic operations and comparisons from the relevant functions in Λ for each problem, then this defines an arithmetic algorithm. If $A \in \mathcal{C}(l^2(\mathbb{N}))$ with non-empty spectrum then there exists $z \in B_m(0)$ with $\gamma(z, A) \leq (m^2 + 1)^{-1}/2$ and, for large n, $z_n \in \texttt{Grid}(n)$ sufficiently close to z with $\gamma(z_n, A) \leq (|z_n|^2 + 1)^{-1}$. Hence, by computing successive $\Gamma_n(A)$, we can assume that $\Gamma_n(A) \neq \emptyset$ without loss of generality (see Remark 3.2.1).

Proposition 3.2.4. *Suppose $A \in \mathcal{C}(l^2(\mathbb{N}))$ with non-empty spectrum and we have a function $\gamma_n(z, A)$ that converges uniformly to $\gamma(z, A)$ on compact subsets of $\mathbb{C}$. Suppose also that (3.1.2) holds, namely*

$$g_m(\mathrm{dist}(z, \mathrm{Sp}(A))) \leq \|R(z, A)\|^{-1}, \quad \forall z \in B_m(0).$$

Then $\Gamma_n(A)$ converges in the Attouch–Wets topology to $\mathrm{Sp}(A)$ (assuming $\Gamma_n(A) \neq \emptyset$ without loss of generality).

Proof. We use the characterisation of the Attouch–Wets topology. Suppose that $m \in \mathbb{N}$ is large such that $B_m(0) \cap \mathrm{Sp}(A) \neq \emptyset$. We must show that given $\delta > 0$, there exists N such that if $n > N$ then $\Gamma_n(A) \cap B_m(0) \subset \mathrm{Sp}(A) + B_\delta(0)$ and $\mathrm{Sp}(A) \cap B_m(0) \subset \Gamma_n(A) + B_\delta(0)$. Throughout the rest of the proof we fix such an m. Let $\epsilon_n = \|\gamma_n(\cdot, A) - \gamma(\cdot, A)\|_{\infty, B_{m+1}(0)}$, where the notation means the supremum norm over the set $B_{m+1}(0)$.

We deal with the second inclusion first. Suppose that $z \in \mathrm{Sp}(A) \cap B_m(0)$, then there exists some $w \in \texttt{Grid}(n)$ such that $|w - z| \leq 1/n$. It follows that

$$\gamma_n(w, A) \leq \gamma(w, A) + \epsilon_n \leq \mathrm{dist}(w, \mathrm{Sp}(A)) + \epsilon_n \leq \epsilon_n + 1/n.$$

By choosing n large, we can ensure that $\epsilon_n < (2m^2+2)^{-1}$ and that $1/n \leq (2m^2+2)^{-1}$ so that $\gamma_n(w, A) < (|w|^2 + 1)^{-1}$. It follows that M_w is non-empty. If $y \in M_w$ then

$$|y - z| \leq |w - z| + |y - w| \leq 1/n + 1/n + g_{\lceil w \rceil}^{-1}(\gamma_n(w, A)).$$

But the g_k's are non-increasing in k, strictly increasing continuous functions with $g_k(0) = 0$. Since $\gamma_n(w, A) \leq \epsilon_n + 1/n$, it follows that

$$|y - z| \leq 2/n + g_{m+1}^{-1}(\epsilon_n + 1/n). \tag{3.2.4}$$

There exists N_1 such that if $n \geq N_1$ then (3.2.4) holds and $2/n + g_{m+1}^{-1}(\epsilon_n + 1/n) \leq \delta$ and this gives the second inclusion.

For the first inclusion, suppose for a contradiction that this is false. Then there exists $n_j \to \infty$, $\delta > 0$ and $z_{n_j} \in \Gamma_{n_j}(A) \cap B_m(0)$ such that $\mathrm{dist}(z_{n_j}, \mathrm{Sp}(A)) \geq \delta$. Then $z_{n_j} \in M_{w_{n_j}}$ for some $w_{n_j} \in \texttt{Grid}(n_j)$. Let

$$I(j) = B_{\texttt{CompInvg}(n_j,\gamma_{n_j}(w_{n_j},A),g_{\lceil |w_{n_j}| \rceil})}(w_{n_j}) \cap \texttt{Grid}(n_j),$$

the set over which we compute minima of γ_{n_j}. Let $y_{n_j} \in \mathrm{Sp}(A)$ be of minimal distance to w_{n_j} (such a y_{n_j} exists since the spectrum restricted to any compact ball is compact). It follows that $\left|y_{n_j} - w_{n_j}\right| \leq g^{-1}_{\lceil |w_{n_j}| \rceil}(\gamma(w_{n_j}, A))$. A simple geometrical argument (which also works when we restrict everything to the real line for self-adjoint operators), shows that there must be a v_{n_j} in $I(j)$ so that

$$\left|v_{n_j} - y_{n_j}\right| \leq \frac{4}{n_j} + g^{-1}_{\lceil |w_{n_j}| \rceil}(\gamma(w_{n_j}, A)) - g^{-1}_{\lceil |w_{n_j}| \rceil}(\gamma_{n_j}(w_{n_j}, A)).$$

Since z_{n_j} minimises γ_{n_j} over $I(j)$ and $M_{w_{n_j}}$ is non-empty, it follows that

$$\gamma(z_{n_j}, A) \leq \gamma_{n_j}(z_{n_j}, A) + \epsilon_{n_j} \leq \min\left\{\frac{1}{\left|w_{n_j}\right|^2 + 1}, \gamma_{n_j}(v_{n_j}, A)\right\} + \epsilon_{n_j}.$$

This implies that

$$\delta \leq \mathrm{dist}(z_{n_j}, \mathrm{Sp}(A)) \leq g_m^{-1}\left(\min\left\{\frac{1}{\left|w_{n_j}\right|^2 + 1}, \gamma_{n_j}(v_{n_j}, A)\right\} + \epsilon_{n_j}\right), \tag{3.2.5}$$

where we recall that g_m^{-1} is continuous. It follows that the w_{n_j} must be bounded and hence so are the v_{n_j}. Due to the local uniform convergence of γ_n to γ, it follows that

$$\frac{4}{n_j} + g^{-1}_{\lceil |w_{n_j}| \rceil}(\gamma(w_{n_j}, A)) - g^{-1}_{\lceil |w_{n_j}| \rceil}(\gamma_{n_j}(w_{n_j}, A)) \to 0, \quad \text{as } n_j \to \infty.$$

But then

$$\gamma(v_{n_j}, A) \leq \mathrm{dist}(v_{n_j}, \mathrm{Sp}(A)) \leq \left|v_{n_j} - y_{n_j}\right| \to 0.$$

Again the local uniform convergence implies that $\gamma_{n_j}(v_{n_j}, A) \to 0$, which contradicts (3.2.5) and completes the proof. □

Next, given such a sequence γ_n, we would like to provide an algorithm for computing the pseudospectrum. However, care must be taken in the unbounded case since the resolvent norm can be constant on open subsets of $\mathbb{C}$ [Sha08]. Simply taking

$$\mathtt{Grid}(n) \cap \{z : \gamma_n(z, A) \leq \epsilon\}$$

is not guaranteed to converge, as can be seen in the case that γ_n is identically γ and A is such that $\|R(z, A)\|^{-1} = \epsilon$ has non-empty interior. To get around this, we will need an extra assumption on the functions γ_n.

Lemma 3.2.5. *Suppose $A \in \mathcal{C}(l^2(\mathbb{N}))$ with non-empty spectrum and let $\epsilon > 0$. Suppose we have a sequence of functions $\gamma_n(z, A)$ that converge uniformly to $\|R(z, A)\|^{-1}$ on compact subsets of $\mathbb{C}$. Set*

$$\Gamma_n^\epsilon(A) = \mathtt{Grid}(n) \cap \{z : \gamma_n(z, A) < \epsilon\}.$$

For large n, $\Gamma_n^\epsilon(A) \neq \emptyset$ so we can assume this without loss of generality. Suppose also $\exists N \in \mathbb{N}$ (possibly dependent on A but independent of z) such that if $n \geq N$ then $\gamma_n(z, A) \geq \|R(z, A)\|^{-1}$. Then $d_{\mathrm{AW}}(\Gamma_n^\epsilon(A), \mathrm{Sp}_\epsilon(A)) \to 0$ as $n \to \infty$.

Proof. Since the pseudospectrum is non-empty, for large n, $\Gamma_n^\epsilon(A) \neq \emptyset$ so by our usual argument of computing successive Γ_n^ϵ (see Remark 3.2.1) we may assume that this holds for all n without loss of generality. We use the characterisation of the Attouch–Wets topology. Suppose that m is large such that

$B_m(0)\cap \mathrm{Sp}_\epsilon(A)\neq\emptyset$. $\exists N\in\mathbb{N}$ such that if $n\geq N$ then $\gamma_n(z,A)\geq \|R(z,A)\|^{-1}$ and hence $\Gamma_n^\epsilon(A)\cap B_m(0)\subset \mathrm{Sp}_\epsilon(A)$. Hence we must show that given $\delta>0$, there exists N_1 such that if $n>N_1$ then $\mathrm{Sp}_\epsilon(A)\cap B_m(0)\subset\Gamma_n^\epsilon(A)+B_\delta(0)$. Suppose for a contradiction that this were false. Then there exists $z_{n_j}\in\mathrm{Sp}_\epsilon(A)\cap B_m(0)$, $\delta>0$ and $n_j\to\infty$ such that $\mathrm{dist}(z_{n_j},\Gamma_{n_j}^\epsilon(A))\geq\delta$. Without loss of generality, we can assume that $z_{n_j}\to z\in\mathrm{Sp}_\epsilon(A)\cap B_m(0)$. There exists some w with $\|R(w,A)\|^{-1}<\epsilon$ and $|z-w|\leq\delta/2$. Assuming $n_j>m+\delta$, there exists $y_{n_j}\in\texttt{Grid}(n_j)$ with $|y_{n_j}-w|\leq 1/n_j$. It follows that

$$\gamma_{n_j}(y_{n_j},A)\leq\left|\gamma_{n_j}(y_{n_j},A)-\gamma(y_{n_j},A)\right|+\left|\gamma(w,A)-\gamma(y_{n_j},A)\right|+\|R(w,A)\|^{-1}.$$

But γ is continuous and γ_{n_j} converges uniformly to γ on compact subsets. Hence for large n_j, it follows that $\gamma_{n_j}(y_{n_j},A)<\epsilon$ so that $y_{n_j}\in\Gamma_{n_j}^\epsilon(A)$. But $|y_{n_j}-z|\leq|z-w|+|y_{n_j}-w|\leq\delta/2+1/n_j$, which is smaller than δ for large n_j. This gives the required contradiction. □

Now suppose that $A\in\hat{\Omega}$ and let $D_{f,n}(A)\leq c_n$. The following shows that we can construct the required sequence $\gamma_n(z,A)$, each function output requiring finitely many arithmetic operations and comparisons of the corresponding input information.

Theorem 3.2.6. *Let $A\in\hat{\Omega}$ and define the function*

$$\tilde{\gamma}_n(z,A)=\min\{\sigma_{\mathrm{inf}}(P_{f(n)}(A-zI)|_{P_n(l^2(\mathbb{N}))}),\sigma_{\mathrm{inf}}(P_{f(n)}(A^*-\bar{z}I)|_{P_n(l^2(\mathbb{N}))})\}.$$

We can compute $\tilde{\gamma}_n$ up to precision $1/n$ using finitely many arithmetic operations and comparisons. We call this approximation $\hat{\gamma}_n$ and set

$$\gamma_n(z,A)=\hat{\gamma}_n(z,A)+c_n+1/n.$$

Then $\gamma_n(z,A)$ converges uniformly to $\gamma(z,A)$ on compact subsets of $\mathbb{C}$ and $\gamma_n(z,A)\geq\gamma(z,A)$.

Proof. We will first prove that $\sigma_{\mathrm{inf}}((A-zI)|_{P_n(l^2(\mathbb{N}))})\downarrow\sigma_{\mathrm{inf}}(A-zI)$ as $n\to\infty$. It is trivial that $\sigma_{\mathrm{inf}}((A-zI)|_{P_n(l^2(\mathbb{N}))})\geq\sigma_{\mathrm{inf}}(A-zI)$ and that $\sigma_{\mathrm{inf}}((A-zI)|_{P_n(l^2(\mathbb{N}))})$ is non-increasing in n. Using Lemma 3.2.3, let $\epsilon>0$ and $x\in\mathcal{D}(A)$ such that $\|x\|=1$ and $\|(A-zI)x\|\leq\sigma_{\mathrm{inf}}(A-zI)+\epsilon$. Since $\mathrm{span}\{e_n:n\in\mathbb{N}\}$ forms a core of A, $AP_{n_j}x_{n_j}\to Ax$ and $P_{n_j}x_{n_j}\to x$ for some $n_j\to\infty$ and some sequence of vectors x_{n_j} that we can assume have norm 1. It follows that for large n_j

$$\sigma_{\mathrm{inf}}((A-zI)|_{P_{n_j}(l^2(\mathbb{N}))})\leq\frac{\|(A-zI)P_{n_j}x_{n_j}\|}{\|P_{n_j}x_{n_j}\|}\to\|(A-zI)x\|\leq\sigma_{\mathrm{inf}}(A-zI)+\epsilon.$$

Since $\epsilon>0$ was arbitrary, this shows the convergence of $\sigma_{\mathrm{inf}}((A-zI)|_{P_n(l^2(\mathbb{N}))})$. The fact that $\mathrm{span}\{e_n:n\in\mathbb{N}\}$ forms a core of A^* can also be used to show that $\sigma_{\mathrm{inf}}((A-zI)^*|_{P_n(l^2(\mathbb{N}))})\downarrow\sigma_{\mathrm{inf}}(A^*-\bar{z}I)$.

Next we will use the assumption of bounded dispersion. For any bounded operators B,C, it holds that $|\sigma_{\mathrm{inf}}(A)-\sigma_{\mathrm{inf}}(B)|\leq\|A-B\|$. The definition of bounded dispersion now implies that

$$\left|\tilde{\gamma}_n(z,A)-\min\{\sigma_{\mathrm{inf}}((A-zI)|_{P_n(l^2(\mathbb{N}))}),\sigma_{\mathrm{inf}}((A-zI)^*|_{P_n(l^2(\mathbb{N}))})\}\right|\leq c_n.$$

The monotone convergence of $\min\{\sigma_{\mathrm{inf}}((A-zI)|_{P_n(l^2(\mathbb{N}))}),\sigma_{\mathrm{inf}}((A-zI)^*|_{P_n(l^2(\mathbb{N}))})\}$, together with Dini's theorem, imply that $\tilde{\gamma}_n(z,A)$ converges uniformly to the continuous function $\gamma(z,A)$ on compact subsets of $\mathbb{C}$ with $\tilde{\gamma}_n(z,A)+c_n\geq\gamma(z,A)$.

The proof will be complete if we can show that we can compute $\tilde{\gamma}_n(z, A)$ to precision $1/n$ using finitely many arithmetic operations and comparisons. To do this, consider the matrices

$$B_n(z) = P_n(A - zI)^* P_{f(n)}(A - zI)P_n, \quad C_n(z) = P_n(A - zI)P_{f(n)}(A - zI)^* P_n.$$

By an interval search routine, we can determine the smallest $l \in \mathbb{N}$ such that at least one of $B_n(z) - (l/n)^2 I$ or $C_n(z) - (l/n)^2 I$ has a negative eigenvalue. We then output l/n to get the $1/n$ bound. □

Note that by taking successive minima, $\upsilon_n(z, A) = \min_{1 \leq j \leq n} \gamma_n(z, A)$, we can obtain a sequence of functions υ_n that converge uniformly on compact subsets of $\mathbb{C}$ to $\gamma(z, A)$ monotonically from above. Hence without loss of generality, we will always assume that γ_n have this property.

Proof of Theorem 3.1.4. By considering bounded diagonal operators, it is straightforward to see that none of the problems (spectra or pseudospectra) lie in Δ_1^G. We first deal with convergence of height one arithmetical towers. For the spectrum, we use the function γ_n described in Theorem 3.2.6 together with Proposition 3.2.4 and its described algorithm. For the pseudospectrum, we use the same function γ_n described in Theorem 3.2.6 and convergence follows from using the algorithm in Proposition 3.2.5.

We are left with proving that our algorithms have Σ_1^A error control. For any $A \in \hat{\Omega}$, the output of the algorithm in Proposition 3.2.5 is contained in the true pseudospectrum since $\gamma_n(z, A) \geq \gamma(z, A) = \|R(z, A)\|^{-1}$. Hence we need only show that the algorithm in Proposition 3.2.4 provides Σ_1^A error control for input $A \in \Omega_g$. Denote the algorithm by Γ_n and set

$$E_n(z) = \texttt{CompInvg}(n, \gamma_n(z, A), g^{-1}_{\lceil |z| \rceil})$$

on $\Gamma_n(A)$ and zero on $\mathbb{C} \backslash \Gamma_n(A)$. Since $\gamma_n(z, A) \geq \|R(z, A)\|^{-1}$, the assumptions on $\{g_m\}$ imply that

$$\mathrm{dist}(z, \mathrm{Sp}(A)) \leq E_n(z), \quad \forall z \in \Gamma_n(A).$$

Suppose for a contradiction that E_n does not converge uniformly to zero on compact subsets of $\mathbb{C}$. Then there exists some compact set K, some $\epsilon > 0$, a sequence $n_j \to \infty$ and $z_{n_j} \in K$ such that $E_{n_j}(z_{n_j}) \geq \epsilon$. It follows that $z_{n_j} \in \Gamma_{n_j}(A)$. Without loss of generality, $z_{n_j} \to z$. By convergence of $\Gamma_{n_j}(A)$, $z \in \mathrm{Sp}(A)$ and hence $\gamma_{n_j}(z_{n_j}, A) \to \gamma(z, A) = 0$. Now choose M large such that $K \subset B_M(0)$. But then

$$E_{n_j}(z_{n_j}) \leq g_M^{-1}(\gamma_{n_j}(z_{n_j}, A)) + \frac{1}{n_j} \to 0,$$

the required contradiction. □

Remark 3.2.7. *The above makes it clear that $E_n(z)$ converges uniformly to the function $g^{-1}_{\lceil |z| \rceil}(\gamma(z, A))$ as $n \to \infty$ on compact subsets of $\mathbb{C}$.*

Finally, we consider the decision problems Ξ_3 and Ξ_4.

Proof of Theorem 3.1.6. It is clearly enough to prove the lower bounds for $\Omega_D \times \mathcal{K}(\mathbb{C})$ and the existence of towers for $\hat{\Omega} \times \mathcal{K}(\mathbb{C})$. The proof of lower bounds for $\Omega_D \times \mathcal{K}(\mathbb{C})$ can also be trivially adapted to the more restrictive versions of the problem described in the theorem.

Step 1: $\{\Xi_3, \Omega_D \times \mathcal{K}(\mathbb{C})\} \notin \Delta_2^G$. Suppose this were false, and Γ_n is a height one tower solving the problem. For every A and n there exists a finite number $N(A, n) \in \mathbb{N}$ such that the evaluations from $\Lambda_{\Gamma_n}(A)$ only take the matrix entries $A_{ij} = \langle Ae_j, e_i \rangle$ with $i, j \leq N(A, n)$ into account. Without loss of generality (by shifting our argument), we assume that $K \cap [0,1] = \{0\}$. We will consider the operators $A_m = \mathrm{diag}\{1, 1/2, ..., 1/m\} \in \mathbb{C}^{m\times m}$, $B_m = \mathrm{diag}\{1, 1, ..., 1\} \in \mathbb{C}^{m\times m}$ and $C = \mathrm{diag}\{1, 1, ...\}$. Set $A = \bigoplus_{m=1}^{\infty}(B_{k_m} \oplus A_{k_m})$, where we choose an increasing sequence k_m inductively as follows.

Set $k_1 = 1$ and suppose that $k_1, ..., k_m$ have been chosen. $\mathrm{Sp}(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C) = \{1, 1/2, ..., 1/m\}$ and hence

$$\Xi_3(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C) = 0,$$

so there exists some $n_m \geq m$ such that if $n \geq n_m$ then

$$\Gamma_n(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C) = 0.$$

Now let $k_{m+1} \geq \max\{N(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C, n_m), k_m + 1\}$. By assumption (iii) in Definition 2.1.1 it follows that $\Lambda_{\Gamma_{n_m}}(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C) = \Lambda_{\Gamma_{n_m}}(A)$ and hence by assumption (ii) in the same definition that $\Gamma_{n_m}(A) = \Gamma_{n_m}(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C) = 0$. But $0 \in \mathrm{Sp}(A)$ and so must have $\lim_{n\to\infty}\Gamma_n(A) = 1$, a contradiction.

Step 2: $\{\Xi_4, \Omega_D\} \notin \Delta_2^G$. The same proof as step 1, but replacing A by $A + \epsilon I$ works in this case.

Step 3: $\{\Xi_3, \hat{\Omega} \times \mathcal{K}(\mathbb{C})\} \in \Pi_2^A$. Recall that we can compute, with finitely many arithmetic operations and comparisons, a function γ_n that converges monotonically down to $\|R(z, A)\|^{-1}$ uniformly on compacts. Set

$$\Gamma_{n_2,n_1}(A) = \text{Does there exist some } z \in K_{n_2} \text{ such that } \gamma_{n_1}(z, A) < 1/2^{n_2}?$$

It is clear that this is an arithmetic algorithm since each K_n is finite and that

$$\lim_{n_1\to\infty}\Gamma_{n_2,n_1}(A) = \text{Does there exist some } z \in K_{n_2} \text{ such that } \|R(z, A)\|^{-1} < 1/2^{n_2}? =: \Gamma_{n_2}(A).$$

If $K \cap \mathrm{Sp}(A) = \emptyset$, then $\|R(z, A)\|^{-1}$ is bounded below on the compact set K and hence for large n_2, $\Gamma_{n_2}(A) = 0$. However, if $z \in \mathrm{Sp}(A) \cap K$ then let $z_{n_2} \in K_{n_2}$ minimise the distance to z. Then

$$\|R(z_{n_2}, A)\|^{-1} \leq \mathrm{dist}(z_{n_2}, \mathrm{Sp}(A)) < 1/2^{n_2}$$

and hence $\Gamma_{n_2}(A) = 1$ for all n_2. This also shows the Π_2^A classification.

Step 4: $\{\Xi_4, \hat{\Omega} \times \mathcal{K}(\mathbb{C})\} \in \Pi_2^A$. Set

$$\Gamma_{n_2,n_1}(A) = \text{Does there exist some } z \in K_{n_2} \text{ such that } \gamma_{n_1}(z, A) < 1/2^{n_2} + \epsilon?,$$

then the same argument used in step 3 works in this case. □

3.3 Proofs: Partial Differential Operators

Here we shall prove Theorem 3.1.9. The constructed algorithms involve technical error estimates with parameters depending on these estimates. In the construction of the algorithms, our strategy will be to reduce the problem to one handled by the proofs in §3.2. To do so, we must first select a suitable basis and then compute matrix values.

3.3.1 Construction of algorithms

We begin with the description for $d = 1$ and comment how this can easily be extended to arbitrary dimensions. As an orthonormal basis of $L^2(\mathbb{R})$ we choose the Hermite functions

$$\psi_m(x) = (2^m m!\sqrt{\pi})^{-1/2} e^{-x^2/2} H_m(x), m \in \mathbb{Z}_{\geq 0},$$

where H_n denotes the n-th Hermite polynomial defined by

$$H_n(x) = (-1)^n \exp(x^2) \frac{d^n}{dx^n} \exp(-x^2).$$

These obey the recurrence relations

$$\psi_m'(x) = \sqrt{\frac{m}{2}}\psi_{m-1}(x) - \sqrt{\frac{m+1}{2}}\psi_{m+1}(x) \tag{3.3.1}$$

$$x\psi_m(x) = \sqrt{\frac{m}{2}}\psi_{m-1}(x) + \sqrt{\frac{m+1}{2}}\psi_{m+1}(x). \tag{3.3.2}$$

We let $C_H(\mathbb{R}) = \mathrm{span}\{\psi_m : m \in \mathbb{Z}_{\geq 0}\}$. Note that since the Hermite functions decay like $e^{-x^2/2}$ (up to polynomials) and the functions a_k and $\widetilde{a}_k$ can only grow polynomially, the formal differential operator T and its formal adjoint T^* make sense as operators from $C_H(\mathbb{R})$ to $L^2(\mathbb{R})$. The next proposition says that we can use the chosen basis.

Proposition 3.3.1. *Consider an operator $T \in \Omega$. Then $C_H(\mathbb{R})$ forms a core of both T and T^*.*

Exercise: Prove Proposition 3.3.1.

The above analysis holds in higher dimensions by considering tensor products

$$C_H(\mathbb{R}^d) := \mathrm{span}\{\psi_{m_1} \otimes ... \otimes \psi_{m_d} \,|\, m_1, ..., m_d \in \mathbb{Z}_{\geq 0}\}$$

of Hermite functions. We will abuse notation and write $\psi_m = \psi_{m_1} \otimes ... \otimes \psi_{m_d}$. It will be clear from the context when we are dealing with the multi-dimensional case. In order to build the required algorithms with Σ_1^A error control, we need to select an enumeration of $\mathbb{Z}^d_{\geq 0}$ in order to represent T as an operator acting on $l^2(\mathbb{N})$. A simple way to do this is to consider successive half spheres $S_n = \{m \in \mathbb{Z}^d_{\geq 0} : |m| \leq n\}$. We list S_1 as $\{e_1, ..., e_{r_1}\}$ and given an enumeration $\{e_1, ..., e_{r_n}\}$ of S_n, we list $S_{n+1} \backslash S_n$ as $\{e_{r_n+1}, ..., e_{r_{n+1}}\}$. We will then list our basis functions as $e_1, e_2, ...$ with $\psi_m = e_{h(m)}$. In practice, it is often more efficient (especially for large d) to consider other orderings such as the hyperbolic cross [Lub08b], or, in the semiclassical regime, to use Hagedorn functions [LL20]. Now that we have a suitable basis, the next question to ask is how to recover the matrix elements of T. In §3.2 the key construction is a function, that can be computed from the information given to us, $\gamma_n(z, T)$, which also converges uniformly from above to $\|R(z, T)\|^{-1}$ on compact subsets of $\mathbb{C}$. Such a sequence of functions is given by

$$\Psi_n(z, T) := \min\{\sigma_{\inf}((T - zI)|_{P_n(l^2(\mathbb{N}))}), \sigma_{\inf}((T^* - \bar{z}I)|_{P_n(l^2(\mathbb{N}))})\}$$

as long as the linear span of the basis forms a core of T and T^*. In §3.2 we used the notion of bounded dispersion to approximate this function. Here we have no such notion, but we can use the information given to us to replace this. It turns out that to approximate $\gamma_n(z, T)$, it suffices to use the following.

Lemma 3.3.2. *Let $\epsilon > 0$ and $n \in \mathbb{N}$, and suppose that we can compute, with finitely many arithmetic operations and comparisons, the matrices*

$$\{W_n(z)\}_{ij} = \langle (T - zI)e_j, (T - zI)e_i \rangle + E_{ij}^{n,1}(z)$$
$$\{V_n(z)\}_{ij} = \langle (T - zI)^* e_j, (T - zI)^* e_i \rangle + E_{ij}^{n,2}(z)$$

for $1 \leq i, j \leq n$ where the entrywise errors $E_{i,j}^{n,1}$ and $E_{i,j}^{n,2}$ have magnitude at most ϵ. Then

$$\left| \Psi_n(z, T)^2 - \min\{\sigma_{\inf}(W_n), \sigma_{\inf}(V_n)\} \right| \leq n\epsilon.$$

It follows that if ϵ is known, we can compute $\Psi_n(z, T)^2$ to within $2n\epsilon$. If ϵ is unknown, then for any $\delta > 0$, we can compute $\Psi_n(z, T)^2$ to within $n\epsilon + \delta$. (In each case with finitely many arithmetic operations and comparisons.)

Proof. Given $\{W_n(z)\}_{ij}$, note that $(\{W_n(z)\}_{ij} + \overline{\{W_n(z)\}_{ji}})/2$ still has an entrywise absolute error bounded by ϵ. Hence without loss of generality we can assume that the approximations $W_n(z)$ and $V_n(z)$ are self-adjoint. Call the matrices with no errors $\tilde{W}_n(z)$ and $\tilde{V}_n(z)$ then note that

$$\min\{\sigma_{\inf}((T - zI)|_{P_n(l^2(\mathbb{N}))}), \sigma_{\inf}((T^* - \bar{z}I)|_{P_n(l^2(\mathbb{N}))})\}^2 = \min\{\sigma_{\inf}(\tilde{W}_n), \sigma_{\inf}(\tilde{V}_n)\}$$

and

$$\left| \min\{\sigma_{\inf}(\tilde{W}_n), \sigma_{\inf}(\tilde{V}_n)\} - \min\{\sigma_{\inf}(W_n), \sigma_{\inf}(V_n)\} \right| \leq \max\left\{ \left\| W_n - \tilde{W}_n \right\|, \left\| V_n - \tilde{V}_n \right\| \right\}. \quad (3.3.3)$$

But for a finite matrix M, we can bound $\|M\|$ by its Frobenius norm $\sqrt{\sum |M_{ij}|^2}$. Hence the right hand side of (3.3.3) is at most $n\epsilon$. In order to use finitely many arithmetic operations and comparisons, we note that given a self-adjoint positive semi-definite matrix M, we can compute $\sigma_{\inf}(M)$ to arbitrary precision using finitely many arithmetic operations and comparisons via the argument in the proof of Theorem 3.2.6. The lemma now follows. □

Finally, we will need some results from the subject of quasi-Monte Carlo numerical integration, which we use to build the algorithm. Note that with either no prior information concerning the coefficients or for large d, this is the type of approach one would use in practice. We start with some definitions and theorems which we include here for completeness. An excellent reference for these results is [Nie92].

Definition 3.3.3. *Let $\{t_1, ..., t_j\}$ be a sequence in $[0, 1]^d$ and let $\mathcal{K}$ denote all subsets of $[0, 1]^d$ of the form $\prod_{k=1}^{d} [0, y_k)$ for $y_k \in (0, 1]$. Then we define the star discrepancy of $\{t_1, ..., t_j\}$ to be*

$$D_j^*(\{t_1, ..., t_j\}) = \sup_{K \in \mathcal{K}} \left| \frac{1}{j} \sum_{k=1}^{j} \chi_K(t_j) - |K| \right|,$$

where χ_K denotes the characteristic function of K.

Definition 3.3.4 ([Hal60]). *For any integer $b \geq 2$, the radical-inverse function η_b is defined on $\mathbb{Z}_{\geq 0}$ by*

$$\eta_b(n) = \sum_{j=0}^{\infty} a_j(n) b^{-j-1},$$

where $n = \sum_{j=0}^{\infty} a_j(n) b^j$ is the (necessarily terminating) digit expansion of n. Given integers $b_1, ..., b_s \geq 2$, the Halton sequence $\{x_n\}_{n \in \mathbb{N}} \subset [0, 1]^s$ in the bases $b_1, ..., b_s$ is defined by

$$x_n = (\eta_{b_1}(n - 1), \eta_{b_2}(n - 1), ..., \eta_{b_s}(n - 1)).$$

Theorem 3.3.5 ([Hal60])**.** *If $\{t_k\}_{k\in\mathbb{N}}$ is the Halton sequence in $[0,1]^d$ in the pairwise relatively prime bases $q_1,...,q_d$, then*

$$D_j^*(\{t_1,...,t_j\}) < \frac{d}{j} + \frac{1}{j}\prod_{k=1}^{d}\left(\frac{q_k-1}{2\log(q_k)}\log(j) + \frac{q_k+1}{2}\right).$$

Note that given d (and suitable q_1,..., q_d), we can easily compute in finitely many arithmetic operations and comparisons a constant $C(d)$ such that the above implies

$$D_j^*(\{t_1,...,t_j\}) < C(d)\frac{(\log(j)+1)^d}{j}. \tag{3.3.4}$$

The following theorem says why this is useful.

Theorem 3.3.6 (Koksma–Hlawka inequality [Nie92])**.** *If f has bounded variation $\mathrm{TV}_{[0,1]^d}(f)$ on the hypercube $[0,1]^d$ then for any $t_1,...,t_j$ in $[0,1]^d$*

$$\left|\frac{1}{j}\sum_{k=1}^{j} f(t_k) - \int_{[0,1]^d} f(x)dx\right| \leq \mathrm{TV}_{[0,1]^d}(f)D_j^*(\{t_1,...,t_j\}).$$

By re-scaling, if f has bounded variation $\mathrm{TV}_{[-r,r]^d}(f)$ and $s_k = 2rt_k - (r,r,...,r)^T$ then we obtain

$$\left|\frac{(2r)^d}{j}\sum_{k=1}^{j} f(s_k) - \int_{[-r,r]^d} f(x)dx\right| \leq (2r)^d\cdot\mathrm{TV}_{[-r,r]^d}(f)D_j^*(\{t_1,...,t_j\}).$$

Finally, in order to deal with our choice of basis, we need the following.

Lemma 3.3.7. *Consider the tensor product $\psi_m(x) := \psi_{m_1}(x_1)\cdot...\cdot\psi_{m_1}(x_d)$ in d dimensions and let $r>0$. Then*

$$\mathrm{TV}_{[-r,r]^d}(\psi_m) \leq \left(1 + 2r\sqrt{2(|m|+1)}\right)^d - 1.$$

Exercise: Prove Lemma 3.3.7.

Proposition 3.3.8. *Given $T\in\Omega_{\mathrm{TV}}^1$ and $\epsilon>0$, we can approximate the matrix values*

$$\langle (T-zI)\psi_m, (T-zI)\psi_n\rangle \quad \textit{and} \quad \langle (T-zI)^*\psi_m, (T-zI)^*\psi_n\rangle$$

to within ϵ using finitely many arithmetical operations and comparisons of the relevant information (captured by Ξ_j^1 in §3.1.2) given to us in each class.

Proof. Let $T\in\Omega_{\mathrm{TV}}^1$ and $\epsilon>0$. Recall that

$$T = \sum_{|k|\leq N} a_k(x)\partial^k, \quad T^* = \sum_{|k|\leq N}\widetilde{a}_k(x)\partial^k,$$

so by expanding out the inner products and also considering the case $a_k = 1$, it is sufficient to approximate

$$\langle a_k\partial^k\psi_m, a_j\partial^j\psi_n\rangle \quad \text{and} \quad \langle \widetilde{a}_k\partial^k\psi_m, \widetilde{a}_j\partial^j\psi_n\rangle$$

for all relevant k, j, m and n. Due to the symmetry in the assumptions of T and T^*, we only need to show that one can compute the first inner product, the proof for the second one is identical. Note that by the specific choice of the basis functions ψ_m, it follows that $\partial^k\psi_m$ can be written as a finite linear combination of tensor products of Hermite functions using the recurrence relations (the coefficients in the linear combinations are thus recursively defined as a function of k). Hence, in the inner product, we can assume that there are no partial derivatives. In doing this, we have assumed that we can compute square roots of integers (which occur in the coefficients) to arbitrary precision (recall we want an arithmetic tower) which can be achieved by a simple interval bisection routine. It follows that we only need to consider approximations of inner products of the form $\langle a_k\psi_m, a_j\psi_n\rangle$.

To do so let $R>1$ then, by Hölder's inequality and the assumption of polynomially bounded growth on the coefficients a_k, we have

$$\begin{aligned}&\int_{|x_i|\geq R}|a_k\overline{a_j}|\,|\psi_m\psi_n|\,dx\\&\leq A_kA_j\left(\int_{|x_i|\geq R}\left(1+|x|^{2B_k}\right)^2\left(1+|x|^{2B_j}\right)^2\psi_m(x)^2dx\right)^{1/2}\left(\int_{|x_i|\geq R}\psi_n(x)^2dx\right)^{1/2}.\end{aligned}$$

The first integral on the right hand side can be bounded by

$$16\int_{\mathbb{R}^d}|x|^{2B}\,\psi_m(x)^2dx\leq 16\int_{\mathbb{R}^d}\left(x_1^2+...+x_d^2\right)^B\psi_m(x)^2dx,$$

for $B=4(B_k+B_j)$, since we restrict to $|x_i|\geq R$ with $R>1$ and $|x|\leq\|x\|_2$. B is even so we can expand out the product $(x_1^2+...+x_d^2)^{B/2}\psi_m$ using the recurrence relations for the Hermite functions. In one dimension this gives

$$\begin{aligned}x\psi_m(x)&=\sqrt{\frac{m}{2}}\psi_{m-1}(x)+\sqrt{\frac{m+1}{2}}\psi_{m+1}(x),\\x^2\psi_m(x)&=\sqrt{\frac{m}{2}}x\psi_{m-1}(x)+\sqrt{\frac{m+1}{2}}x\psi_{m+1}(x),\\&=\sqrt{\frac{m}{2}}\left(\sqrt{\frac{m-1}{2}}\psi_{m-2}(x)+\sqrt{\frac{m}{2}}\psi_m(x)\right)+\sqrt{\frac{m+1}{2}}\left(\sqrt{\frac{m+1}{2}}\psi_m(x)+\sqrt{\frac{m+2}{2}}\psi_{m+2}(x)\right),\end{aligned}$$

and so on. We can do the same for tensor products of Hermite functions. In particular, multiplying a tensor product of Hermite functions, ψ_m, by $(x_1^2+...+x_d^2)$ induces a linear combination of at most $4d$ such tensor products, each with a coefficient of magnitude at most $(|m|+2)^2$ and index with l^∞ norm bounded by $|m|+2$ (allowing repetitions). It follows that $(x_1^2+...+x_d^2)^{B/2}\psi_m$ can be written as a linear combination of at most $(4d)^{B/2}$ such tensor products, each with a coefficient of magnitude at most $(|m|+B)^B$. Squaring this and integrating, the orthogonality and normalisation of the tensor product of Hermite functions implies that

$$16\int_{\mathbb{R}^d}(x_1^2+...+x_d^2)^B\psi_m(x)^2dx\leq 16(4d)^{B/2}(|m|+B)^{2B}=:p_1(|m|).$$

For the other integral, define $p_2(|n|):=4d(|n|+2)^4$. We then have

$$\int_{|x_i|\geq R}\psi_n^2dx\leq\frac{1}{R^4}\int_{\mathbb{R}^d}|x|^4\,\psi_n^2dx\leq\frac{p_2(|n|)}{R^4},$$

by using the same argument as above but with $B=2$.

So given $\delta>0$ and n,m,B,A_k,A_j, (and d) we can choose $r\in\mathbb{N}$ large such that

$$\int_{|x_i|\geq r}|a_k\overline{a_j}|\,|\psi_m\psi_n|\,dx\leq A_kA_j\frac{p_1(|m|)^{1/2}p_2(|n|)^{1/2}}{r^2}\leq\delta.$$

We now have to consider the cases $T\in\Omega_{\mathrm{TV}}^1$ or $T\in\Omega_{\mathrm{AN}}^1$ separately, noting that it is sufficient to approximate the integral $\int_{|x_i|\leq r}a_k\overline{a_j}\psi_m\psi_n dx$ to any given precision. For notational convenience, let

$$L_r(m)=\left[1+\sigma\left(\left(1+2r\sqrt{2(|m|+1)}\right)^d-1\right)\right]$$

so that with $\sigma = 3^d + 1$ as in the definition of $\|\cdot\|_{\mathcal{A}_r}$, we have via Lemma 3.3.7 that $\|\psi_m\|_{\mathcal{A}_r} \leq L_r(m)$.

Given k, j, m, n, δ and $r \in \mathbb{N}$ as above, choose M large such that

$$(2r)^d \cdot \frac{C(d)\big(\log(M)+1\big)^d}{M} \cdot c_r^2 \cdot L_r(m) \cdot L_r(n) \leq \delta/2, \tag{3.3.5}$$

where $C(d)$ is as (3.3.4) and c_r controls the total variation as in (3.1.5). Again, note that such an M can be chosen in finitely many arithmetic operations and comparisons with the given data and assuming that logarithms and square roots can be computed to arbitrary precision (say by a power series representation and bound on the remainder). Using the fact that $\mathcal{A}_r$ is a Banach algebra (in particular we can bound the norms of product of functions by the product of their norms) and Theorem 3.3.6, it follows that

$$\left| \frac{(2r)^d}{M} \sum_{l=1}^{M} a_k(s_l)\overline{a_j}(s_l)\psi_m(s_l)\psi_n(s_l) - \int_{|x_i|\leq r} a_k\overline{a_j}\psi_m\psi_n dx \right| \leq \delta/2,$$

where $s_l = 2rt_l - (r, r, ..., r)^T$ are the rescaled Halton points. Hence it is enough to show that each product $a_k(s_l)\overline{a_j}(s_l)\psi_m(s_l)\psi_n(s_l)$ can be computed to a given accuracy using finitely many arithmetic operations and comparisons. Since each $s_l \in \mathbb{Q}^d$ we can evaluate $a_k(s_l)\overline{a_j}(s_l)$. Note that we can compute $\exp(-x^2/2)$ to arbitrary precision with finitely many arithmetic operations and comparisons (again say by a power series representation and bound on the remainder) and that we can compute the coefficients of the polynomials Q_m with $\psi_m(x) = Q_m(x)\exp(-x^2/2)$, using the recursion formulae to any given precision, it follows that we can compute $\psi_m(s_l)\psi_n(s_l)$ to a given accuracy using finitely many arithmetic operations and comparisons. Using the bounds on the a_k and $\overline{a_j}$ and Cramér's inequality, we can bound the error in the product and hence the result follows. □

We can now prove the positive parts of Theorem 3.1.9.

Proof of inclusions in Theorem 3.1.9. **Step 1**: $\{\Xi_1^1, \Omega_{\mathrm{TV}}^1\} \in \Sigma_A^1$. The proof of this simply strings together the above results. The linear span of $\{e_1, e_2, ...\}$ (the reordered Hermite functions) is a core of T and T^* by Proposition 3.3.1. By Proposition 3.3.8, we can compute the inner products $\langle (T - zI)e_j, (T - zI)e_i\rangle$ and $\langle (T - zI)^*e_j, (T - zI)^*e_i\rangle$ up to arbitrary precision with finitely many arithmetic operations and comparisons. Using Lemma 3.3.2, given $z \in \mathbb{C}$, we can compute some approximation $\upsilon_n(z, T)$ in finitely many arithmetic operations and comparisons such that

$$\left|\upsilon_n(z,T)^2 - \min\{\sigma_{\mathrm{inf}}((T - zI)|_{P_n(l^2(\mathbb{N}))}), \sigma_{\mathrm{inf}}((T^* - \bar{z}I)|_{P_n(l^2(\mathbb{N}))})\}^2\right| \leq \frac{1}{n^2}.$$

We now set

$$\gamma_n(z, T) = \upsilon_n(z, T) + 1/n. \tag{3.3.6}$$

Then γ_n satisfies the hypotheses of Proposition 3.2.4. The proof of Theorem 3.1.4 also makes clear that we have error control since $\gamma_n(z, T) \geq \|R(z, T)\|^{-1}$.

Step 2: $\{\Xi_2^1, \Omega_{\mathrm{TV}}^1\} \in \Sigma_A^1$. Consider the sequence of functions γ_n defined by equation (3.3.6). These converge uniformly to $\|R(z, T)\|^{-1}$ on compact subsets of $\mathbb{C}$ and satisfy $\gamma_n(z, T) \geq \|R(z, T)\|^{-1}$. We can now apply Proposition 3.2.5.

Step 3: $\{\Xi_1^2, \Omega_{\mathrm{TV}}^2\}, \{\Xi_2^2, \Omega_{\mathrm{TV}}^2\} \in \Delta_2^A$. Let $T \in \Omega_{\mathrm{TV}}^2$. Our strategy will be to compute the inner products $\langle (T - zI)e_j, (T - zI)e_i\rangle$ and $\langle (T - zI)^*e_j, (T - zI)^*e_i\rangle$ to an error which decays rapidly enough as we let the cut-off parameter r tend to ∞. We follow the proof of Proposition 3.3.8 closely. Recall that given n, m, we can choose $r \in \mathbb{N}$ large such that

$$\int_{|x_i|\geq r} |a_k\overline{a_j}|\,|\psi_m\psi_n|\,dx \leq A_kA_j\frac{p_1(|m|)^{1/2}p_2(|n|)^{1/2}}{r^2},$$

with the crucial difference that now we do not assume we can compute A_k, A_j, p_1 or p_2. It follows that there exists some polynomial p_3, with coefficients not necessarily computable from the given information, such that

$$\int_{|x_i|\geq r} |a_k\overline{a_j}|\,|\psi_m\psi_n|\,dx \leq \frac{p_3(|m|,|n|)}{r^2},$$

for all $|j|, |k| \leq N$. Now we use the sequence b_r to bound the error in the integral over the compact cube asymptotically. We assume without loss of generality that b_r is increasing monotonically to ∞ with r. Using Halton sequences and the same argument in the proof of Proposition 3.3.8, we can approximate $\int_{|x_i|\leq r} a_k\overline{a_j}\psi_m\psi_n dx$, with an error that, asymptotically up to some unknown constant, is bounded by

$$r^d\cdot\frac{\big(\log(M)+1\big)^d}{M}\cdot b_r^2\cdot L_r(m)\cdot L_r(n), \tag{3.3.7}$$

where M is the number of Halton points. We can let M depend on r, n and m such that (3.3.7) is bounded by a constant times $1/r^2$. It follows that we can bound the total error in approximating $\langle a_k\psi_m, a_j\psi_n\rangle$ for any j, k by $p_3(|m|,|n|)/r^2$, by making the coefficients of p_3 larger if necessary. We argue similarly for the adjoint and note that $\langle(T-zI)\psi_m,(T-zI)\psi_n\rangle$ and $\langle(T-zI)^*\psi_m,(T-zI)^*\psi_n$ are both approximated to within

$$(1+|z|^2)\frac{P(|m|,|n|)}{r^2},$$

for some unknown polynomial P. Hence we can apply Lemma 3.3.2 (the form where we do not know the error in inner product estimates), changing the polynomial P to take into account the basis mapping from $\mathbb{Z}^d_{\geq 0}$ to $\mathbb{N}$ to some polynomial Q, to gain some approximation $\upsilon_n(z,T)$ in finitely many arithmetic operations and comparisons such that

$$\big|\upsilon_n(z,T)^2-\min\{\sigma_{\mathrm{inf}}((T-zI)|_{P_n(l^2(\mathbb{N}))}),\sigma_{\mathrm{inf}}((T^*-\bar{z}I)|_{P_n(l^2(\mathbb{N}))})\}^2\big| \leq \frac{n(1+|z|^2)Q(n)}{r(n,z)^2}+\frac{1}{n^3}. \tag{3.3.8}$$

We now choose $r(z,n)$ larger if necessary such that $r(z,n)\geq(1+|z|^2)\exp(n)$. We now set $\gamma_n(z,T)=\upsilon_n(z,T)+1/n$. Then γ_n satisfies the hypotheses of Proposition 3.2.4 and Proposition 3.2.5 since the error in (3.3.8) decays faster than $1/n^2$. We can use these propositions to build the required arithmetical algorithm. □

3.3.2 Proofs of impossibility results

Recall the maps

$$\Xi_j^1,\Xi_j^2:\Omega_{\mathrm{TV}}^1,\Omega_{\mathrm{TV}}^2\ni T\mapsto\begin{cases}\mathrm{Sp}(T)\in\mathcal{M}_{\mathrm{AW}} & j=1\\ \mathrm{Sp}_\epsilon(T)\in\mathcal{M}_{\mathrm{AW}} & j=2,\end{cases}$$

We split up the arguments to deal with Ω_{TV}^1 and then Ω_{TV}^2.

Proof that $\{\Xi_j,\Omega_{\mathrm{TV}}^1\}\notin\Delta_1^G$. Suppose first for a contradiction that a height one tower, Γ_n, exists for the problem $\{\Xi_1,\Omega_{\mathrm{TV}}^1\}$ such that $d_{\mathrm{AW}}(\Gamma_n(T),\Xi_1(T))\leq 2^{-n}$. We will deal with the one-dimensional case and higher dimensions are similar. Let $\rho(x)$ be any smooth bump function with maximum value 1, minimum value 0 and support $[0,1]$. Let ρ_n denote the translation of ρ to have support $[n,n+1]$. We will consider the two (self-adjoint and bounded) operators

$$(T_0u)(x)=0, \qquad (T_mu)(x)=\rho_m(x)u(x),$$

which have spectra $\{0\}$ and $[0,1]$ respectively. For these we can take the polynomial bound (the $\{A_k\}$ and $\{B_k\}$) to be 1 and the total variation bound to be $c_r=1+\sigma\mathrm{TV}_{[0,1]}(\rho)$. When we compute $\Gamma_2(T_0)$, we only use finitely many evaluations of the coefficient function $a_0(x)=0$ (as well as the other given information). We can then choose m large such that the support of ρ_m does not intersect the points of evaluation. By assumptions (ii) and (iii) in Definition 2.1.1, $\Gamma_2(T_m)=\Gamma_2(T_0)$. But this contradicts the triangle inequality since $d_{\mathrm{AW}}(\{0\},[0,1])\geq 1$

To argue for the pseudospectrum let $\epsilon>0$ and note that $2\epsilon\notin\mathrm{Sp}_\epsilon(T_0)$ but $2\epsilon\in\mathrm{Sp}_\epsilon(\epsilon T_m)$. We now alter the given c_r to $\epsilon(1+\sigma\mathrm{TV}_{[0,1]}(\rho))$ and the polynomial bound to ϵ. The argument is now exactly as before. Namely, we choose n large such that

$$d_{\mathrm{AW}}(\Gamma_n(T_0),[-\epsilon,2\epsilon])>2^{-n}$$

then choose m large such that $\Gamma_n(T_0)=\Gamma_n(\epsilon T_m)$. □

Exercise: Prove that $\{\Xi_j,\Omega^2_{\mathrm{TV}}\}\notin\Sigma_1^G\cup\Pi_1^G$.

3.4 Numerical Examples and Applications

We now demonstrate the broad applicability of the algorithm(s) of this chapter by a few test examples. Examples of discrete operators are given first, including quasicrystals, the NSA Anderson model and open systems in optics. We end with a selection of examples of PDOs.

3.4.1 Quasicrystals

We first revisit the quasicrystal example from Chapter 1. The free Hamiltonian H_0 (Laplacian) is given by

$$(H_0\psi)_i=\sum_{i\sim j}(\psi_j-\psi_i)\,, \tag{3.4.1}$$

with the notation $i\sim j$ meaning sites i and j are connected by an edge and hence summation is over nearest neighbour sites (vertices). Previous numerical methods study the eigenvalues of the Hamiltonian restricted to a finite portion of the tiling with a choice of boundary conditions at the edges (finite section method). However, this causes additional eigenvalues (spectral pollution or 'edge states') to appear, which are not in the spectrum of H_0 acting on the infinite tiling. We will compare our method to finite section with open boundary conditions (truncating the tile and the corresponding matrix without applying additional boundary conditions), and the method of approximating an aperiodic tiling by periodic approximants [TFUT91].

Figure 3.1 (left) shows the output of the algorithm of this chapter for $n=10^5$ and the two finite section methods, with n the number of vertices used in the computation. It is important to note that the new algorithm uses the same number of vertices of the tile as the finite section method for a given n. The error estimate, computed for both the new algorithm as well as the finite section methods using the method in

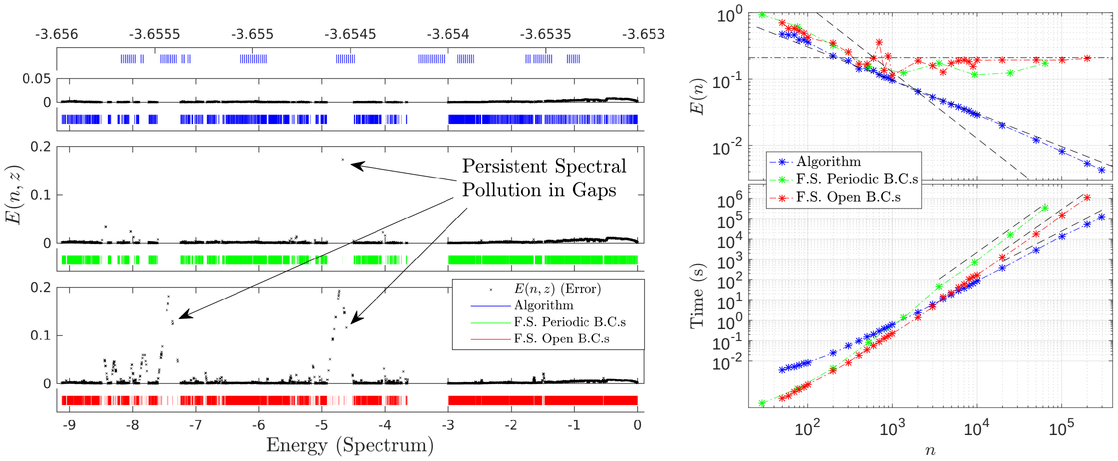


Figure 3.1: Left: Large scale experiment with $n = 10^5$ for the algorithm of this chapter and finite section with open boundary conditions and periodic approximants, applied to the operator H_0 in (3.4.1). The top row shows a magnified section of the approximation provided by the new algorithm and the high resolution obtained. The approximation computed with the finite section methods produces spurious points in band gaps with large errors ~ 0.2. Right: The maximum errors as well as time of outputs for the algorithm of this chapter (blue) and finite section methods (red for open BCs, green for periodic).

the proof of Theorem 3.2.6, is also shown. This error estimate converges uniformly to the true error on compact subsets of $\mathbb{R}$. Finite section methods produce spurious points in the gaps of the spectrum, and the frequency of spectral pollution is lower for the periodic approximants. The hat shape of the error function in the figure also suggests that our error estimate has converged in the gaps of the spectrum.

The time taken for our algorithm and for the finite section methods to reach the final output (shown in Figure 3.1) suggests a speed-up of about 20 times. Moreover, the time for the finite section method appears to grow $\sim \mathcal{O}(n^{2.9})$, $\mathcal{O}(n^{3.0})$ for open and periodic boundary conditions respectively, whereas the time for our algorithm grows $\sim \mathcal{O}(n^{2.1})$. This predicts even larger differences in computation time for larger n, and meant we were able to compute the spectrum for very large n only using the new algorithm.

3.4.2 Superconductors and the non-Hermitian Anderson model

Hatano and Nelson initiated the study of the non-Hermitian Anderson model in the context of vortex pinning in type-II superconductors [HN96]. Their model showed that an imaginary gauge field in a disordered one-dimensional lattice can induce a delocalisation transition. While synthesising such an imaginary vector potential is a challenge in condensed-matter physics, this phenomenon has been investigated in the field of optics [LGDV15]. From a computational point of view, non-Hermitian Hamiltonians pose a serious challenge, as no previous algorithm converges to the pseudospectra of infinite-dimensional non-Hermitian operators nor provides error bounds.[1] The operator on $l^2(\mathbb{Z})$ can be written as

$$(Hx)_n = e^{-\tau}x_{n-1} + e^{\tau}x_{n+1} + V_n x_n,$$

where $\tau > 0$ and V is a random potential.

[1] Computations of spectra of non-normal operators are also well-known to suffer from numerical instability, even in finite dimensions. For finite section computations, we checked answers using extended precision. This was not an issue for our pseudospectra calculations which are stable (pseudospectra also behave continuously under perturbations).

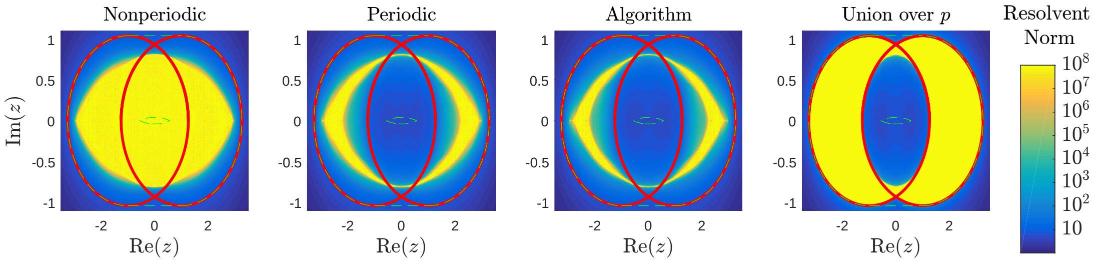


Figure 3.2: Pseudospectra of the finite section method with non-periodic boundary conditions shown as contours of the resolvent norm $\|(H_n - zI)^{-1}\|$ for $n = 10^6$. Similar plots for periodic boundary conditions, the new algorithm with and without varying p. Bounds on the spectrum are shown in green and the set $E + M$ in red.

Spectral computations of H are delicate. Once truncated to a finite lattice of size n, the spectrum and pseudospectrum of the finite section H_n depend on the boundary conditions imposed. Non-periodic boundary conditions (standard finite section) yield an entirely real spectrum, completely ignoring the instability of the model and utterly different from the complex spectrum of H. Hatano and Nelson argued that a more physical model would be periodic boundary conditions. In our case, periodic boundary conditions lead to spectra that converge to a curve in the complex plane strictly contained in the spectrum [GK98].

If $(V_n)_{n\in\mathbb{Z}}$ are i.i.d. random variables, then $\mathrm{Sp}(H)$ and $\mathrm{Sp}_\epsilon(H)$ only depend on the support of the potential, M, almost surely. We consider the Bernoulli case $M = \{\pm 1\}$ where $V_n = 1$ with probability $p \in (0, 1)$. This choice ensures the spectrum has a hole in it by a standard series argument. Defining the ellipse $E = \{e^{\tau+\mathrm{i}\theta} + e^{-\tau-\mathrm{i}\theta} : \theta \in [0, 2\pi)\}$, we also have $E \pm 1 \subset \mathrm{Sp}(H)$ which is contained in the convex hull of $E + [-1, 1]$. Figure 3.2 shows the result of the finite section, i.e. the pseudospectra of H_n for $n = 10^6$ (corresponding to a matrix size of $2n + 1$) and the new algorithm with $\tau = 1/2$ and $p = 1/2$. The spectra of finite sections with non-periodic boundary conditions give the wrong set in the limit $n \to \infty$, filling the hole in the spectrum and converging to the interval $[-3, 3]$ (this can be proven). Pseudospectra for periodic boundary conditions fare much better, as proven for a large class of operators in [Col20b].

We can take advantage of the fact that, ignoring round-off errors, our algorithm has zero error in its output and that the pseudospectrum is invariant under changes in $p \in (0, 1)$. Thus, we have also shown the output over a union of varying p. This gives an excellent estimate of the spectrum and the pseudospectrum.

3.4.3 Open systems in optics

Open systems typically yield non-Hermitian Hamiltonians as there is no guaranteed energy preservation. However, non-Hermitian Hamiltonians can posses real spectra when they respect parity–time (PT) symmetry [BB98, KGM08, Ben07]. A Hamiltonian $H = p^2/2 + V(x)$ is said to be PT-symmetric if it commutes with the action of the operator PT where P is the parity operator $\hat{x} \to -\hat{x}, \hat{p} \to -\hat{p}$ and T the time operator $\hat{p} \to -\hat{p}, i \to -i$. Many PT-symmetric Hamiltonians possess the remarkable property that their spectra are real for small enough $\mathrm{Im}(V)$ but that the spectrum becomes complex above a certain threshold. This phase transition is known as symmetry breaking.

Detecting when symmetry breaking occurs poses a substantial challenge since it is very sensitive to surface/edge states arising from standard truncations. We discuss PT-symmetry breaking for the case of an

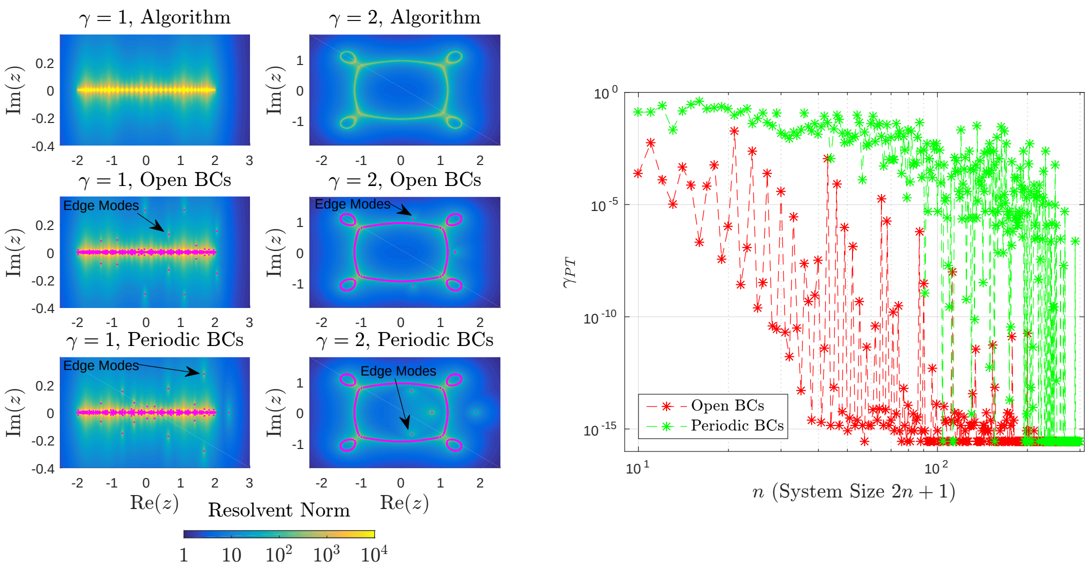


Figure 3.3: Left: Pseudospectra of H computed with the new algorithm and finite sections with different BCs (in magenta). We can easily detect edge modes with the new algorithm, whereas the finite section approach produces incorrect solutions (edge modes). In the periodic case we have no edge, and rather these modes are due to the jump in the potential between the two end sites. Right: Fragile PT-symmetric phase as we increase the system size due to edge states with complex eigenvalues, which verifies the failure of finite sections.

aperiodic potential on a discrete lattice:

$$(Hx)_n = x_{n-1} + x_{n+1} + V_n x_n$$

acting on $l^2(\mathbb{Z})$ where $V_n = \cos(n) + i\gamma\sin(n)$ and $\gamma \geq 0$. Here the aperiodicity occurs due to the incommensurability of the potential and lattice. We stress that the new algorithm can handle any type of potential (such as additional defects modelled by random potentials).

In the limit of increasing system size, the critical parameter γ_{PT} depends on the boundary conditions imposed, often decreasing as the number of sites increases with a fragile PT-symmetric phase. This limit can differ from the value γ_{PT} on the infinite lattice due to surface/edge states [BFKS09]. Using our algorithm gives an estimate for γ_{PT} in the infinite lattice case avoiding this fragility, suggesting that symmetry breaking occurs at $\gamma_{PT} \approx 1 \pm 0.05$. This allows us to detect edge states rigorously (spectral pollution) and the corresponding edge modes. Figure 3.3 shows pseudospectral plots generated by our algorithm for $\gamma = 1, 2$ as well as the plots for finite chains of length 2001 for open and periodic boundary conditions. We can easily use the new algorithm to separate bulk states from edge states. We have also shown the values of γ_{PT} for the finite chains showing the fragility of the PT-symmetric phase.

3.4.4 Partial differential operators

We demonstrate the algorithms of this chapter on PDOs on $L^2(\mathbb{R}^d)$. For the examples with polynomial coefficients in this section, all error bounds and results were *verified rigorously with interval arithmetic*. We also consider non-polynomial coefficients in §3.4.4.

Anharmonic oscillators

First, consider operators of the form

$$H = -\Delta + V(x) = -\Delta + \sum_{j=1}^{d} (a_j x_j + b_j x_j^2) + \sum_{\alpha \in \mathbb{Z}_{\geq 0}^d, |\alpha| \leq M} c(\alpha) x^\alpha,$$

where $a_j, b_j, c(\alpha) \in \mathbb{R}$ and the multi-indices α are chosen such that $\sum_{|\alpha| \leq M} c(\alpha) x^\alpha$ is bounded from below. The Faris–Lavine theorem [RS75, Theorem X.28] shows that such operators are self-adjoint.

We begin with comparisons to some known results in one dimension:

$$\begin{aligned} &V_1(x) = x^2 - 4x^4 + x^6 && E_0 = -2 \\ &V_2(x) = 4x^2 - 6x^4 + x^6 && E_1 = -9 \\ &V_3(x) = (105/64)x^2 - (43/8)x^4 + x^6 - x^8 + x^{10} && E_0 = 3/8 \\ &V_4(x) = (169/64)x^2 - (59/8)x^4 + x^6 - x^8 + x^{10} && E_1 = 9/8. \end{aligned}$$

These examples have discrete spectra and, following the physicists' convention, we list the energy levels as $E_0 \leq E_1 \leq E_2 \leq ....$ We found that the grid resolution of the search routine and the search accuracy for the smallest singular values, not the matrix size, were the main deciding factors in the error bound. Clearly, once we know roughly where the eigenvalues are, we can speed up computations using the fact that the algorithm is local. Furthermore, the search routine's computational time only grows *logarithmically* in its precision. Hence we set the grid spacing and the spacing of the search routine to be $10^5 n$. Table 3.1 shows the results and all values were computed rapidly using a local search grid.

Potential	Exact	$n = 500$	$n = 1000$
V_1	-2	$-2 \pm 2 \times 10^{-8}$	-2 ± 10^{-8}
V_2	-9	$-9 \pm 2 \times 10^{-8}$	-9 ± 10^{-8}
V_3	0.375	$0.375 \pm 1.6192 \times 10^{-4}$	$0.375 \pm 1 \times 10^{-7}$
V_4	1.125	$1.125 \pm 6.013 \times 10^{-4}$	$1.125 \pm 2.4 \times 10^{-7}$

Table 3.1: Test run of algorithm on some potentials with known eigenvalues. Note that we quickly converge to the eigenvalue with error bounds computed by the algorithm and using interval arithmetic.

Next, we consider the operator

$$H_1 = -\Delta + x_1^2 x_2^2,$$

on $L^2(\mathbb{R}^2)$, which is a classic example of a potential that does not blow up at ∞ in every direction, yet still induces an operator with compact resolvent and hence discrete spectrum [Sim83]. Figure 3.4 shows the convergence of the estimate of $\|R(z, H_1)\|^{-1}$ from above as well as finite section estimates. As expected from variational methods, the finite section method produces eigenvalues converging to the true eigenvalues from above (there is no essential spectrum and the operator is positive). Furthermore, the areas where `DistSpec` has converged correspond to areas where finite section has converged. One expects that the time taken for finite section grows somewhere between quadratically and cubically, whereas the new algorithm grows at most $\mathcal{O}(n^{2.75})$ up to logarithmic factors (if one does not take advantage of previous estimates and compact resolvent to reduce the interval length of searches). This is also shown in Figure 3.4, where we found that the finite section method grew roughly cubically whereas our algorithm grew roughly as

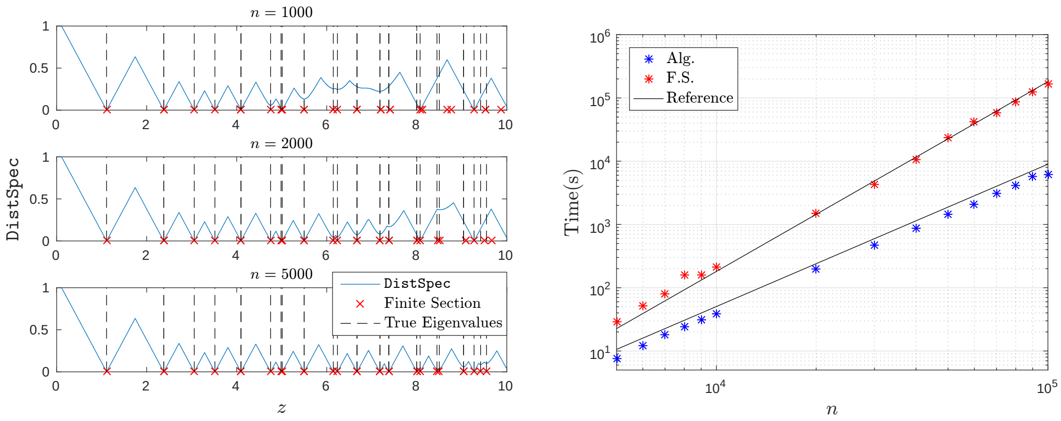


Figure 3.4: Two-dimensional example. Left: The convergence of our algorithm (shown as `DistSpec`) and finite section to the true eigenvalues on the interval $[0, 10]$. Note that points with reliable finite section eigenvalues correspond to points where the estimate of the resolvent norm is well-resolved. Right: Time taken (when not using interval arithmetic) for both methods over a range of n (100 cores) showing near cubic growth for finite section and $\mathcal{O}(n^{2.25})$ growth for our algorithm (reference lines).

$\mathcal{O}(n^{2.25})$ (both shown as reference lines). The speed-up for our algorithm, compared with $\mathcal{O}(n^{2.75})$, was due to the AMD basis ordering used.

Schrödinger operator with constant magnetic field

In this example, we demonstrate that the algorithm of this chapter for computing the spectrum does not suffer from spectral pollution, which is often found in other methods used for self-adjoint operators when there is a gap in the essential spectrum. We will demonstrate this on the Schrödinger operator with constant magnetic field ($B \in \mathbb{R},\ B \neq 0$) in $\mathbb{R}^2$,

$$H_B = \left(-i\partial_{x_1} - \frac{Bx_2}{2}\right)^2 + \left(-i\partial_{x_2} + \frac{Bx_1}{2}\right)^2,$$

which is essentially self-adjoint [RS75] and plays an important role in superconductivity theory [FH10]. It can be shown via unitary transformations that

$$\mathrm{Sp}(H_B) = \{(2k-1)\,|B| : k \in \mathbb{N}\},$$

(see [Hel13]) with each element of the spectrum being an eigenvalue of infinite multiplicity (so that the above agrees with the essential spectrum). Figure 3.5 (left) shows the output of finite section over a range of n and $B = 1$. As expected, there is no spectral pollution below the essential spectrum, but there is heavy spectral pollution in the gaps of the essential spectrum. Figure 3.5 (right) shows the output of our algorithm. This avoids spectral pollution whilst converging to the true spectrum.

This is a simple example since one can analytically diagonalise the operator. However, given an operator, it can be hard to choose an appropriate basis such that finite section avoids spectral pollution (in fact this is, in general, impossible in a precise sense - see §7.1) and the above example demonstrates that we do not have to worry about this when using our algorithm. This will also be revisited for Dirac operators [STY$^+$04] in §4.6.2, where we compute highly oscillatory bounded modes.

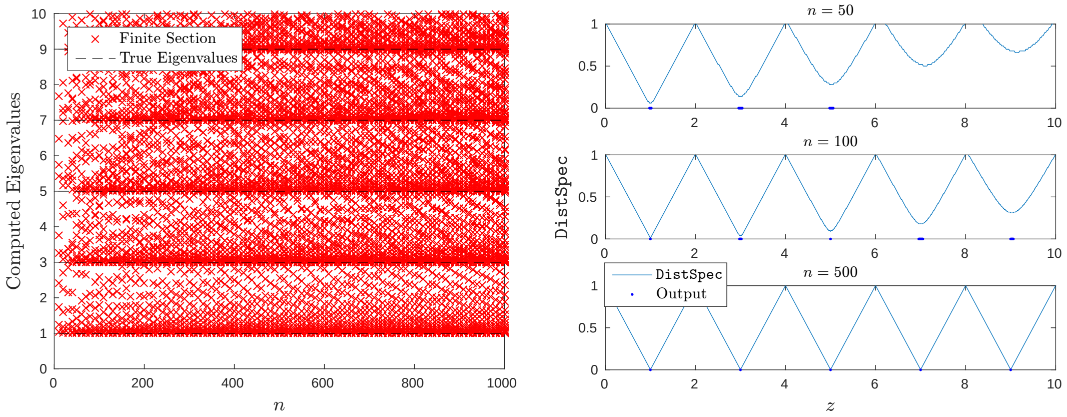


Figure 3.5: Left: Finite section for various n. Note the extremely heavy spectral pollution, although eigenvalues do appear to cluster around the true spectrum. Right: The estimates provided by DistSpec. The estimate converges quickly to the true value from above. The output of our algorithm can be spotted by eye and corresponds to the local minima of the curves below the cut-off 0.5 in this case.

Potential V	E_0	E_1	E_2	E_3	E_4
$\cos(x)$	1.7561051579	3.3447026910	5.0606547136	6.8649969390	8.7353069954
$\tanh(x)$	0.8703478514	2.9666370800	4.9825969775	6.9898951678	8.9931317537
$\exp(-x^2)$	1.6882809272	3.3395578680	5.2703748823	7.2225903394	9.1953373991
$(1+x^2)^{-1}$	1.7468178026	3.4757613534	5.4115076464	7.3503220313	9.3168983920

Table 3.2: Computed eigenvalues for different potentials (first five shown). Each eigenvalue E_n, computed with an error bound at most 10^{-9} via DistSpec, is a shift of the harmonic oscillator eigenvalue $2n+1$

General coefficients: perturbed harmonic oscillator

As a simple set of examples, we consider

$$T = -\Delta + x^2 + V(x),$$

on $L^2(\mathbb{R})$, where V is a bounded potential (for more examples with general coefficients, see [CHns]). Such operators have discrete spectra, however, the perturbation V causes the eigenvalues to shift relative to the classical harmonic oscillator (whose spectrum is the set of odd positive integers). Table 3.2 shows the first five eigenvalues for a range of potentials, computed with an error bound at most 10^{-9}.

Pseudospectra and PT-symmetry

We now turn to the pseudospectrum and consider PT-symmetric non-self-adjoint operators T. The first example is the imaginary cubic oscillator defined formally (in one dimension) by

$$H_2 = -d^2/dx^2 + ix^3.$$

This operator is the most studied example of a PT-symmetric operator (a concept met previously in §3.4.3) [BB98, BBJ02], as well as appearing in statistical physics and quantum field theory [Fis78]. It is known that the resolvent is compact [CGM80] with all eigenvalues simple and residing in $\mathbb{R}_{\geq 0}$ [DDT01, Tai06].

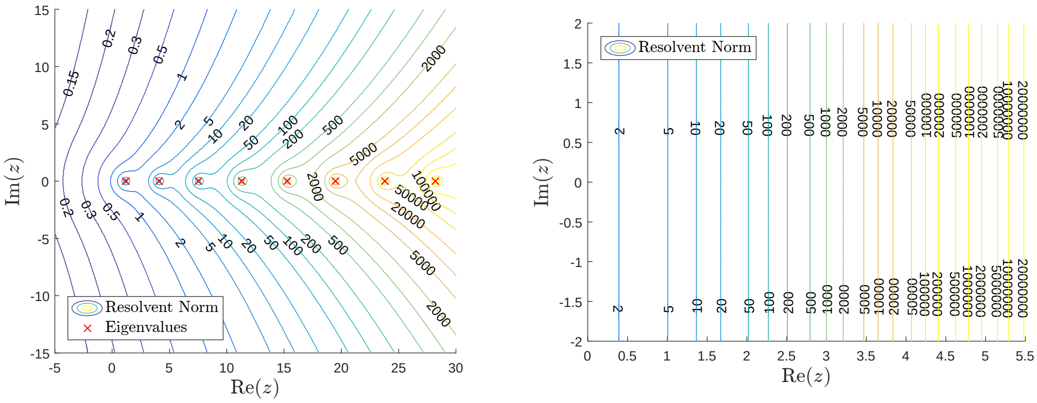


Figure 3.6: Left: Calculated pseudospectrum for the imaginary cubic oscillator. Note the clear presence of eigenvalues. Right: Calculated pseudospectrum for imaginary Airy operator. Both figures were produced with $n = 1000$.

The eigenvectors are complete but do not form a Riesz basis [SK12]. Figure 3.6 shows the pseudospectrum computed using $n = 1000$. This demonstrates the instability of the spectrum of the operator.

Next, we consider the imaginary Airy operator

$$H_3 = -d^2/dx^2 + ix,$$

since this is known to have empty spectrum [Hel13], demonstrating that the algorithm is effective in this case. Note that any finite section method will overestimate the pseudospectrum due to the presence of false eigenvalues. H_3 is PT-symmetric and has compact resolvent. The resolvent norm $\|R(z, H_3)\|$ only depends on the real part of z and blows up exponentially as $\text{Re}(z) \to +\infty$. We have shown the computed pseudospectrum for $n = 1000$ in Figure 3.6.

Chapter 4

Computing Spectral Measures

Any normal operator A has an associated projection-valued measure, E^A, whose existence is guaranteed by the spectral theorem and whose support is $\mathrm{Sp}(A)$ [KR97a, KR97b, RS80]. This allows the representation of the operator A as an integral over $\mathrm{Sp}(A)$, analogous to the finite-dimensional case of diagonalisation:

$$Ax = \int_{\mathrm{Sp}(A)} \lambda dE^A(\lambda)x, \quad \forall x \in \mathcal{D}(A),$$

where $\mathcal{D}(A)$ denotes the domain of A. For example, if A is compact, then E^A corresponds to projections onto eigenspaces, familiar from the finite-dimensional setting. However, in general, the situation is more complicated with different types of spectra. The computation of E^A, along with its various decompositions and their supports, is of great applicative and theoretical interest. For example, spectral measures are related to the autocorrelation function in signal processing, resonance phenomena in scattering theory, and stability analysis for fluids and many other quantities [KM71, GS03, Ros91, ELOB07, ELO94, ELS19, BP84, HHK72, LSY16, WC15, KS03, DN86, DS06a, TOD12]. Moreover, the computation of E^A allows computation of additional objects, such as the functional calculus and the Radon–Nikodym derivative of the absolutely continuous component.

In this chapter, based on [Col21, CHT21], we provide algorithms for the computation of spectral measures for a large class of self-adjoint operators. We classify the computation of measures, measure decompositions, functional calculus and Radon–Nikodym derivatives in the SCI hierarchy. The central ingredient is the computation of the resolvent operator with error control. We also discuss how to improve the convergence rates by using rational convolution kernels. For a given desired accuracy, one may evaluate the resolvent at a much larger distance from the spectrum than in the case of a first-order method. The examples highlight that the new algorithms can easily be used in tandem with any numerical procedure that computes the action of the resolvent with asymptotic error control. This gives great flexibility to the methods. The reader is encouraged to explore the software package `SpecSolve`: https://github.com/SpecSolve/SpecSolve, which supports general ODEs, PDEs, integral operators and lattice operators. Further examples of the use of these algorithms can be found in [JCN$^+$21, CHTW21].

4.1 Background and Summary

We consider the canonical separable Hilbert space $\mathcal{H} = l^2(\mathbb{N})$, the set of square summable sequences with canonical basis $\{e_n\}_{n\in\mathbb{N}}$. By a choice of basis our results extend to any separable Hilbert space. For example, we can handle partial differential operators through spectral methods. The algorithms can be made to work with any method that computes the resolvent with an asymptotic form of error control - a matrix representation is not needed. Let $\mathcal{C}(l^2(\mathbb{N}))$ be the set of closed densely defined linear operators A such that $\mathrm{span}\{e_n : n \in \mathbb{N}\}$ forms a core of A and A^*. The point spectrum (the set of eigenvalues) will be denoted by $\mathrm{Sp}_{\mathrm{p}}(A)$. We will focus on the subclass $\Omega_{\mathrm{N}} \subset \mathcal{C}(l^2(\mathbb{N}))$ of normal operators, those for which $\mathcal{D}(A) = \mathcal{D}(A^*)$ and $\|Ax\| = \|A^*x\|$ for all $x \in \mathcal{D}(A)$. The subclass $\subset \Omega_{\mathrm{N}}$ of self-adjoint (again allowing unbounded operators) operators will be denoted by Ω_{SA}. Recall that for $A \in \Omega_{\mathrm{SA}}$, $\mathrm{Sp}(A) \subset \mathbb{R}$.

Given $A \in \Omega_{\mathrm{N}}$ and a Borel set B, E_B^A will denote the projection $E^A(B)$. Given $x, y \in l^2(\mathbb{N})$, we can define a bounded (complex-valued) measure $\mu_{x,y}^A$ via the formula

$$\mu_{x,y}^A(B) = \langle E_B^A x, y\rangle.$$

Via the Lebesgue decomposition theorem [Hal50], $\mu_{x,y}^A$ can be decomposed into three parts

$$\mu_{x,y}^A = \mu_{x,y,\mathrm{ac}}^A + \mu_{x,y,\mathrm{sc}}^A + \mu_{x,y,\mathrm{pp}}^A,$$

the absolutely continuous part of the measure (with respect to the Lebesgue measure), the singular continuous part (singular with respect to the Lebesgue measure and atomless) and the pure point part. When considering Ω_{SA}, we will consider Lebesgue measure on $\mathbb{R}$ and let

$$\rho_{x,y}^A(\lambda) = \frac{d\mu_{x,y,\mathrm{ac}}^A}{dm}(\lambda), \tag{4.1.1}$$

the Radon–Nikodym derivative of $\mu_{x,y,\mathrm{ac}}^A$ with respect to Lebesgue measure. Of course this can be extended to the unitary (and, more generally, normal) case. This naturally gives a decomposition of the Hilbert space $\mathcal{H} = l^2(\mathbb{N})$. For $\mathcal{I} = \mathrm{ac}, \mathrm{sc}$ and pp, we let $\mathcal{H}_{\mathcal{I}}$ consist of vectors x whose measure $\mu_{x,x}^A$ is absolutely continuous, singular continuous and pure point respectively. This gives rise to the orthogonal decomposition

$$\mathcal{H} = \mathcal{H}_{\mathrm{ac}} \oplus \mathcal{H}_{\mathrm{sc}} \oplus \mathcal{H}_{\mathrm{pp}} \tag{4.1.2}$$

whose associated projections will be denoted by P_{ac}^A, P_{sc}^A and P_{pp}^A respectively. These projections commute with A and the projections obtained through the projection-valued measure. Of particular interest is the spectrum of A restricted to each $\mathcal{H}_{\mathcal{I}}$, which will be denoted by $\mathrm{Sp}_{\mathcal{I}}(A)$. These different sets and subspaces often, but not always, characterise different physical properties in quantum mechanics (such as the famous RAGE theorem [Rue69, AG74, Ens78]), where a system is modelled by some Hamiltonian $A \in \Omega_{\mathrm{SA}}$ [CFKS87, Com93, GKP91, Las96]. For example, pure point spectrum implies the absence of ballistic motion for many Schrödinger operators [Sim90].

4.1.1 Algorithmic set-up

Given an operator $A \in \mathcal{C}(l^2(\mathbb{N}))$, we can view it as an infinite matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots \\ a_{21} & a_{22} & a_{23} & \cdots \\ a_{31} & a_{32} & a_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

through the inner products $a_{ij} = \langle Ae_j, e_i\rangle$. To be precise about the information needed to compute spectral properties, we define the class of evaluation functions $\Lambda_1 = \{\langle Ae_j, e_i\rangle : i, j \in \mathbb{N}\}$. For discrete operators, this information is often given to us, for example, in tight-binding models in physics, and hence it is natural to seek to compute spectral properties from matrix values. For partial differential operators, such information is often given through inner products with a suitable basis, and, in this case, the inexact input model is needed due to approximating the integrals.

We will be concerned with operators whose matrix representation has a known asymptotic rate of column/off-diagonal decay. Namely, let $f:\mathbb{N}\rightarrow\mathbb{N}$ with $f(n)>n$ and let $\alpha=\{\alpha_n\}_{n\in\mathbb{N}},\beta=\{\beta_n\}_{n\in\mathbb{N}}$ be null sequences[1] of non-negative real numbers. We then define

$$\begin{aligned}\Omega_{f,\alpha,\beta}=\{A\in\Omega_{\mathrm{SA}}:\|(P_{f(n)}-I)AP_n\|=\mathcal{O}(\alpha_n),\text{ as } n\rightarrow\infty\}\\ \times\{x\in l^2(\mathbb{N}):\|P_nx-x\|=\mathcal{O}(\beta_n),\text{ as } n\rightarrow\infty\},\end{aligned}\tag{4.1.3}$$

where P_n denotes the orthogonal projection onto $\mathrm{span}\{e_1,...,e_n\}$. We will also use

$$\Omega_{f,\alpha}=\{A\in\Omega_{\mathrm{SA}}:\|(P_{f(n)}-I)AP_n\|=\mathcal{O}(\alpha_n),\text{ as } n\rightarrow\infty\}.$$

The collection of vectors in $l^2(\mathbb{N})$ satisfying $\|P_nx-x\|=\mathcal{O}(\beta_n)$ will be denoted by V_β. Finally, when $\alpha_n\equiv 0$, we will abuse notation slightly in requiring the stronger condition

$$\|(P_{f(n)}-I)AP_n\|=0.$$

Thus $\Omega_{f,0}$ is the class of self-adjoint operators whose matrix sparsity structure is captured by the function f. For example, if $f(n)=n+1$ we recover the class of self-adjoint tridiagonal matrices, the most studied class of operators. When discussing classes that include vectors $x\in l^2(\mathbb{N})$, we extend Λ_1 to include pointwise evaluations of the coefficients of x.

4.1.2 A motivating example

Consider a Jacobi operator with matrix

$$J=\begin{pmatrix}b_1 & a_1 & & \\ a_1 & b_2 & a_2 & \\ & a_2 & b_3 & \ddots \\ & & \ddots & \ddots\end{pmatrix}$$

where $a_j,b_j\in\mathbb{R}$ and $a_j>0$. An enormous amount of work exists on the study of these operators, and the correspondence between bounded Jacobi matrices and probability measures with compact support [Tes00, Dei99]. The entries in the matrix provide the coefficients in the recurrence relation for the corresponding orthonormal polynomials. To study the canonical measure μ_J, one usually considers the principal resolvent function defined on $\mathbb{C}\backslash\mathrm{Sp}(J)$ via

$$G(z):=\langle R(z,J)e_1,e_1\rangle=\int_{\mathbb{R}}\frac{d\mu_J(\lambda)}{\lambda-z},$$

and then takes z close to the real axis. The function G is also known in the differential equations and Schrödinger communities as the Weyl m-function [Tes00, GS97a] and one can develop the discrete analogue of what is known as Weyl–Titchmarsh–Kodaira theory for Sturm–Liouville operators. Going back

[1] We use the term 'null sequence' for a sequence converging to zero.

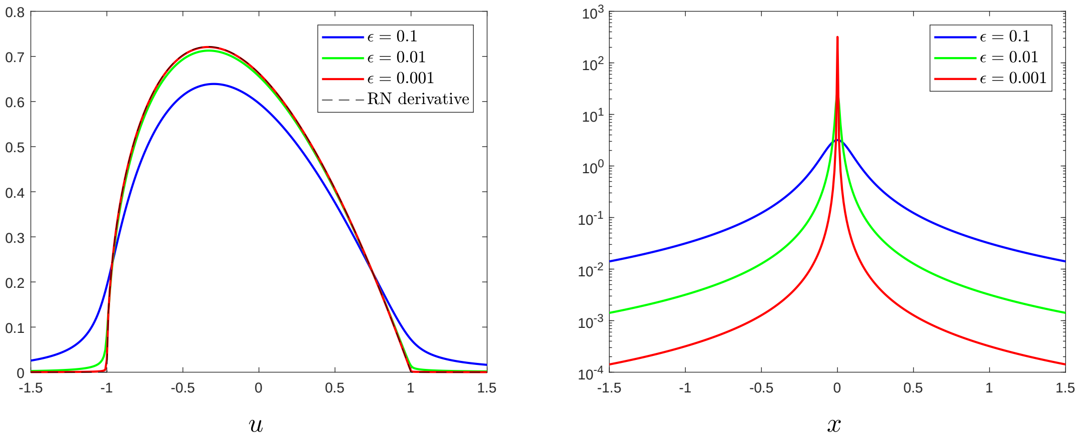


Figure 4.1: Smoothed approximations of the Radon–Nikodym derivative for the Jacobi operator associated to Jacobi polynomials with $\alpha = 1$, $\beta = 1/2$. Here the measure is absolutely continuous and supported on $[-1, 1]$. Left: Convolutions $K_H(u + i\epsilon; J, e_1)$ for different ϵ using the methods of this chapter. Right: The associated Poisson kernel $\pi^{-1}\epsilon/(\epsilon^2 + x^2)$ which approaches a Dirac delta distribution as $\epsilon \downarrow 0$.

to the work of Stieltjes [Sti94] (see also [Akh65, Wal48]), there is a representation of G as a continued fraction:

$$G(z) := \frac{1}{-z + b_1 - \frac{a_1^2}{-z+b_2-\dots}}. \tag{4.1.4}$$

One can also approximate G via finite truncated matrices [Tes00].

However, there are two obstacles to overcome when using (4.1.4) and its variants as a means to compute measures. First, this representation of the principal resolvent function is structurally dependent. For example, (4.1.4) is valid for the restricted case of Jacobi operators and hence one is led to seek different methods for different operators (such as tight-binding Hamiltonians on two-dimensional lattices, which have a growing bandwidth when represented as an infinite matrix). Second, this would seem to give the wrong classification of the difficulty of the problem in the SCI hierarchy, giving rise to a tower of algorithms with two limits. One first takes a truncation parameter n to infinity to compute $G(z)$ for $\mathrm{Im}(z) > 0$, and then a second limit as z approaches the real axis. One of the main messages of this chapter is that both of these issues can be overcome. Measures can be computed in one limit via an algorithm Γ_n and for a large class of operators. The only restriction is a known asymptotic decay rate of the off-diagonal entries.

Consider the Poisson kernel for the half-plane defined respectively by

$$P_H(x, y) = \frac{1}{\pi}\frac{y}{x^2 + y^2},$$

where (x, y) denote the usual Cartesian coordinates. Let A be a normal operator, then for $z \notin \mathrm{Sp}(A)$, we have from the functional calculus that

$$R(z, A) = \int_{\mathrm{Sp}(A)} \frac{1}{\lambda - z} dE^A(\lambda).$$

For self-adjoint A, $z = u + iv \in \mathbb{C}\backslash\mathbb{R}$ ($u, v \in \mathbb{R}$) and $x \in l^2(\mathbb{N})$ we define

$$\begin{aligned} K_H(z; A, x) :&= \frac{1}{2\pi i}[R(z, A) - R(\overline{z}, A)]x \\ &= \frac{1}{2\pi i}\int_{-\infty}^{\infty}\left[\frac{1}{\lambda - z} - \frac{1}{\lambda - \overline{z}}\right] dE^A(\lambda)x = \int_{-\infty}^{\infty} P_H(u - \lambda, v) dE^A(\lambda)x. \end{aligned}$$

We see that the computation of the resolvent with error control allows the computation of $G(z)$ with error control through taking inner products. By considering $G(z) - G(\overline{z})$, this allows the computation of the convolution of the measure μ_J with the Poisson kernel P_H. In other words, we can compute a smoothed version of the measure μ_J with error control. Figure 4.1 demonstrates this for a typical example. We will see also in §4.5, that kernels different to the Poisson kernel allow improved rates of convergence.

4.2 Approximating the Resolvent

The algorithms built in this chapter rely on the ability to compute the action of the resolvent operator $R(z, A) = (A - z)^{-1}$ for $z \notin \mathrm{Sp}(A)$ with error control.

Proposition 4.2.1. *Let $A \in \Omega_{\mathrm{N}}$, $z \in \mathbb{C}\backslash\mathrm{Sp}(A)$ and $x \in l^2(\mathbb{N})$. Suppose that the following hold for constants C_1 and C_2 (that may depend on A and x and may be unknown), together with null sequences $\{\alpha_n\}_{n\in\mathbb{N}}$ and $\{\beta_n\}_{n\in\mathbb{N}}$ independent of A and x:*

1. *For $f : \mathbb{N} \to \mathbb{N}$ with $f(n) > n$, $\|(I - P_{f(n)})AP_n\| \leq C_1\alpha_n$,*
2. *$\|P_n x - x\| \leq C_2\beta_n$,*
3. *For $\delta > 0$, $\mathrm{dist}(z, \mathrm{Sp}(A)) \geq \delta$.*

Then there exists a sequence of arithmetic algorithms $\Gamma_n(A, x, z)$ mapping into $l^2(\mathbb{N})$, each of which use the evaluation functions in Λ_1, such that each vector $\Gamma_n(A, x, z)$ has finite support with respect to the canonical basis for each n and $\Gamma_n(A, x, z) \to R(z, A)x$. Moreover, the following error bound holds

$$\|\Gamma_n(A, x, z) - R(z, A)x\| \leq \frac{C_2\beta_{f(n)} + C_1\alpha_n\|\Gamma_n(A, x, z)\| + \|P_{f(n)}(A - zI)\Gamma_n(A, x, z) - P_{f(n)}x\|}{\delta}. \tag{4.2.1}$$

If a bound on C_1 and C_2 are known, this error bound can be computed to arbitrary accuracy using finitely many arithmetic operations and comparisons. In the more general case for a fixed $\{\alpha_n\}$, $\{\beta_n\}$ and f, this gives an asymptotic error bound holding for all A, x and z which satisfy the above assumptions.

Proof. We have that $n = \mathrm{rank}(P_n) = \mathrm{rank}((A - zI)P_n) = \mathrm{rank}(P_{f(n)}(A - zI)P_n)$ for large n since $\sigma_{\mathrm{inf}}(A - zI) > 0$ and $\|(I - P_{f(n)})(A - zI)P_n\| \leq C_1\alpha_n \to 0$. Hence we can define

$$\widetilde{\Gamma}_n(A, x, z) := \begin{cases} 0, & \text{if } \sigma_{\mathrm{inf}}(P_n(A^* - \overline{z}I)P_{f(n)}(A - zI)|_{P_n(l^2(\mathbb{N}))}) \leq \frac{1}{n} \\ [P_n(A^* - \overline{z}I)P_{f(n)}(A - zI)P_n]^{-1}P_n(A^* - \overline{\mathrm{z}}I)P_{f(n)}x, & \text{otherwise.} \end{cases}$$

Suppose that n is large enough so that $\sigma_{\mathrm{inf}}(P_n(A^* - \overline{z}I)P_{f(n)}(A - zI)|_{P_n(l^2(\mathbb{N}))}) > 1/n$. Then $\widetilde{\Gamma}_n(A, x, z)$ is a (least-squares) solution of the optimisation problem $\mathrm{argmin}_y\|P_{f(n)}(A - zI)P_n y - x\|$. The linear space $\mathrm{span}\{e_n : n \in \mathbb{N}\}$ forms a core of A and hence of $A - zI$. It follows by invertibility of $A - zI$ that given any $\epsilon > 0$, there exists an $m = m(\epsilon)$ and a $y = y(\epsilon)$ with $P_m y = y$ such that

$$\|(A - zI)y - x\| \leq \epsilon.$$

It follows that for all $n \geq m$,

$$\begin{aligned}\|(A-zI)\widetilde{\Gamma}_n(A,x,z)-x\| &\leq \|P_{f(n)}(A-zI)\widetilde{\Gamma}_n(A,x,z)-x\|+C_1\alpha_n\|\widetilde{\Gamma}_n(A,x,z)\|\\ &\leq \|P_{f(n)}(A-zI)y-x\|+C_1\alpha_n\|\widetilde{\Gamma}_n(A,x,z)\|\\ &\leq \|P_{f(n)}(A-zI)y-P_{f(n)}x\|+C_2\beta_{f(n)}+C_1\alpha_n\|\widetilde{\Gamma}_n(A,x,z)\|\\ &\leq \epsilon+C_2\beta_{f(n)}+C_1\alpha_n\|\widetilde{\Gamma}_n(A,x,z)\|.\end{aligned}$$

This implies that

$$\begin{aligned}\|\widetilde{\Gamma}_n(A,x,z)-R(z,A)x\| &\leq \|R(z,A)\|\|(A-zI)\widetilde{\Gamma}_n(A,x,z)-x\|\\ &\leq \|R(z,A)\|\left(\epsilon+C_2\beta_{f(n)}+C_1\alpha_n\|\widetilde{\Gamma}_n(A,x,z)\|\right).\end{aligned}$$

In particular, since α_n and β_n are null, this implies that $\|\widetilde{\Gamma}_n(A,x,z)\|$ is uniformly bounded in n. Since $\epsilon>0$ was arbitrary, we also see that $\widetilde{\Gamma}_n(A,x,z)$ converges to $R(z,A)x$.

Define the matrices

$$B_n=P_n(A^*-\overline{z}I)P_{f(n)}(A-zI)P_n,\quad C_n=P_n(A^*-\mathrm{z}I)P_{f(n)}.$$

Given the evaluation functions in Λ_1, we can compute the entries of these matrices to any given accuracy and hence also to arbitrary accuracy in the operator norm (say using the Frobenius norm to bound the operator norm), using finitely many arithmetic operations and comparisons. Denote the approximations of B_n and C_n by $\widetilde{B}_n$ and $\widetilde{C}_n$ respectively and assume that

$$\|B_n-\widetilde{B}_n\|\leq u_n,\quad \|C_n-\widetilde{C}_n\|\leq v_n,$$

for null sequences $\{u_n\},\{v_n\}$. Note that $\widetilde{B}_n^{-1}$ can be computed using finitely many arithmetic operations and comparisons. So long as u_n is small enough, the resolvent identity implies that

$$\|B_n^{-1}-\widetilde{B}_n^{-1}\|\leq\frac{\|\widetilde{B}_n^{-1}\|^2u_n}{1-u_n\|\widetilde{B}_n^{-1}\|}=:w_n.$$

By taking u_n and v_n smaller if necessary (so that the algorithm is adaptive and it is straightforward to bound the norm of a finite matrix from above), we can ensure that $\|\widetilde{B}_n^{-1}\|v_n\leq n^{-1}$ and $(\|\widetilde{C}_n\|+v_n)w_n\leq n^{-1}$. We can compute $\sigma_{\inf}(P_n(A^*-\overline{z}I)P_{f(n)}(A-zI)|_{P_n(l^2(\mathbb{N}))})$ to arbitrary accuracy using finitely many arithmetic operations and comparisons. Suppose this is done to an accuracy $1/n^2$ and denote the approximation via τ_n. We then define

$$\Gamma_n(A,x,z):=\begin{cases}0, & \text{if } \tau_n\leq\frac{1}{n}\\ \widetilde{B}_n^{-1}\widetilde{C}_n\widetilde{x}_n, & \text{otherwise,}\end{cases}$$

where $\widetilde{x}_n=P_{f(n)}x$. It follows that $\Gamma_n(A,x,z)$ can be computed using finitely many arithmetic operations and, for large n,

$$\|\Gamma_n(A,x,z)-\widetilde{\Gamma}_n(A,x,z)\|\leq\left(\|\widetilde{B}_n^{-1}\|v_n+(\|\widetilde{C}_n\|+v_n)w_n\right)\|x\|\rightarrow 0,$$

so that $\Gamma_n(A,x,z)$ converges to $R(z,A)x$.

Furthermore, the following error bound holds (which also holds if $\tau_n\leq 1/n$)

$$\begin{aligned}\|\Gamma_n(A,x,z)-R(z,A)x\| &\leq \|R(z,A)\|\|(A-zI)\Gamma_n(A,x,z)-x\|\\ &\leq\frac{C_2\beta_{f(n)}+C_1\alpha_n\|\Gamma_n(A,x,z)\|+\|P_{f(n)}(A-zI)\Gamma_n(A,x,z)-P_{f(n)}x\|}{\mathrm{dist}(z,\mathrm{Sp}(A))},\end{aligned}$$

since A is normal so that $\|R(z,A)\| = \mathrm{dist}(z,\mathrm{Sp}(A))^{-1}$. This bound converges to 0 as $n \to \infty$. If the C_1 and C_2 are known it can be approximated to arbitrary accuracy using finitely many arithmetic operations and comparisons. □

Note that if A is banded with bandwidth m, then we can take $f(n) = n+m$ and the above computation can be done in $\mathcal{O}(nm^2)$ operations [GVL13].

Corollary 4.2.2. *There exists a sequence of arithmetic algorithms*

$$\Gamma_n : \Omega_{f,\alpha,\beta} \times \mathbb{C}\backslash\mathbb{R} \to l^2(\mathbb{N})$$

with the following properties:

1. *For all $(A,x) \in \Omega_{f,\alpha,\beta}$ and $z \in \mathbb{C}\backslash\mathbb{R}$, $\Gamma_n(A,x,z)$ converges to $R(z,A)x$ in $l^2(\mathbb{N})$ as $n \to \infty$.*

2. *For any $(A,x) \in \Omega_{f,\alpha,\beta}$, there exists a constant $C(A,x)$ such that for all $z \in \mathbb{C}\backslash\mathbb{R}$,*

$$\|\Gamma_n(A,x,z) - R(z,A)x\| \leq \frac{C(A,x)}{|\mathrm{Im}(z)|}\left[\alpha_n + \beta_n\right].$$

Exercise: Prove Corollary 4.2.2 using Proposition 4.2.1.

Finally, we will need Stone's famous formula.

Proposition 4.2.3 (Stone's formula [Sto90])**.** *Recalling the definition of K_H in §4.1.2. Let $A \in \Omega_{\mathrm{SA}}$. Then for any $-\infty \leq a < b \leq \infty$ and $x \in l^2(\mathbb{N})$,*

$$\lim_{\epsilon \downarrow 0} \int_a^b K_H(u+i\epsilon; A, x)du = E^A_{(a,b)}x + \frac{1}{2}E^A_{\{a,b\}}x.$$

Exercise: Prove Stone's formula using the dominated convergence theorem.

4.3 Computation of Measures

We start by considering the computation of $E^A_U x$ where $U \subset \mathbb{R}$ is a non-trivial open set. The collection of these subsets will be denoted by $\mathcal{U}$. To be precise, we assume that we have access to a finite or countable collection $a_m(U), b_m(U) \in \mathbb{R} \cup \{\pm\infty\}$ such that U can be written as a disjoint union

$$U = \bigcup_m (a_m(U), b_m(U)). \tag{4.3.1}$$

With an abuse of notation, we add this information as evaluation functions to Λ_1 to form $\widetilde{\Lambda}_1$.

Theorem 4.3.1 (Computation of measures on open sets)**.** *Given the above set-up, consider the map*

$$\begin{aligned} \Xi_{\mathrm{meas}} : \Omega_{f,\alpha,\beta} \times \mathcal{U} &\to l^2(\mathbb{N}) \\ (A,x,U) &\to E^A_U x. \end{aligned}$$

Then $\{\Xi_{\mathrm{meas}}, \Omega_{f,\alpha,\beta} \times \mathcal{U}, \widetilde{\Lambda}_1\} \in \Delta^A_2$. In other words, we can construct a convergent sequence of arithmetic algorithms for the problem.

Proof. Let $A \in \Omega_{\mathrm{SA}}$ and $z_1, z_2 \in \mathbb{C}\backslash\mathbb{R}$. By the resolvent identity and self-adjointness of A,

$$\|R(z_1, A) - R(z_1, A)\| \leq |\mathrm{Im}(z_1)|^{-1} |\mathrm{Im}(z_2)|^{-1} |z_1 - z_2| .$$

Hence, for $z = u+i\epsilon$ with $\epsilon > 0$, the vector-valued function $K_H(u+i\epsilon; A, x)$ (considered with argument u) is Lipschitz continuous with Lipschitz constant bounded by $\epsilon^{-2}\|x\|/\pi$. Now consider the class $\Omega_{f,\alpha,\beta} \times \mathcal{U}$ and let $(A, x, U) \in \Omega_{f,\alpha,\beta} \times \mathcal{U}$. From Corollary 4.2.2, we can construct a sequence of arithmetic algorithms, $\widehat{\Gamma}_n$, such that

$$\|\widehat{\Gamma}_n(A, u, z) - K_H(u + i\epsilon; A, x)\| \leq \frac{C(A, x)}{\epsilon} (\alpha_n + \beta_n)$$

for all $(A, x) \in \Omega_{f,\alpha,\beta}$. It follows from standard quadrature rules and taking subsequences if necessary (using that $\{\alpha_n\}$ and $\{\beta_n\}$ are null), that for $-\infty < a < b < \infty$, the integral

$$\int_a^b K_H \left(u + \frac{i}{n}; A, x\right) du$$

can be approximated to an accuracy $\widehat{C}(A, x)/n$ using finitely many arithmetic operations and comparisons and the relevant set of evaluation functions $\widetilde{\Lambda}_1$ (the constant C now becomes $\widehat{C}$ due to not knowing the exact value of $\|x\|$).

Recall that we assumed the disjoint union

$$U = \bigcup_m (a_m, b_m)$$

where $a_m, b_m \in \mathbb{R} \cup \{\pm\infty\}$ and the union is at most countable. Without loss of generality, we assume that the union is over $m \in \mathbb{N}$. We then let $a_{m,n}, b_{m,n} \in \mathbb{Q}$ be such that $a_{m,n} \downarrow a_m$ and $b_{m,n} \uparrow b_m$ as $n \to \infty$ with $a_{m,n} < b_{m,n}$ and hence $(a_{m,n}, b_{m,n}) \subset (a_m, b_m)$. Let

$$U_n = \bigcup_{m=1}^{n} (a_{m,n}, b_{m,n}),$$

then the proof of Stone's formula in Proposition 4.2.3 (essentially an application of the dominated convergence theorem) can be easily adapted to show that

$$\lim_{n\to\infty} \int_{U_n} K_H \left(u + \frac{i}{n}; A, x\right) du = E_U^A x.$$

Note that we do not have to worry about contributions from endpoints of the intervals (a_m, b_m) since we approximate strictly from within. To finish the proof, we simply let $\Gamma_n(A, x, U)$ be an approximation of the integral

$$\int_{U_n} K_H \left(u + \frac{i}{n}; A, x\right) du$$

to within accuracy $\widehat{C}(A, x)/n$ (which by the above remarks can be computed using finitely many arithmetic operations and comparisons and the relevant set of evaluation functions $\widetilde{\Lambda}_1$). □

Recall from §4.1 that $P_{\mathcal{I}}^A$ denotes the orthogonal projection onto the space $\mathcal{H}_{\mathcal{I}}^A$, where $\mathcal{I}$ denotes a generic type (ac, sc, pp, c or s). We have included the continuous and singular parts denoted by c or s which correspond to $\mathcal{H}_{\mathrm{ac}} \oplus \mathcal{H}_{\mathrm{sc}}$ and $\mathcal{H}_{\mathrm{sc}} \oplus \mathcal{H}_{\mathrm{pp}}$ respectively. These are often encountered in mathematical physics.

Theorem 4.3.2. *Given the above set-up, consider the map*

$$\begin{aligned} \Xi_{\mathcal{I}} : \Omega_{f,\alpha,\beta} \times V_\beta \times \mathcal{U} &\to \mathbb{C} \\ (A, x, y, U) &\to \langle P_{\mathcal{I}}^A E_U^A x, y\rangle = \mu_{x,y,\mathcal{I}}^A(U), \end{aligned}$$

for $\mathcal{I} = \mathrm{ac}, \mathrm{sc}, \mathrm{pp}, \mathrm{c}$ *or* s*. Then*

$$\Delta_2^G \not\ni \{\Xi_{\mathcal{I}}, \Omega_{f,\alpha,\beta} \times V_\beta \times \mathcal{U}, \widetilde{\Lambda}_1\} \in \Delta_3^A.$$

To prove this theorem, it is enough, by the polarisation identity, to consider $x = y$ (note that all the projections commute). We will split the proof into two parts - the Δ_3^A inclusion and the Δ_2^G exclusion.

Proof of inclusion in Theorem 4.3.2

Since $P_{\mathrm{pp}}^A = I - P_{\mathrm{c}}^A$, $P_{\mathrm{ac}}^A = I - P_{\mathrm{s}}^A$ and $P_{\mathrm{sc}}^A = P_{\mathrm{s}}^A - P_{\mathrm{pp}}^A$, it is enough to consider only $\mathcal{I} = \mathrm{c}$ and $\mathcal{I} = \mathrm{s}$.

Step 1: We first deal with $\mathcal{I} = \mathrm{c}$, where we shall use a similar argument to the proof of Theorem 4.4.1 (which is more general than what we need). We recall the RAGE theorem [Rue69, AG74, Ens78] as follows. Let Q_n denote the orthogonal projection onto vectors in $l^2(\mathbb{N})$ with support outside the subset $\{1, ..., n\} \subset \mathbb{N}$. Then for any $x \in l^2(\mathbb{N})$,

$$\begin{aligned}\langle P_{\mathrm{c}}^A E_U^A x, x\rangle = \|P_{\mathrm{c}}^A E_U^A x\|^2 &= \lim_{n\to\infty}\lim_{t\to\infty}\frac{1}{t}\int_0^t \left\|Q_n e^{-iAs} E_U^A x\right\|^2 ds \\ &= \lim_{n\to\infty}\lim_{t\to\infty}\frac{1}{t}\int_0^t \left\|Q_n e^{-iAs}\chi_U(A)x\right\|^2 ds.\end{aligned}$$

The proof of Theorem 4.4.1 is easily adapted to show that there exists arithmetic algorithms $\widetilde{\Gamma}_{n,m}$ using $\widetilde{\Lambda}_1$ such that

$$\|Q_n e^{-iAs}\chi_U(A)x - \widetilde{\Gamma}_{n,m}(A, x, U, s)\| \leq \frac{C(A, x, U)}{m}$$

for all $(A, x, U, s) \in \Omega_{f,\alpha,\beta} \times \mathcal{U} \times \mathbb{R}$. Note that this bound can be made independent of s (as we have written above) by sufficiently approximating the function $\exp(-its)\chi_U(t)$ (it has known total variation for a given s and uniform bound). We now define

$$\Gamma_{n,m}(A, x, U) = \frac{1}{m^2}\sum_{j=1}^{m^2} \|\widetilde{\Gamma}_{m,n}(A, x, U, j/m)\|^2.$$

Using the fact that for $a, b \in l^2(\mathbb{N})$,

$$|\langle a, a\rangle - \langle b, b\rangle| \leq \|a - b\|\left(2\|a\| + \|a - b\|\right), \tag{4.3.2}$$

it follows that

$$\left|\|Q_n e^{-iAs}\chi_U(A)x\|^2 - \|\widetilde{\Gamma}_{n,m}(A, x, U, s)\|^2\right| \leq \frac{C(A, x, U)}{m}\left(2\|x\| + \frac{C(A, x, U)}{m}\right).$$

Hence we have shown that

$$\begin{aligned}\left|\Gamma_{n,m}(A, x, U) - \frac{1}{m}\int_0^m \left\|Q_n e^{-iAs}\chi_U(A)x\right\|^2 ds\right| &\leq \frac{1}{m^2}\sum_{j=1}^{m^2}\frac{C(A, x, U)}{m}\left(2\|x\| + \frac{C(A, x, U)}{m}\right) \\ &+ \frac{1}{m^2}\sum_{j=1}^{m^2}\left|g_n(j/m) - m\int_{\frac{j-1}{m}}^{\frac{j}{m}} g_n(s)ds\right|,\end{aligned}$$

where $g_n(s) = \|Q_n e^{-iAs}\chi_U(A)x\|^2$. Clearly the first term converges to 0 as $m \to \infty$ so we only need to consider the second. Using (4.3.2), it follows that for any $\epsilon > 0$ that

$$|g_n(s) - g_n(s+\epsilon)| \leq 4\|Q_n e^{-iAs}(e^{-iA\epsilon} - I)\chi_U(A)x\|\|x\| \leq 4\|x\|\|(e^{-iA\epsilon} - I)\chi_U(A)x\|.$$

But $e^{-iA\epsilon} - I$ converges strongly to 0 as $\epsilon \downarrow 0$ and hence the quantity

$$\left| g_n(j/m) - m \int_{\frac{j-1}{m}}^{\frac{j}{m}} g_n(s)ds \right| \to 0$$

as $m \to \infty$ uniformly in j. It follows that

$$\lim_{m\to\infty} \Gamma_{n,m}(A,x,U) = \lim_{t\to\infty} \frac{1}{t}\int_0^t \left\|Q_n e^{-iAs}E_U^A x\right\|^2 ds$$

and hence

$$\lim_{n\to\infty}\lim_{m\to\infty} \Gamma_{n,m}(A,x,U) = \langle P_{\mathrm{c}}^A E_U^A x, x\rangle.$$

Step 2: Next we deal with the case $\mathcal{I} = s$. Note that for $z \in \mathbb{C}\backslash\mathbb{R}$, $\langle R(z,A)x,x\rangle$ is simply the Stieltjes transform (also called the Borel transform) of the positive measure $\mu_{x,x}^A$

$$\langle R(z,A)x,x\rangle = \int_{\mathbb{R}} \frac{1}{\lambda - z} d\mu_{x,x}^A(\lambda).$$

The Hilbert transform of $\mu_{x,x}^A$ is given by the limit

$$H_{\mu_{x,x}^A}(t) = \frac{1}{\pi}\lim_{\epsilon\downarrow 0}\mathrm{Re}\left(\langle R(t+i\epsilon,A)x,x\rangle\right),$$

with the limit existing (Lebesgue) almost everywhere. This object was studied in [PSZ10, Pol96], where we shall use the result (since the measure is positive) that for any bounded continuous function f,[2]

$$\lim_{s\to\infty}\frac{\pi s}{2}\int_{\mathbb{R}} f(t)\chi_{\{w:|H_{\mu_{x,x}^A}(w)|\geq s\}}(t)dt = \int_{\mathbb{R}} f(t)d\mu_{x,x,\mathrm{s}}^A(t). \tag{4.3.3}$$

Now let $(A,x,U) \in \Omega_{f,\alpha,\beta} \times \mathcal{U}$ with

$$U = \bigcup_m (a_m, b_m)$$

where $a_m, b_m \in \mathbb{R}\cup\{\pm\infty\}$ and the disjoint union is at most countable as in (4.3.1). Without loss of generality, we assume that the union is over $m \in \mathbb{N}$. Due to the possibility of point spectra at the endpoints a_m, b_m, we cannot simply replace f by χ_U in the above limit (4.3.3). However, this can be overcome in the following manner.

Let ∂U denote the boundary of U defined by $\overline{U}\backslash U$ and let ν denote the measure $\mu_{x,x}^A|_{\partial U}$. Let f_s denote a pointwise increasing sequence of continuous functions, converging everywhere up to χ_U, such that the support of each f_s is contained in

$$[-s,s]\bigcap\left(\bigcup_{m=1}^{\lceil s\rceil}\left(a_m + 1/\sqrt{s}, b_m - 1/\sqrt{s}\right)\right).$$

Such a sequence exists (and can easily be explicitly constructed) precisely because U is open. We first claim that

$$\lim_{s\to\infty}\frac{\pi s}{2}\int_{\mathbb{R}} f_s(t)\chi_{\{w:|H_{\mu_{x,x}^A}(w)|\geq s\}}(t)dt = \mu_{x,x,\mathrm{s}}^A(U). \tag{4.3.4}$$

To see this note that for any $u \in \mathbb{R}$, the following inequalities hold

$$\begin{aligned}\liminf_{s\to\infty}\frac{\pi s}{2}\int_{\mathbb{R}} f_s(t)\chi_{\{w:|H_{\mu_{x,x}^A}(w)|\geq s\}}(t)dt &\geq \liminf_{s\to\infty}\frac{\pi s}{2}\int_{\mathbb{R}} f_u(t)\chi_{\{w:|H_{\mu_{x,x}^A}(w)|\geq s\}}(t)dt \\ &= \int_{\mathbb{R}} f_u(t)d\mu_{x,x,\mathrm{s}}^A(t).\end{aligned}$$

[2] Note that this is stronger than weak* convergence which in this case means restricting to continuous functions vanishing at infinity. That the result holds for arbitrary bounded continuous functions is due to the tightness condition that the result holds for the function identically equal to 1.

Taking $u \to \infty$ gives that

$$\liminf_{s\to\infty} \frac{\pi s}{2} \int_{\mathbb{R}} f_s(t)\chi_{\{w:|H_{\mu_{x,x}^A}(w)|\geq s\}}(t)dt \geq \mu_{x,x,s}^A(U), \tag{4.3.5}$$

so we are left with proving a similar bound for the limit supremum. Note that any point in the support of f_s is of distance at least $1/\sqrt{s}$ from ∂U. It follows that there exists a constant C independent of t such that for any $t \in \operatorname{supp}(f_s)$,

$$|H_\nu(t)| \leq C\sqrt{s}$$

Now let $\epsilon \in (0,1)$. Then, for large s, $s - C\sqrt{s} \geq (1-\epsilon)s$ and hence

$$\operatorname{supp}(f_s) \cap \{w : |H_{\mu_{x,x}^A}(w)| \geq s\} \subset \operatorname{supp}(f_s) \cap \{w : |H_{\mu_{x,x}^A - \nu}(w)| \geq (1-\epsilon)s\}. \tag{4.3.6}$$

Now let f be any bounded continuous function such that $f \geq \chi_U$. Then using (4.3.6),

$$\begin{aligned}
&\limsup_{s\to\infty} \frac{\pi s}{2} \int_{\mathbb{R}} f_s(t)\chi_{\{w:|H_{\mu_{x,x}^A}(w)|\geq s\}}(t)dt \\
&\quad\leq \limsup_{s\to\infty} \frac{1}{1-\epsilon}\frac{\pi(1-\epsilon)s}{2} \int_{\mathbb{R}} f_s(t)\chi_{\{w:|H_{\mu_{x,x}^A-\nu}(w)|\geq(1-\epsilon)s\}}(t)dt \\
&\quad\leq \limsup_{s\to\infty} \frac{1}{1-\epsilon}\frac{\pi(1-\epsilon)s}{2} \int_{\mathbb{R}} f(t)\chi_{\{w:|H_{\mu_{x,x}^A-\nu}(w)|\geq(1-\epsilon)s\}}(t)dt \\
&\quad= \frac{1}{1-\epsilon} \int_{\mathbb{R}} f(t)d([\mu_{x,x}^A - \nu]_s)(t).
\end{aligned}$$

Now we let $f \downarrow \chi_{\overline{U}}$, with pointwise convergence everywhere. This is possible since the complement of $\overline{U}$ is open. By the dominated convergence theorem, and since ϵ was arbitrary, this yields

$$\limsup_{s\to\infty} \frac{\pi s}{2} \int_{\mathbb{R}} f_s(t)\chi_{\{w:|H_{\mu_{x,x}^A}(w)|\geq s\}}(t)dt \leq [\mu_{x,x}^A - \nu]_s(\overline{U}) = \mu_{x,x,s}^A(U),$$

where the last equality follows from the definition of ν. The claim (4.3.4) now follows.

Let χ_n be a sequence of non-negative continuous piecewise affine functions on $\mathbb{R}$, bounded by 1 and such that $\chi_n(t) = 0$ if $t \leq n-1$ and $\chi_n(t) = 1$ if $t \geq n+1$. Consider the integrals

$$I(n,m) = \frac{\pi n}{2} \int_{\mathbb{R}} f_n(t)\chi_n(|F_m(t)|)dt,$$

where $F_m(t)$ is an approximation of

$$\frac{1}{\pi}\mathrm{Re}\left(\left\langle R\left(t + \frac{i}{m}, A\right)x, x\right\rangle\right)$$

to pointwise accuracy $O(m^{-1})$ over $t \in [-n,n]$. Note that a suitable piecewise linear function f_n can be constructed using $\widetilde{\Lambda}_1$, as can suitable χ_n, and a suitable approximation function F_m can be pointwise evaluated using $\widetilde{\Lambda}_1$ (again by Corollary 4.2.2). It follows that there exists arithmetic algorithms $\Gamma_{n,m}(A,x,U)$ using $\widetilde{\Lambda}_1$ such that

$$|I(n,m) - \Gamma_{n,m}(A,x,U)| \leq \frac{C(A,x,U)}{m}.$$

The dominated convergence theorem implies that

$$\lim_{m\to\infty} \Gamma_{n,m}(A,x,U) = \lim_{m\to\infty} I(n,m) = \frac{\pi n}{2} \int_{\mathbb{R}} f_n(t)\chi_n(|H_{\mu_{x,x}^A}(t)|)dt.$$

Note that continuity of χ_n is needed to gain convergence almost everywhere and prevent possible oscillations about the level set $\{H_{\mu_{x,x}^A}(t) = n\}$. We also have

$$\chi_{\{w:|H_{\mu_{x,x}^A}(w)|\geq n+1\}}(t) \leq \chi_n(|H_{\mu_{x,x}^A}(t)|) \leq \chi_{\{w:|H_{\mu_{x,x}^A}(w)|\geq n-1\}}(t)$$

The same arguments used to prove (4.3.4), therefore show that

$$\lim_{n\to\infty} \frac{\pi n}{2} \int_{\mathbb{R}} f_n(t)\chi_n(|H_{\mu^A_{x,x}}(t)|)dt = \mu^A_{x,x,\mathrm{s}}(U).$$

Hence,

$$\lim_{n\to\infty} \lim_{m\to\infty} \Gamma_{n,m}(A,x,U) = \mu^A_{x,x,\mathrm{s}}(U),$$

completing the proof of inclusion in Theorem 4.3.2. □

Proof of exclusion in Theorem 4.3.2

To prove the exclusion, we need two results which will also be used in Chapter 5. Namely, a result connected to Anderson localisation (Theorem 5.2.1) and a result concerning sparse potentials of discrete Schrödinger operators (Theorem 5.3.3). We also introduce some notation which will also be used in Chapter 5. Consider a connected, undirected graph G, such that the degree of each vertex is bounded by some constant C_G and such that the set of vertices $V(G)$ is countably infinite. We also assume that there exists at most one edge between two vertices and no edges from a vertex to itself. We use the abuse of notation by identifying each $x \in V$ with its canonical vector in $l^2(V(G)) \cong l^2(\mathbb{N})$. The notation $x \sim y$ means there is an edge in G connecting vertices x and y. We will use $|x-y|$ to denote the length of a shortest path between vertices x, y (which always exists since the graph is connected), and $\zeta(x)$ to denote the valence of x. An arbitrary base vertex x_0 is chosen and we define $|x| = |x - x_0|$.

The (negative) discrete Laplacian or free Hamiltonian H_0 acts on $\psi \in l^2(V(G))$ via

$$\{H_0\psi\}(x) = -\sum_{y\sim x}[\psi(y) - \psi(x)].$$

Since the vertex degree is bounded, H_0 is a bounded operator. We define a Schrödinger operator on G to be an operator of the form

$$H_v = H_0 + v,$$

where v is a bounded (real-valued) multiplication operator

$$\{v\psi\}(x) = v(x)\psi(x).$$

Proof of exclusion in Theorem 4.3.2. Since $P^A_{\mathrm{pp}} = I - P^A_{\mathrm{c}}$, $P^A_{\mathrm{ac}} = I - P^A_{\mathrm{s}}$ and $P^A_{\mathrm{sc}} = P^A_{\mathrm{s}} - P^A_{\mathrm{pp}}$, it is enough, by Theorem 4.3.1, to consider $\mathcal{I} = \mathrm{pp}$, ac and sc. We restrict the proof to considering bounded Schrödinger operators H_v acting on $l^2(\mathbb{N})$, which are clearly a subclass of $\Omega_{f,0}$ for $f(n) = n+1$. In this distinguished case, we truncate the operator naturally defined on $l^2(\mathbb{Z})$ and define

$$H_0 = \begin{pmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & \ddots \\ & & \ddots & \ddots \end{pmatrix}$$

We also set $x = e_1$, with the crucial properties that this vector is cyclic and hence $\mu^{H_v}_{e_1,e_1}$ has the same support as $\mathrm{Sp}(H_v)$, and that $x \in V_0$. Throughout, we also take $U = (0,4)$.

Step 1: We begin with P^A_{pp}. Suppose for a contradiction that there does exist a sequence of general algorithms Γ_n such that

$$\lim_{n\to\infty} \Gamma_n(H_v) = \langle P^{H_v}_{\mathrm{pp}} E^{H_v}_{(0,4)} e_1, e_1\rangle.$$

We take a general algorithm, denoted $\widehat{\Gamma}_n$, from Theorem 4.3.1 which has

$$\lim_{n\to\infty} \widehat{\Gamma}_n(H_v) = \mu_{e_1,e_1}^{H_v}((0,4)).$$

Since e_1 is cyclic, this limit is non-zero if $(0,4) \cap \mathrm{Sp}(H_v) \neq \emptyset$. We therefore define

$$\widetilde{\Gamma}_n(H_v) = \begin{cases} 0 & \text{if } \widehat{\Gamma}_n(H_v) = 0 \\ \frac{\Gamma_n(H_v)}{\widehat{\Gamma}_n(H_v)} & \text{otherwise.} \end{cases}$$

We will use Theorem 5.2.1 and the following well-known facts:

1. If for any $l \in \mathbb{N}$ there exists m_l such that $v(m_l + 1) = v(m_l + 2) = ... = v(m_l + l) = 0$, then $(0,4) \subset \mathrm{Sp}(H_v)$.

2. If there exists $N \in \mathbb{N}$ such that $v(n)$ is 0 for $n \geq N$, then $\mathrm{Sp}_{\mathrm{pp}}(H_v) \cap (0,4) = \emptyset$ [Rem98], but $[0,4] \subset \mathrm{Sp}(H_v)$ (the potential acts as a compact perturbation so the essential spectrum is $[0,4]$).

3. If we are in the setting of Theorem 5.2.1, then the spectrum of $H_{v_\omega} + A$ is pure point almost surely. Moreover, if $\rho = \chi_{[-c,c]}/(2c)$ for some constant c, then $[-c, 4+c] \subset \mathrm{Sp}_{\mathrm{pp}}(H_{v_\omega} + A)$ almost surely.

The strategy will be to construct a potential v such that $(0,4) \subset \mathrm{Sp}(H_v)$, yet $\widetilde{\Gamma}_n(H_v)$ does not converge. This is a contradiction since by our assumptions, for such a v we must have

$$\widetilde{\Gamma}_n(H_v) \to \frac{\langle P_{\mathrm{pp}}^{H_v} E_{(0,4)}^{H_v} e_1, e_1 \rangle}{\mu_{e_1,e_1}^{H_v}((0,4))}.$$

To do this, choose $\rho = \chi_{[-c,c]}/(2c)$ for some constant c such that the conditions of Theorem 5.2.1 hold and define the potential v inductively as follows.

Let v_1 be a potential of the form v_ω (with the density ρ) such that $\mathrm{Sp}(H_{v_1})$ is pure point. Such a v_1 exists by Theorem 5.2.1 and we have $\langle P_{\mathrm{pp}}^{H_{v_1}} E_{(0,4)}^{H_{v_1}} e_1, e_1 \rangle = \mu_{e_1,e_1}^{H_{v_1}}((0,4))$. Hence for large enough n it must hold that $\widetilde{\Gamma}_n(H_{v_1}) > 3/4$. Fix $n = n_1$ such that this holds. Then $\Gamma_{n_1}(H_{v_1})$ only depends on $\{v_1(j) : j \leq N_1\}$ for some integer N_1 by (i) of Definition 2.1.1. Define the potential v_2 by $v_2(j) = v_1(j)$ for all $j \leq N_1$ and $v_2(j) = 0$ otherwise. Then by fact (2) above, $\langle P_{\mathrm{pp}}^{H_{v_2}} E_{(0,4)}^{H_{v_2}} e_1, e_1 \rangle = 0$ but $\mu_{e_1,e_1}^{H_{v_2}}((0,4)) \neq 0$, and hence $\widetilde{\Gamma}_n(H_{v_2}) < 1/4$ for large n, say for $n = n_2 > n_1$. But then $\Gamma_{n_2}(H_{v_2})$ only depends on $\{v_2(j) : j \leq N_2\}$ for some integer N_2.

We repeat this process inductively switching between potentials which induce $\widetilde{\Gamma}_{n_k}(H_{v_k}) < 1/4$ for k even and potentials which induce $\widetilde{\Gamma}_{n_k}(H_{v_k}) > 3/4$ for k odd. Explicitly, if k is even then define a potential v_{k+1} by $v_{k+1}(j) = v_k(j)$ for all $j \leq N_k$ and $v_{k+1}(j) = v_\omega(j)$ (with the density ρ) otherwise such that the spectrum of H_{v_k} is pure point. Such a ω exists from Theorem 5.2.1 applied with the perturbation A to match the potential for $j \leq N_k$. If k is odd then we define v_{k+1} by $v_{k+1}(j) = v_k(j)$ for all $j \leq N_k$ and $v_{k+1}(j) = 0$ otherwise. We can then choose n_{k+1} such that the above inequalities hold and N_{k+1} such that $\Gamma_{n_{k+1}}(H_{v_{k+1}})$ only depends on $\{v_{k+1}(j) : j \leq N_{k+1}\}$. We also ensure that $N_{k+1} \geq N_k + k$.

Finally set $v(j) = v_k(j)$ for $j \leq N_k$. It is clear from (iii) of Definition 2.1.1, that $\widetilde{\Gamma}_{n_k}(H_v) = \widetilde{\Gamma}_{n_k}(H_{v_k})$ and this implies that $\widetilde{\Gamma}_{n_k}(H_v)$ cannot converge. However, since $N_{k+1} \geq N_k + k$, for any k odd we have $v(N_k + 1) = v(N_k + 2) = ... = v(N_k + k) = 0$. Fact (1) implies that $(0,4) \subset \mathrm{Sp}(H_v)$, hence $\mu_{e_1,e_1}^{H_v}((0,4)) \neq 0$ and therefore $\widetilde{\Gamma}_n(H_v)$ converges. This provides the required contradiction.

Step 2: Next we deal with $\mathcal{I} = \mathrm{ac}$. To prove that one limit will not suffice, our strategy will be to reduce a certain decision problem to the computation of Ξ_{ac}. Let $(\mathcal{M}', d')$ be the discrete space $\{0,1\}$, let

Ω' denote the collection of all infinite sequence $\{a_j\}_{j\in\mathbb{N}}$ with entries $a_j \in \{0,1\}$ and consider the problem function

$$\Xi'(\{a_j\}): \text{ Does } \{a_j\} \text{ have infinitely many non-zero entries?}$$

In [Colns], it was shown that $\mathrm{SCI}(\Xi',\Omega')_G = 2$ (where the evaluation functions consist in component-wise evaluation of the array $\{a_j\}$). Suppose for a contradiction that Γ_n is a height one tower of general algorithms such that

$$\lim_{n\to\infty} \Gamma_n(H_v) = \langle P_{\mathrm{ac}}^{H_v} E_{(0,4)}^{H_v} e_1, e_1\rangle.$$

We will gain a contradiction by using the supposed tower to solve $\{\Xi',\Omega'\}$.

Given $\{a_j\}\in\Omega'$, consider the operator H_v, where the potential is of the following form:

$$v(m) = \sum_{k=1}^{\infty} a_k \delta_{m,k!}. \tag{4.3.7}$$

Then by Theorem 5.3.3, $\langle P_{\mathrm{ac}}^{H_v} E_{(0,4)}^{H_v} e_1, e_1\rangle = \mu_{e_1,e_1}^{H_v}((0,4))$ if $\sum_k a_k < \infty$ (that is, if $\Xi'(\{a_j\}) = 0$) and $\langle P_{\mathrm{ac}}^{H_v} E_{(0,4)}^{H_v} e_1, e_1\rangle = 0$ otherwise. Note that in either case we have $\mu_{e_1,e_1}^{H_v}((0,4)) \neq 0$. We follow Step 1 and take a general algorithm, denoted $\widehat{\Gamma}_n$, from Theorem 4.3.1 which has

$$\lim_{n\to\infty} \widehat{\Gamma}_n(H_v) = \mu_{e_1,e_1}^{H_v}((0,4)).$$

Since e_1 is cyclic, this limit is non-zero for H_v, where v is of the form (4.3.7). We therefore define

$$\widetilde{\Gamma}_n(H_v) = \begin{cases} 0 & \text{if } \widehat{\Gamma}_n(H_v) = 0 \\ \frac{\Gamma_n(H_v)}{\widehat{\Gamma}_n(H_v)} & \text{otherwise.}\end{cases}$$

It follows that

$$\lim_{n\to\infty} \widetilde{\Gamma}_n(H_v) = \begin{cases} 1 & \text{if } \Xi'(\{a_j\}) = 0 \\ 0 & \text{otherwise.}\end{cases}$$

Given N we can evaluate any matrix value of H using only finitely many evaluations of $\{a_j\}$ and hence the evaluation functions $\widetilde{\Lambda}_1$ can be computed using component-wise evaluations of the sequence $\{a_j\}$. We now set

$$\overline{\Gamma}_n(\{a_j\}) = \begin{cases} 0 & \text{if } \Gamma_n(H_v) > \frac{1}{2} \\ 1 & \text{otherwise.}\end{cases}$$

The above comments show that each of these is a general algorithm and it is clear that it converges to $\Xi'(\{a_j\})$ as $n\to\infty$, the required contradiction.

Step 3: Finally, we must deal with $\mathcal{I} = \mathrm{sc}$. The argument is the same as Step 2, but now with replacing $\langle P_{\mathrm{ac}}^{H_v} E_{(0,4)}^{H_v} e_1, e_1\rangle$ with $\langle P_{\mathrm{sc}}^{H_v} E_{(0,4)}^{H_v} e_1, e_1\rangle$ and the resulting $\widetilde{\Gamma}_n(H_v)$ with $1 - \widetilde{\Gamma}_n(H_v)$. □

4.4 Two Important Applications

Theorem 4.3.1 can be extended to computing the functional calculus. Recall that given a (possibly unbounded complex-valued) Borel function F, defined on $\mathbb{C}$, and $A\in\Omega_{\mathrm{N}}$, $F(A)$ is defined by

$$F(A) = \int_{\mathrm{Sp}(A)} F(\lambda)\, dE^A(\lambda).$$

$F(A)$ is a densely defined closed normal operator with dense domain given by

$$\mathcal{D}(F(A))=\left\{x\in l^2(\mathbb{N}):\int_{\mathrm{Sp}(A)}\left|F(\lambda)\right|^2 d\mu_{x,x}^A(\lambda)<\infty\right\}.$$

For simplicity, we will only deal with the case that F is a bounded continuous function on $\mathbb{R}$, that is, $F\in C_b(\mathbb{R})$. In this case $\mathcal{D}(F(A))$ is the whole of $l^2(\mathbb{N})$ (the measures $\mu_{x,y}^A$ are finite) and we can use standard properties of the Poisson kernel. We assume that given $F\in C_b(\mathbb{R})$ we have access to piecewise constant functions F_n supported in $[-n,n]$ such that $\|F-F_n\|_{l^\infty([-n,n])}\leq n^{-1}$. Clearly other suitable data also suffices and as usual we abuse notation slightly by adding this information to Λ_1 to define $\widetilde{\Lambda}_1$.

Theorem 4.4.1 (Computation of the functional calculus)**.** *Consider the map*

$$\begin{aligned}\Xi_{\mathrm{fun}}:\Omega_{f,\alpha,\beta}\times C_b(\mathbb{R})&\to l^2(\mathbb{N})\\(A,x,F)&\to F(A)x.\end{aligned}$$

Then $\{\Xi_{\mathrm{fun}},\Omega_{f,\alpha,\beta}\times C_b(\mathbb{R}),\widetilde{\Lambda}_1\}\in\Delta_2^A$.

Proof. Let $(A,x,F)\in\Omega_{f,\alpha,\beta}\times C_b(\mathbb{R})$ then by Fubini's theorem,

$$\int_{-n}^{n}K_H(u+i/n;A,x)F_n(u)du=\int_{-\infty}^{\infty}\int_{-n}^{n}P_H(u-\lambda,1/n)F_n(u)du\,dE^A(\lambda)x.$$

The inner integral is bounded since F is bounded and the Poisson kernel integrates to 1 along the real line. It also converges to $F(\lambda)$ everywhere. Hence by the dominated convergence theorem

$$\lim_{n\to\infty}\int_{-n}^{n}K_H(u+i/n;A,x)F_n(u)du=F(A)x.$$

We now use the same arguments used to prove Theorem 4.3.1. Using Corollary 4.2.2, together with $\|K_H(u+i/n;A,x)\|_{l^\infty(\mathbb{R})}\leq nC_1$ and the fact that $K_H(u+i/n;A,x)$ is Lipschitz continuous with Lipschitz constant n^2C_2 for some (possibly unknown) constants C_1 and C_2, we can approximate this integral with an error that vanishes in the limit $n\to\infty$. □

Recall the definition of the Radon–Nikodym derivative in (4.1.1) and note that $\rho_{x,y}^A\in L^1(\mathbb{R})$ for $A\in\Omega_{\mathrm{SA}}$. We consider its computation in L^1 sense in the following theorem, where, as before, we assume (4.3.1), adding the approximations of U to our evaluation set along with component-wise evaluations of a given vector y to form $\widetilde{\Lambda}_1$. However, we must consider the computation away from the singular part of the spectrum - this is also reflected in the results of §4.5.2.

Theorem 4.4.2 (Computation of the Radon–Nikodym derivative)**.** *Consider the map*

$$\begin{aligned}\Xi_{\mathrm{RN}}:\Omega_{f,\alpha,\beta}\times l^2(\mathbb{N})\times\mathcal{U}&\to L^1(\mathbb{R})\\(A,x,y,U)&\to\rho_{x,y}^A|_U.\end{aligned}$$

We restrict this map to the quadruples (A,x,y,U) such that U is strictly separated from $\mathrm{supp}(\mu_{x,y,\mathrm{sc}}^A)\cup\mathrm{supp}(\mu_{x,y,\mathrm{pp}}^A)$ *and denote this subclass by $\widetilde{\Omega}_{f,\alpha,\beta}$. Then $\{\Xi_{\mathrm{RN}},\widetilde{\Omega}_{f,\alpha,\beta},\widetilde{\Lambda}_1\}\in\Delta_2^A$. Furthermore, each output $\Gamma_n(A,x,y,U)$ consists of a piecewise linear function, supported in U with rational knots and taking (complex) rational values at these knots.*

Proof. Let $(A, x, y, U) \in \widetilde{\Omega}_{f,\alpha,\beta}$. For $u \in U$ we decompose as follows

$$\begin{aligned}\langle K_H(u+i\epsilon; A, x), y\rangle =& \frac{1}{\pi}\int_{\mathbb{R}} \frac{\epsilon}{(\lambda-u)^2+\epsilon^2}\rho^A_{x,y}(\lambda)d\lambda \\ &+ \frac{1}{\pi}\int_{\mathbb{R}\backslash U} \frac{\epsilon}{(\lambda-u)^2+\epsilon^2}\left\{d\mu^A_{x,y,\mathrm{sc}}(\lambda) + d\mu^A_{x,y,\mathrm{pp}}(\lambda)\right\}.\end{aligned} \tag{4.4.1}$$

The first term converges to $\rho^A_{x,y}|_U$ in $L^1(U)$ as $\epsilon \downarrow 0$ since $\rho^A_{x,y}|_U \in L^1(U)$. Since we assumed that U is separated from $\mathrm{supp}(\mu^A_{x,y,\mathrm{sc}}) \cup \mathrm{supp}(\mu^A_{x,y,\mathrm{pp}})$, it follows that the second term of (4.4.1) converges to 0 in $L^1(U)$ as $\epsilon \downarrow 0$. Hence we are done if we can approximate $\langle K_H(u+i/n; A, x), y\rangle$ in $L^1(U)$ with an error converging to zero as $n \to \infty$.

Recall that $K_H(u+i/n; A, x)$ is Lipschitz continuous with Lipschitz constant at most $n^2\|x\|/\pi$. By assumption, and using Corollary 4.2.2, we can approximate $K_H(u+i/n; A, x)$ to asymptotic precision with vectors of finite support. Hence the inner product

$$f_n(u) := \langle K_H(u+i/n; A, x), y\rangle$$

can be approximated to asymptotic precision (now with a possibly unknown constant also depending on $\|y\|$) and f_n is Lipschitz continuous with Lipshitz constant at most $n^2\|x\|\|y\|/\pi$.

Recall that U can be written as the disjoint union

$$U = \bigcup_m (a_m, b_m)$$

where $a_m, b_m \in \mathbb{R} \cup \{\pm\infty\}$ and the union is at most countable. Without loss of generality, we assume that the union is over $m \in \mathbb{N}$. Given an interval (a_m, b_m), let $a_m < z_{m,1,n} < z_{m,2,n} < ... < z_{m,r_m,n} < b_m$ be such that we have $z_{m,j,n} \in \mathbb{Q}$ and $|z_{m,j,n} - z_{m,j+1,n}| \leq (b_m - a_m)^{-1}n^{-3}m^{-2}$ and $|a_m - z_{m,1,n}|, |b_m - z_{m,r_m,n}| \leq n^{-1}$. We also let $f_{m,n}$ be a piecewise affine interpolant with knots $z_{m,1,n}, ..., z_{m,r_m,n}$ supported on $(z_{m,1,n}, z_{m,r_m,n})$ with the property that $|f_{m,n}(z_{m,j,n}) - f_n(z_{m,j,n})| < C(b_m - a_m)^{-1}n^{-1}m^{-2}$. Here C is some unknown constant which occurs from the asymptotic approximation of f_n that arises from Corollary 4.2.2 and we can always compute such $f_{m,n}$ in finitely many arithmetic operations and comparisons.

Let $\Gamma_n(A, x, y, U)$ be the function that agrees with $f_{m,n}$ on (a_m, b_m) for $m \leq n$ and is zero elsewhere. Clearly the nodes of $\Gamma_n(A, x, y, U)$ can be computed using finitely many arithmetic operations and comparisons and the relevant set of evaluation functions $\widetilde{\Lambda}_1$. A simple application of the triangle inequality implies that

$$\begin{aligned}\int_U \left|\Gamma_n(A, U, x, y)(u) - \rho^A_{x,y}(u)\right| du \leq & \sum_{m>n}\int_{(a_m,b_m)} \left|\rho^A_{x,y}(u)\right| du \\ &+ \sum_{m\leq n}\int_{(a_m,b_m)\backslash(z_{m,1,n},z_{m,r_m,n})} \left|\rho^A_{x,y}(u)\right| du \\ &+ \sum_{m\leq n}\int_{(z_{m,1,n},z_{m,r_m,n})} \left|\rho^A_{x,y}(u) - f_n(u)\right| du + \frac{\widetilde{C}(x,y,A)}{n}\sum_{m\leq n}\frac{1}{m^2},\end{aligned}$$

where the last term is due to the piecewise linear interpolant. The bound converges to zero as required. □

4.5 High-order Kernels

4.5.1 Motivation

As an example, consider $L^2([-1,1])$ and the operator defined by

$$\mathcal{L}q(x) = xq(x) + \int_{-1}^{1} e^{-(x^2+y^2)}q(y)\,dy, \qquad x \in [-1,1]. \tag{4.5.1}$$

The operator $\mathcal{L}$ in (4.5.1) has continuous spectrum in $[-1,1]$, due to the multiplicative $xq(x)$ term, and discrete spectrum in $\mathbb{R}\setminus[-1,1]$ from the integral that acts as a compact perturbation. We discretise $\mathcal{L}$ with an $N \times N$ matrix corresponding to an adaptive Chebyshev collocation scheme. For efficient storage and computation of the resolvent, we exploit low numerical rank structure in the discretisation of the smooth kernel [TT13]. We apply a Clenshaw–Curtis quadrature rule to compute the inner products [Tre19] required to sample the scalar spectral measures.

There are two limits to take: $N \to \infty$ and $\epsilon \downarrow 0$. These two limits must be taken with considerable care [Col21]. If N is kept fixed as one takes $\epsilon \downarrow 0$, then the computed samples get polluted by the discrete spectrum of the discretisation. Instead, as one takes $\epsilon \downarrow 0$, one must appropriately increase N too. In practice, we increase N by selecting it adaptively to ensure that we adequately approximate the resolvent. Proposition 4.2.1 gives us a handle on how to choose N adaptively as $\epsilon \downarrow 0$. However, there is a numerical trade-off. Ideally, we would like to take ϵ small to recover a more accurate approximation of the spectral measure. On the other hand, we wish to evaluate the resolvent as far away from the spectrum as possible since, typically, evaluating nearer the spectrum requires larger discretisation sizes.[3]

For example, Figure 4.3 (left) shows the discretisation sizes, N, needed to evaluate the Radon–Nikodym derivative of the spectral measure convolved with the Poisson kernel accurately. Here, we evaluate at $x_0 = 1/2 \in [-1,1]$ and consider $\mu_{f,f}^{\mathcal{L}}$ with $f(x) = \sqrt{3/2}x$. For the operator in (4.5.1) and $\epsilon = 0.05, 0.01$, and 0.005, we need $N = 400$, 1700, and 3100, respectively. We have also shown (Figure 4.3 (right)) the error in the convolution approximation of the Radon–Nikodym derivative, which is of order $\mathcal{O}(\epsilon\log(\epsilon^{-1}))$ (see Theorem 4.5.2 below) for the Poisson kernel ($m = 1$). Unfortunately, to obtain samples of the spectral measure that have two digits of relative accuracy, we require that $\epsilon \approx 0.01$. Since we require $N \approx 20/\epsilon$ for small $\epsilon > 0$, it is computationally infeasible to obtain more than five or six digits of accuracy with the Poisson kernel. We have also shown the relative errors when using the high-order kernels developed in this section. The order is denoted by m, and the plot corresponds to $\mathcal{O}(\epsilon^m\log(\epsilon^{-1}))$ when m is odd and a $\mathcal{O}(\epsilon^m)$ when m is even. A sixth-order kernel enables us to achieve about 11 digits of accuracy without decreasing ϵ below 0.01. Although using a sixth-order kernel requires six times as many resolvent evaluations as that of the Poisson kernel (see below), this is typically favourable because the cost of evaluating the resolvent near the continuous spectrum of $\mathcal{L}$ increases as $\epsilon \downarrow 0$.

4.5.2 High-order kernels, high-order convergence and error control

It is well-known in signal processing and statistics that the convergence rate of convolutions is determined by the number of vanishing moments of the kernel. We therefore make the following definition:

Definition 4.5.1 (mth order kernel)**.** *Let m be a positive integer and $K \in L^1(\mathbb{R})$. We say that K is an mth order kernel if it has the following three properties:*

[3] Two reasons for this, explored in more detail in [CHT21], are the formation of interior layers and oscillatory behaviour of the solutions of the corresponding linear systems. This problem of needing large discretisations is distinct from, though related to, the problem of conditioning. If $x_0 \in \mathrm{Sp}(A)$, then $\|R(x_0 + i\epsilon, A)\| = \epsilon^{-1}$ and the shifted linear systems become increasingly ill-conditioned as $\epsilon \downarrow 0$. This can limit the attainable accuracy and is also important if one solves the shifted linear systems using iterative methods (more iterations may be required).

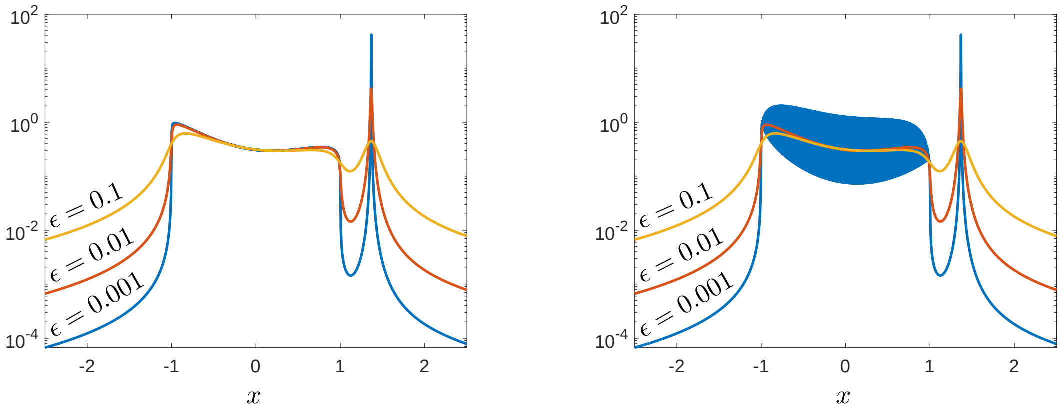
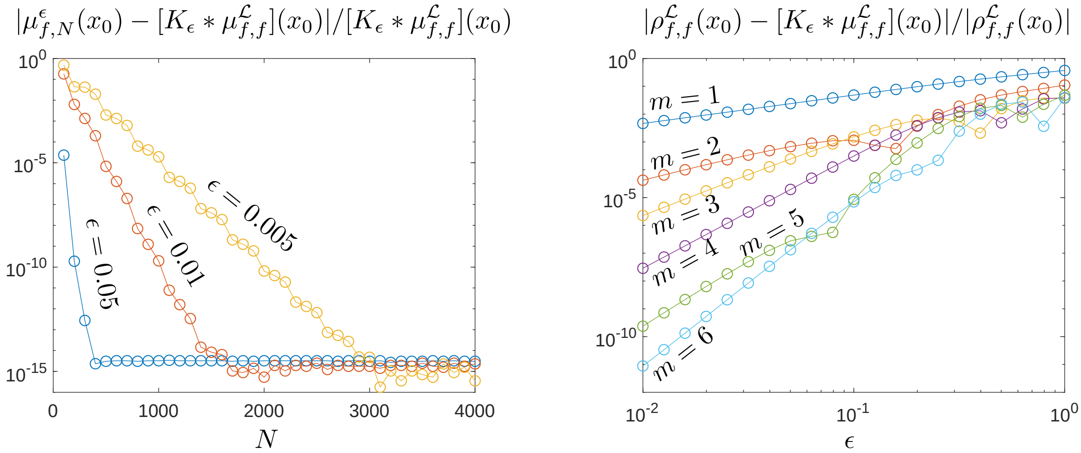


Figure 4.2: Left: The smoothed approximation $[K_\epsilon * \mu_{f,f}^{\mathcal{L}}]$ (K_ϵ denotes the rescaled Poisson kernel) for the integral operator in (4.5.1) and different ϵ. The discretisation sizes for solving the shifted linear systems are adaptively selected. Right: The same computation except with a fixed discretisation size of $N = 500$.


$|\mu_{f,N}^{\epsilon}(x_0) - [K_\epsilon * \mu_{f,f}^{\mathcal{L}}](x_0)|/[K_\epsilon * \mu_{f,f}^{\mathcal{L}}](x_0)$

$|\rho_{f,f}^{\mathcal{L}}(x_0) - [K_\epsilon * \mu_{f,f}^{\mathcal{L}}](x_0)|/|\rho_{f,f}^{\mathcal{L}}(x_0)|$

Figure 4.3: Left: The relative error in the numerical approximation, denoted by $\mu_{f,N}^{\epsilon}$, corresponding to discretisation size N, of the smoothed measure $[K_\epsilon * \mu_{f,f}^{\mathcal{L}}](x_0)$ (K_ϵ denotes the rescaled Poisson kernel) for the operator in (4.5.1) with $\epsilon = 0.05$, $\epsilon = 0.01$, and $\epsilon = 0.005$. Right: The pointwise relative error in smoothed measures of the operator in (4.5.1) computed using the high-order kernels with poles in (4.5.21) for $1 \leq m \leq 6$ (K_ϵ denotes the rescaled kernels). The relative errors are computed by comparing with numerical solutions that are resolved to machine precision.

(i) Normalised: $\int_{\mathbb{R}} K(x)dx = 1$.

(ii) Zero moments: $K(x)x^j$ *is integrable and* $\int_{\mathbb{R}} K(x)x^j dx = 0$ *for* $0 < j < m$.

(iii) Decay at $\pm\infty$*: There is a constant* C_K*, independent of* x*, such that*

$$|K(x)| \leq \frac{C_K}{(1+|x|)^{m+1}}, \qquad x \in \mathbb{R}. \tag{4.5.2}$$

We denote the rescaled kernel $\epsilon^{-1}K(\epsilon^{-1}\cdot)$ by K_ϵ. For example, the Poisson kernel used previously in this chapter is a first-order kernel and is not a second-order kernel. In contrast, the Gaussian kernel, $h(x) = (2\pi)^{-1/2}e^{-x^2/2}$, is a second-order kernel which plays an important role in density of states calculations [LSY16] and kernel density estimation [Sil18]. However, it is not particularly useful in our setting since it is not clear how to approximate the convolutions $h_\epsilon * \mu_{x,y}^A$. We will see in §4.5.3 that rational kernels are much more useful in this regard since we can compute the convolution by computing the action of the resolvent with error control, just like we did for the Poisson kernel.

The results of this subsection are stated in terms of convergence of convolutions for probability measures. However, by rescaling and the polar identity, corresponding results for the spectral measures $\mu_{x,y}^A$ can easily be obtained. We let $C^{k,\alpha}(I)$ denote the Hölder space of functions that are k times continuously differentiable on an interval I with an α-Hölder continuous kth derivative [Eva10]. For $h_1 \in C^{0,\alpha}(I)$ and $h_2 \in C^{k,\alpha}(I)$ we set

$$|h_1|_{C^{0,\alpha}(I)} = \sup_{x\neq y\in I} \frac{|h_1(x)-h_1(y)|}{|x-y|^\alpha}, \quad \|h_2\|_{C^{k,\alpha}(I)} = |h_2^{(k)}|_{C^{0,\alpha}(I)} + \max_{0\leq j\leq k} \|h_2^{(j)}\|_{\infty,I}.$$

The following theorem describes the pointwise convergence rates.

Theorem 4.5.2. *Let* K *be an* m*th order kernel,* μ *denote a probability measure on* $\mathbb{R}$ *and let* $\epsilon, \eta > 0$*. Suppose that* $x \in \mathbb{R}$ *is such that* μ *is absolutely continuous on the interval* $I = [x-\eta, x+\eta]$ *with* $C^{n,\alpha}(I)$ *Radon–Nikodym derivative* $\rho|_I$ *(with respect to Lebesgue measure), where* $n \in \mathbb{N}_{\geq 0}$*,* $\alpha \in [0,1)$ *and* $n+\alpha > 0$*. Then*

(i) If $n+\alpha < m$*, then, for a constant* $C(n,\alpha)$ *depending only on* n *and* α*,*

$$|\rho|_I(x) - K_\epsilon * \mu(x)| \leq \frac{C_K\epsilon^m}{(\epsilon+\frac{\eta}{2})^{m+1}} + C(n,\alpha)\|\rho|_I\|_{C^{n,\alpha}(I)} \int_{\mathbb{R}} |K(y)||y|^{n+\alpha}\, dy \left(1+\eta^{-n-\alpha}\right)\epsilon^{n+\alpha}.$$

(ii) If $n+\alpha \geq m$*, then, for a constant* $C(m)$ *depending only on* m*,*

$$|\rho|_I(x) - K_\epsilon * \mu(x)| \leq \frac{C_K\epsilon^m}{(\epsilon+\frac{\eta}{2})^{m+1}} + C(m)\|\rho|_I\|_{C^m(I)} \left(C_K + \int_{-\frac{\eta}{\epsilon}}^{\frac{\eta}{\epsilon}} |K(y)||y|^m\, dy\right)\left(1+\eta^{-m}\right)\epsilon^m.$$

Here, C_K *denotes the constant in (4.5.2).*

Remark 4.5.3. *If we fix* η *and consider small* ϵ*, then we obtain rates* $\mathcal{O}(\epsilon^{n+\alpha})$ *and* $\mathcal{O}(\epsilon^m \log(\epsilon^{-1}))$ *in cases (i) and (ii) respectively. One can show that these rates are, in general, sharp. Note that the error bound deteriorates when* η *becomes small (as expected).*

Proof. We first decompose

$$\rho|_I = g_1 + g_2,$$

where $g_1, g_2 \in C^{n+\alpha}(I)$ are both non-negative, g_1 is compactly supported in $(x-\eta, x+\eta)$ and g_2 is identically zero on $(x-\eta/2, x+\eta/2)$. Moreover, we can select g_1 so that in case (i) of the theorem,

$$\frac{1}{n!}\left|g_1^{(n)}\right|_{C^{0,\alpha}(I)} \leq C(n,\alpha)\|\rho|_I\|_{C^{n,\alpha}(I)}\left(1+\eta^{-n-\alpha}\right),$$

for some universal constant $C(n,\alpha)$ that only depends on n and α, whereas in case (ii),

$$2e\left\|g_1^{(m)}\right\|_\infty \leq C(m)\|\rho|_I\|_{C^m(I)}\left(1+\eta^{-m}\right),$$

for some universal constant $C(m)$ that only depends on m. Existence of such decompositions follows from standard arguments with cut-off functions.

First we deal with case (i) and assume that $\alpha > 0$. The case of $\alpha = 0$ is almost identical with some changes of indices. We use the following form of Taylor's theorem,

$$g_1(x+y) - g_1(x) = \sum_{j=1}^{n} \frac{g_1^{(j)}(x)}{j!} y^j + \int_0^y \int_0^{t_1} .... \int_0^{t_{n-1}} \left[g_1^{(n)}(t_n + x) - g_1^{(n)}(x)\right] dt_1...dt_n.$$

For notational convenience, let

$$M_n(x,y;g_1) = \int_0^y \int_0^{t_1} .... \int_0^{t_{n-1}} \left[g_1^{(n)}(t_n + x) - g_1^{(n)}(x)\right] dt_1...dt_n.$$

Substituting this into the convolution equation yields

$$K_\epsilon * g_1(x) - g_1(x) = \sum_{j=1}^{n} \frac{g_1^{(j)}(x)}{j!}\epsilon^{-1}\int_{\mathbb{R}} K\left(\frac{-y}{\epsilon}\right) y^j dy + \epsilon^{-1}\int_{\mathbb{R}} K\left(\frac{-y}{\epsilon}\right) M_n(x,y;g_1)dy. \quad (4.5.3)$$

Using the Hölder condition and direct integration, we have that

$$|M_n(x,y;g_1)| \leq \frac{|y|^{n+\alpha}}{(\alpha+1)\cdots(\alpha+n)}\left|g_1^{(n)}\right|_{C^{0,\alpha}(I)}.$$

Hence, by a change of variables $y \to -y$, the last integral in (4.5.3) is bounded by

$$\epsilon^{-1}\left|g_1^{(n)}\right|_{C^{0,\alpha}(I)}\int_{\mathbb{R}}\left|K\left(\frac{y}{\epsilon}\right)\right|\frac{|y|^{n+\alpha}}{(\alpha+1)\cdots(\alpha+n)}dy \leq \frac{\int_{\mathbb{R}}|K(y)|\,|y|^{n+\alpha}\,dy}{n!}\left|g_1^{(n)}\right|_{C^{0,\alpha}(I)} \cdot \epsilon^{n+\alpha}.$$

Since $n < m$ (recall that $\alpha > 0$ in the case we are dealing with), it follows that (again by a change of variables $y \to -y$) all the other integrals in (4.5.3) vanish and hence we have

$$|K_\epsilon * g_1(x) - g_1(x)| \leq \frac{\int_{\mathbb{R}}|K(y)|\,|y|^{n+\alpha}\,dy}{n!}\left|g_1^{(n)}\right|_{C^{0,\alpha}(I)} \cdot \epsilon^{n+\alpha}. \quad (4.5.4)$$

Due to the fact that g_1 and g_2 are non-negative, it follows that the measure $\mu - g_1 dx$ is non-negative, supported on the closure of $(x-\eta/2, x+\eta/2)^c$ and has total variation at most 1. Linearity of convolutions now implies that

$$|\rho|_I(x) - K_\epsilon * \mu(x)| \leq \frac{C_K\epsilon^m}{(\epsilon+\frac{\eta}{2})^{m+1}} + |K_\epsilon * g_1(x) - g_1(x)|.$$

Together with (4.5.4), this yields the result.

For case (ii), we use Taylor's theorem to obtain

$$\left|g_1(x+y) - \sum_{j=0}^{m-1}\frac{g_1^{(j)}(x)}{j!}y^j\right| \leq \frac{\|g_1^{(m)}\|_\infty\,|y|^m}{m!}.$$

We then split the range of integration, noting that $g_1(x+y)=0$ if $|y|>\eta$, to obtain

$$\begin{aligned}|K_\epsilon * g_1(x) - g_1(x)| &\leq |g_1(x)|\,\epsilon^{-1}\left|\int_{|y|\geq\eta} K\left(\frac{y}{\epsilon}\right)dy\right| \\ &+ \sum_{j=1}^{m-1}\frac{\left|g_1^{(j)}(x)\right|}{j!}\epsilon^{-1}\left|\int_{|y|\leq\eta} K\left(\frac{y}{\epsilon}\right)y^j\,dy\right| + \frac{\|g_1^{(m)}\|_\infty}{m!}\epsilon^{-1}\int_{|y|\leq\eta}\left|K\left(\frac{y}{\epsilon}\right)\right||y|^m\,dy.\end{aligned}$$

Due to the vanishing moments condition and decay (4.5.2), if $1\leq j<m$ then

$$\epsilon^{-1}\left|\int_{|y|\leq\eta} K\left(\frac{y}{\epsilon}\right)y^j\,dy\right| = \epsilon^{-1}\left|\int_{|y|\geq\eta} K\left(\frac{y}{\epsilon}\right)y^j\,dy\right| \leq \frac{2C_K}{m-j}\epsilon^j\left(\frac{\epsilon}{\eta}\right)^{m-j},$$

where the last equality follows by a change of variables. We can write out $g_1^{(j)}(x)$ as an iterated integral of $g_1^{(m)}$, to obtain $|g_1^{(j)}(x)|\leq\eta^{m-j}\|g_1^{(m)}\|_\infty$. It follows that

$$\begin{aligned}|K_\epsilon * g_1(x) - g_1(x)| &\leq \frac{\|g_1^{(m)}\|_\infty}{m!}\epsilon^m\int_{|y|\leq\frac{\eta}{\epsilon}}|K(y)|\,|y|^m\,dy + \sum_{j=0}^{m-1}\frac{\left|g_1^{(j)}(x)\right|}{j!}\cdot\frac{2C_K}{m-j}\cdot\epsilon^j\left(\frac{\epsilon}{\eta}\right)^{m-j} \\ &\leq \frac{\|g_1^{(m)}\|_\infty}{m!}\epsilon^m\int_{|y|\leq\frac{\eta}{\epsilon}}|K(y)|\,|y|^m\,dy + 2eC_K\|g_1^{(m)}\|_\infty\epsilon^m.\end{aligned}$$

We now argue as before to finish the proof. □

As well as pointwise error estimates, we can obtain L^p estimates which are useful when the Radon–Nikodym derivative has integrable singularities or in applications where the spectral measure is a probability measure (and hence L^1 convergence is natural). The convergence in L^p is most easily studied through the Fourier transform of the kernel, which in this section we define as

$$\widehat{K}(\omega) = \int K(x)\exp(2\pi i x\omega)dx.$$

Lemma 4.5.4. *Let K be an mth order kernel. Then $\widehat{K}$ is $m-1$ times continuously differentiable, $(\widehat{K})^{(j)}$ is bounded for $j=0,...,m-1$, and $(\widehat{K})^{(j)}(0)=0$ for $j=1,...,m-1$. Furthermore, for any $\alpha\in(0,1)$, $\widehat{K}\in C^{m-1,\alpha}(\mathbb{R})$.*

Exercise: Prove Lemma 4.5.4.

For an mth order kernel K, we define the function

$$\widehat{G_{m,K}}(\omega) := \frac{\widehat{K}(\omega)-1}{(2\pi i\omega)^m}.$$

Lemma 4.5.4 shows that $\widehat{G_{m,K}}\in L^2(\mathbb{R})$ and we denote its inverse Fourier transform by $G_{m,K}$. The following theorem gives the convergence rates of our smoothed approximation in the L^p sense.

Theorem 4.5.5. *Let K be an mth order kernel, μ denote a probability measure on $\mathbb{R}$ and let $\epsilon,\eta>0$. Then $G_{m,K}$ is bounded and satisfies*

$$|G_{m,K}(x)| \leq \frac{C_K}{m!(1+|x|)}. \tag{4.5.5}$$

Let $1\leq p<\infty$ and suppose that μ is absolutely continuous on the interval $I=(a-\eta,b+\eta)$ for $\eta>0$ and some $a<b$. Let ρ denote the Radon–Nikodym derivative of the absolutely continuous component of μ,

and suppose that $\rho_I := \rho|_I \in W^{m,p}(I)$. *Then*

$$\begin{aligned}\|\rho_I - [K_\epsilon * \mu]\|_{L^p,[a,b]} \leq & \frac{C_K(b-a)^{1/p}}{(\epsilon+\eta/2)^{m+1}}\epsilon^m \\ & + C(m)\int_{-((b-a)+2\eta)/\epsilon}^{((b-a)+2\eta)/\epsilon} |G_{m,K}(x)|\,dx \cdot (1+\eta^{-m}) \cdot \|\rho_I\|_{W^{m,p}(I)} \cdot \epsilon^m,\end{aligned} \tag{4.5.6}$$

where $C(m)$ *denotes a constant depending only on* m. *In particular, as* $\epsilon \downarrow 0$

$$\|\rho_I - [K_\epsilon * \mu]\|_{L^p,[a,b]} = \mathcal{O}(\epsilon^m \log(1/\epsilon)). \tag{4.5.7}$$

If there exists $\delta > 0$ *such that* $|K(x)(1+|x|)^{m+1+\delta}|$ *is bounded, then* $|G_{m,K}(x)(1+|x|)^{1+\delta}|$ *is also bounded and*

$$\|\rho_I - [K_\epsilon * \mu]\|_{L^p,[a,b]} = \mathcal{O}(\epsilon^m). \tag{4.5.8}$$

Proof of Theorem 4.5.5. We first argue for convolutions with smooth compactly supported functions and then take a limit. Let $g \in C_0^\infty$, the space of smooth compactly supported functions on $\mathbb{R}$, and let L denote the diameter of the support of g. For a function $F \in L^1(\mathbb{R})$, define the function

$$\phi_F(x) = \begin{cases} \int_{-\infty}^x F(t)dt - \int_\mathbb{R} F(t)dt, & \text{if } x > 0, \\ \int_{-\infty}^x F(t)dt, & \text{otherwise,} \end{cases}$$

which induces a map $\phi : F \to \phi_F$. Note that ϕ_F is bounded and decays at infinity. We let $\phi_{n,F}$ denote the n-fold iteration of ϕ applied to F (assuming that all of F, ϕ_F,...,$\phi_{n-1,F} \in L^1(\mathbb{R})$). The purpose of this map is that, in the sense of distributions, we have

$$F - \int_\mathbb{R} F(t)dt \cdot \delta_0 = \phi_F'$$

and hence

$$[F * g](x) - \int_\mathbb{R} F(t)dt \cdot g(x) = [-\phi_F * g'](x).$$

Applying this to $F = K_\epsilon$, we see that

$$[K_\epsilon * g](x) - g(x) = [-\phi_{K_\epsilon} * g'](x)$$

Note that if

$$F(x) \leq \frac{C}{(1+|x|)^{m+1}} \tag{4.5.9}$$

for some constant C, then

$$\phi_F(x) \leq \frac{C}{m(1+|x|)^m}. \tag{4.5.10}$$

Hence if $m > 1$, $\phi_{K_\epsilon} \in L^1(\mathbb{R})$ and we can apply the map again to obtain

$$[K_\epsilon * g](x) - g(x) = [\phi_{2,K_\epsilon} * g''](x) - \int_\mathbb{R} \phi_{K_\epsilon}(t)dt \cdot g'(x).$$

Inductively, we can apply the above argument to obtain the expression

$$[K_\epsilon * g](x) - g(x) = (-1)^m[\phi_{m,K_\epsilon} * g^{(m)}](x) + \sum_{j=1}^{m-1}(-1)^j \int_\mathbb{R} \phi_{j,K_\epsilon}(t)dt \cdot g^{(j)}(x). \tag{4.5.11}$$

Note that since K is an mth order kernel, ϕ_{m,K_ϵ} is bounded by a constant multiple of $(1+|x|)^{-1}$ and hence $\phi_{m,K_\epsilon} \in L^2(\mathbb{R})$. We can apply the convolution theorem, taking Fourier transforms, to obtain

$$(\widehat{K_\epsilon}(\omega)-1)\widehat{g}(\omega) = (-1)^m\widehat{\phi_{m,K_\epsilon}}(\omega)[-2\pi i\omega]^m\widehat{g}(\omega) + \sum_{j=1}^{m-1}(-1)^j\int_{\mathbb{R}}\phi_{j,K_\epsilon}(t)dt[-2\pi i\omega]^j\widehat{g}(\omega). \tag{4.5.12}$$

Since $g \in C_0^\infty(\mathbb{R})$ was arbitrary (and we can take $\widehat{g}(\omega) \neq 0$), it follows that

$$(-1)^m\widehat{\phi_{m,K_\epsilon}}(\omega) = \frac{(\widehat{K_\epsilon}(\omega)-1)}{(-2\pi i\omega)^m} - \sum_{j=1}^{m-1}(-1)^j\frac{\int_{\mathbb{R}}\phi_{j,K_\epsilon}(t)dt}{(-2\pi i\omega)^{m-j}}.$$

Since $\phi_{m,K_\epsilon} \in L^2(\mathbb{R})$, it follows that $\widehat{\phi_{m,K_\epsilon}} \in L^2(\mathbb{R})$. However, by Lemma 4.5.4, as $\omega \to 0$, $|\widehat{K_\epsilon}(\omega)-1| = \mathcal{O}(\omega^{m-1+\alpha})$ for any $\alpha \in (0,1)$. It follows that

$$\int_{\mathbb{R}}\phi_{j,K_\epsilon}(t)dt = 0$$

for $j = 1,...,m-1$. Hence we have $\phi_{m,K_1} = \phi_{m,K} = G_{m,K}$. Iterating (4.5.9) and (4.5.10) implies (4.5.5).

Now suppose that x lies in the support of g, then we can replace $\phi_{m,K_\epsilon}(x)$ by $\chi_{[-L,L]}(x)\phi_{m,K_\epsilon}(x)$ in (4.5.11), where χ_U denotes the indicator function of a set U. By Hölder's inequality, $\chi_{[-L,L]}\phi_{m,K_\epsilon} \in L^1(\mathbb{R})$ and hence, by Young's convolution inequality, it follows that

$$\begin{aligned}\|K_\epsilon * g - g\|_{L^p,\mathrm{supp}(g)} &\leq \left\|[\chi_{[-L,L]}\phi_{m,K_\epsilon}] * g^{(m)}\right\|_{L^p} &(4.5.13)\\ &\leq \int_{-L}^{L}|\phi_{m,K_\epsilon}(x)|\,dx\cdot\|g^{(m)}\|_{L^p}. &(4.5.14)\end{aligned}$$

Furthermore, we have by a simple change of variables that

$$\phi_{K_\epsilon}(x) = \epsilon\left(\epsilon^{-1}\phi_K(\epsilon^{-1}x)\right).$$

Iterating, we see that $\phi_{m,K_\epsilon}(x) = \epsilon^{m-1}\phi_{m,K}(\epsilon^{-1}x) = \epsilon^{m-1}G_{m,K}(\epsilon^{-1}x)$. By a change of variables in the integral expression in (4.5.14), it follows that

$$\|K_\epsilon * g - g\|_{L^p,\mathrm{supp}(g)} \leq \epsilon^m\int_{-L/\epsilon}^{L/\epsilon}|G_{m,K}(x)|\,dx\cdot\|g^{(m)}\|_{L^p}. \tag{4.5.15}$$

We can pass to a limit of approximating functions to see that the bound in (4.5.15) also holds for any $g \in W^{m,p}(\mathbb{R})$ of compact support, where L denotes the diameter of the support.

Let $I' = (a-\eta/2, b+\eta/2)$. Since $\rho_I \in W^{m,p}(I)$, we can decompose $\rho_I = g_1 + g_2$ such that g_1 is non-negative, supported in I with $\|g_1^{(m)}\|_{L^p(\mathbb{R})} \leq C(m)\|\rho_I\|_{W^{m,p}(I)}(1+\eta^{-m})$ for some constant $C(m)$ (that depends only on m) and g_2 is non-negative with support contained in $\mathbb{R}\setminus I'$. Therefore, $\rho_I = g_1$ on (a,b) and for almost any $x \in (a,b)$

$$|\rho_I(x) - [K_\epsilon * \mu](x)| \leq \epsilon^{-1}\frac{C_K}{(1+\frac{\eta}{2\epsilon})^{m+1}} + |[K_\epsilon * g_1](x) - g_1(x)|.$$

By the triangle inequality, this implies that

$$\|\rho_I - [K_\epsilon * \mu]\|_{L^p,[a,b]} \leq \frac{C_K(b-a)^{1/p}}{(\epsilon+\eta/2)^{m+1}}\epsilon^m + \int_{-((b-a)+2\eta)/\epsilon}^{((b-a)+2\eta)/\epsilon}|G_{m,K}(x)|\,dx\cdot\|g_1^{(m)}\|_{L^p}\cdot\epsilon^m, \tag{4.5.16}$$

since

$$\left\|\epsilon^{-1}\frac{C_K}{(1+\frac{\eta}{2\epsilon})^{m+1}}\right\|_{L^p,[a,b]} = \frac{C_K(b-a)^{1/p}}{(\epsilon+\eta/2)^{m+1}}\epsilon^m.$$

The bound (4.5.16) then implies (4.5.6).

Finally, (4.5.7) follows from (4.5.5) and (4.5.6) through bounding the integral

$$\int_{-((b-a)+2\eta)/\epsilon}^{((b-a)+2\eta)/\epsilon} |G_{m,K}(x)|\,dx \leq \int_{-((b-a)+2\eta)/\epsilon}^{((b-a)+2\eta)/\epsilon} \frac{C_K}{m!(1+|x|)}dx = \mathcal{O}(\log(1/\epsilon)).$$

If $|K(x)|(1+|x|)^{m+1+\delta}$ is bounded for $\delta > 0$, then the same argument used for (4.5.9) and (4.5.10) implies that $|G_{m,K}(x)(1+|x|)^{1+\delta}|$ is also bounded and hence $G_{m,K} \in L^1(\mathbb{R})$. The rate (4.5.8) follows since

$$\lim_{\epsilon \downarrow 0} \int_{-((b-a)+2\eta)/\epsilon}^{((b-a)+2\eta)/\epsilon} |G_{m,K}(x)|\,dx < \infty$$

and the other terms are $\mathcal{O}(\epsilon^m)$. □

As well as increasing the rate of convergence for computing Radon–Nikodym derivatives, high-order kernels increase the rate of convergence for computing the functional calculus. However, no regularity assumptions on μ are needed. Instead, one can apply Fubini's theorem and (strictly speaking the proofs of) Theorems 4.5.2 and 4.5.5 to obtain high-order convergence through regularity of the function F. For example, if K is an mth order kernel and $F \in C^{n,\alpha}(\mathbb{R})$, then for any probability measure μ, regardless of the regularity of μ, we have

$$\left| \int F(x) d\mu(x) - \int F(x) d\left[K_\epsilon * \mu\right](x) \right| = \mathcal{O}(\epsilon^{n+\alpha}) + \mathcal{O}(\epsilon^m \log(\epsilon^{-1})).$$

As expected, when F is analytic, we can do even better.

4.5.3 Constructing rational kernels

Theorems 4.5.2 and 4.5.5 show that the convolution with the Poisson kernel has a pointwise and L^p local rate of convergence of $\mathcal{O}(\epsilon \log(\epsilon^{-1}))$ for regular enough measures. In designing a kernel suitable for numerical computations, we note that the results of §4.2 allow the computation of $R(z, A)x$ with error control for any $z \notin \mathbb{R}$ and $(A, x) \in \Omega_{f,\alpha,\beta}$ assuming that we have explicit bounds on $\|(I - P_n)AP_n\|$ and $\|P_n x - x\|$. To avoid compounding errors (and requiring larger n to solve the relevant systems), it is beneficial to avoid evaluating squares and higher powers of the resolvent. This leads us to kernels of the form

$$K(u) = \frac{1}{2\pi i} \sum_{j=1}^{n_1} \frac{\alpha_j}{u - a_j} - \frac{1}{2\pi i} \sum_{j=1}^{n_2} \frac{\beta_j}{u - b_j}, \tag{4.5.17}$$

where $a_1, ..., a_{n_1}$ are distinct points in the upper half-plane and $b_1, ..., b_{n_2}$ are distinct points in the lower half-plane. We can then compute the convolution $\mu_{x,y}^A * K_\epsilon$ with error control through the formula

$$\mu_{x,y}^A * K_\epsilon(u) = \frac{-1}{2\pi i} \left[\sum_{j=1}^{n_1} \alpha_j \langle R(u - \epsilon a_j, A)x, y\rangle - \sum_{j=1}^{n_2} \beta_j \langle R(u - \epsilon b_j, A)x, y\rangle \right]. \tag{4.5.18}$$

By considering the Fourier transform of K at zero frequency and matching the left and right derivatives of the Fourier transform, a straightforward calculation shows that the first $m - 1$ moments of K exist and are zero (excluding the 0th order which must be 1 to achieve convergence), if and only if

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_{n_1} \\ a_1^2 & a_2^2 & \cdots & a_{n_1}^2 \\ \vdots & \vdots & & \vdots \\ a_1^{m-1} & a_2^{m-1} & \cdots & a_{n_1}^{m-1} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{n_1} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{4.5.19}$$

m	$\pi K(u)\prod_{j=1}^m(u-a_j)(u-\overline{a_j})$	$\{\alpha_1,\dots,\alpha_{\lceil m/2\rceil}\}$
2	$\frac{20}{9}$	$\left\{\frac{1+3i}{2}\right\}$
3	$-\frac{5}{4}u^2+\frac{65}{16}$	$\{-2+i,5\}$
4	$-\frac{3536}{625}u^2+\frac{21216}{3125}$	$\left\{\frac{-39-65i}{24},\frac{17+85i}{8}\right\}$
5	$\frac{130}{81}u^4-\frac{12350}{729}u^2+\frac{70720}{6561}$	$\left\{\frac{15-10i}{4},\frac{-39+13i}{2},\frac{65}{2}\right\}$
6	$\frac{1287600}{117649}u^4-\frac{34336000}{823543}u^2+\frac{667835200}{40353607}$	$\left\{\frac{725+1015i}{192},\frac{-2775-6475i}{192},\frac{1073+7511i}{96}\right\}$

Table 4.1: The numerators and residues of the first six rational kernels with equispaced poles (see (4.5.21)). We give the first $\lceil m/2\rceil$ residues because the others follow by the symmetry $\alpha_{m+1-j}=\overline{\alpha_j}$.

with a similar system holding for the β_j and b_j. By considering the 2nd to (n_1+1)th rows, this (transposed) Vandermonde system cannot have a solution if $n_1 < m$. We therefore set $n_1 = n_2 = m$. In the case that $x = y$, a further numerical saving can be made by letting $b_j = \overline{a_j}$ and noting that in this case

$$\mu_{x,x}^A * K_\epsilon(u) = \frac{-1}{\pi}\text{Im}\left[\sum_{j=1}^m \alpha_j \langle R(u-\epsilon a_j, A)x, x\rangle\right], \tag{4.5.20}$$

meaning that we only need m resolvent evaluations per point of evaluation.

The location of the poles in the upper half-plane is entirely flexible. As a natural extension of the Poisson kernel, whose two poles are at $\pm i$, we consider the family of mth order kernels with equispaced poles in the upper and lower half-planes given by

$$a_j = \frac{2j}{m+1} - 1 + i, \qquad b_j = \overline{a_j}, \qquad 1 \leq j \leq m. \tag{4.5.21}$$

Empirically, the choice in (4.5.21) performed slightly better than other natural choices such as Chebyshev points with an offset $+i$ or rotated roots of unity. The ill-conditioning of the Vandermonde system does not play a role for the values of m used (typically at most $m = 10$). Moreover, equispaced poles are particularly useful when one wishes to sample the smoothed measure $K_\epsilon * \mu_{x,y}^A$ over an interval, since samples of the resolvent can be reused for different points in the interval. The first ten kernels are plotted in Figure 4.4 (left) and the first six are explicitly written down in Table 4.1.

4.5.4 Jacobi operator examples

Let J be a Jacobi matrix

$$J = \begin{pmatrix} b_1 & a_1 & & \\ a_1 & b_2 & a_2 & \\ & a_2 & b_3 & \ddots \\ & & \ddots & \ddots \end{pmatrix}$$

with $a_j, b_j \in \mathbb{R}$ and $a_j > 0$. Under suitable conditions, the probability measure $\mu_J := \mu_{e_1,e_1}^J$ is exactly the probability measure associated with the orthonormal polynomials defined by

$$P_{-1}(x) = 0, \quad P_0(x) = 1, \quad xP_k(x) = a_{k+1}P_{k+1}(x) + b_{k+1}P_k(x) + a_kP_{k-1}(x).$$

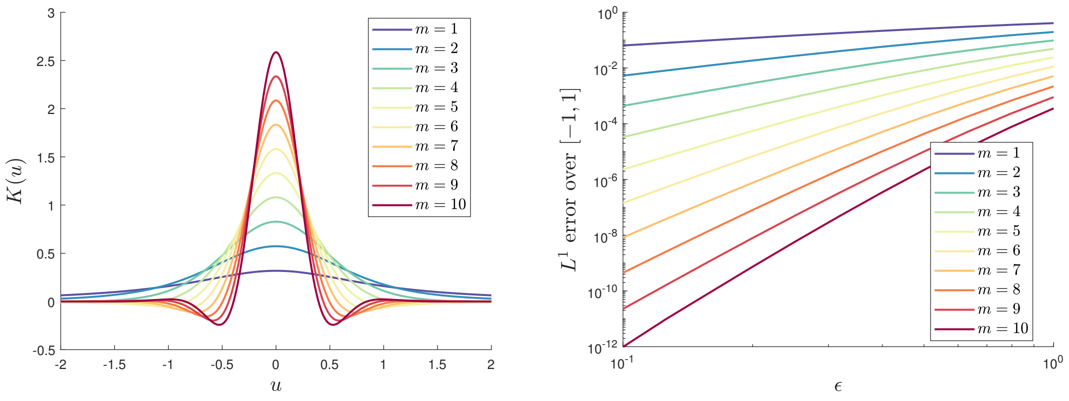


Figure 4.4: Left: Kernels used for convolution. Right: Convergence for Gaussian measure.

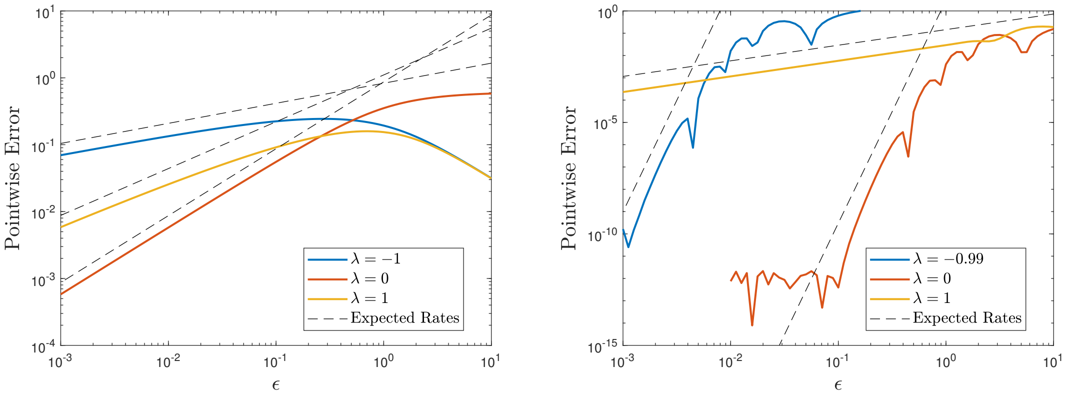


Figure 4.5: Left: Pointwise errors for $\lambda = -1, 0, 1$ for $m = 1$ and $\alpha = 0.7$, $\beta = 0.3$. Right: Pointwise errors for $\lambda = -0.99, 0, 1$ for $m = 10$ and $\alpha = 0.7$, $\beta = -0.3$.

As a simple example,consider $a_k = \sqrt{k/2}$ and $b_k = 0$, corresponding to the famous measure $d\mu_J = \exp(-\lambda^2)/\sqrt{\pi}d\lambda$, which induces the Hermite polynomials. We have shown the convergence (measured via the L^1 error over $[-1, 1]$) of our method using §4.2, for different values of m in terms of the distance of the poles to the real line ($= \epsilon$) in Figure 4.4 (right). We can clearly see the convergence rates $\mathcal{O}(\epsilon^m)$ (up to logarithmic factors)[4] from Theorems 4.5.2 and 4.5.5.

As a second example, consider the Jacobi polynomials defined for $\alpha, \beta > -1$ which have

$$a_k = 2\sqrt{\frac{k(k+\alpha)(k+\beta)(k+\alpha+\beta)}{(2k+\alpha+\beta-1)(2k+\alpha+\beta)^2(2k+\alpha+\beta+1)}}, \quad b_k = \frac{\beta^2-\alpha^2}{(2k+\alpha+\beta)(2k-2+\alpha+\beta)}$$

and measure on the interval $[-1, 1]$ given by

$$d\mu_J = \frac{(1-\lambda)^\alpha(1+\lambda)^\beta}{N(\alpha,\beta)}d\lambda = f_{\alpha,\beta}(\lambda)d\lambda,$$

where $N(\alpha, \beta)$ is a normalising constant, ensuring the measure is a probability measure. Figure 4.5 (left) shows the pointwise convergence at $\lambda = -1, 0, 1$ for $m = 1$ and $\alpha = 0.7$, $\beta = 0.3$. The approximation converges at the expected rates (corresponding to the relevant Hölder regularity) from Theorem 4.5.2. Figure 4.5 (right) shows a similar plot for $\lambda = -0.99, 0, 1$ for $m = 10$ and $\alpha = 0.7$, $\beta = -0.3$. The rate of

[4] There are no logarithmic factors when m is even. However, an extra $\log(\epsilon^{-1})$ factor appears when m is odd (owing to the non-integrability of $u^m K(u)$). More generally, by analysing the solution of the system (4.5.19), the logarithmic factors disappear precisely when $\prod_{j=1}^{m} a_j = \prod_{j=1}^{m} b_j$.

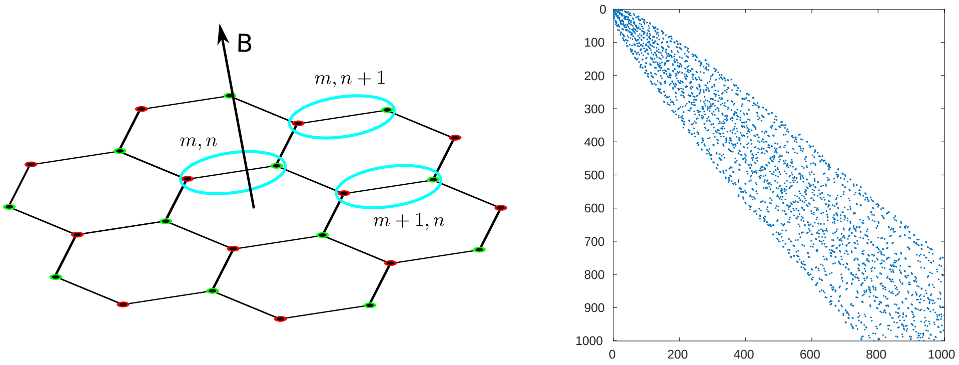


Figure 4.6: Left: Honeycomb structure of graphene as a bipartite graph. The spinor structure is shown via circled lattice vertices. The arrow shows the perpendicular magnetic field **B**. Right: Sparsity structure of the first $10^3 \times 10^3$ block of the infinite matrix, and the corresponding growing local bandwidth.

convergence is increased to order 10 for $\lambda = -0.99$ and $\lambda = 0$ where the measure is locally smooth, but remains order α at $\lambda = 1$. The error at $\lambda = -0.99$ is larger than at $\lambda = 0$ due to being much nearer the singularity at -1, which corresponds to a smaller η in Theorem 4.5.2.

4.6 Numerical Examples

4.6.1 Magneto-graphene Schrödinger operator

We apply the method to a magnetic tight-binding model of graphene, which involves a discrete graph operator [AEG14]. Graphene is a two-dimensional material with carbon atoms situated at the vertices of a honeycomb lattice (Figure 4.6), whose unusual properties are studied in condensed-matter physics [NGP$^+$09, Nov11]. Magnetic properties of graphene are well-studied and include experimental observations of the quantum Hall effect and Hofstadter's butterfly [PGY$^+$13], and twistronics [Cha19, LSY$^+$19].

A honeycomb lattice can be decomposed into two bipartite sub-lattices (shown via the red and green dots in Figure 4.6 (left)) and thus the wave function of an electron can be modelled as the spinor [AEG14]

$$\psi_{m,n} = (\psi^{[1]}_{m,n}, \psi^{[2]}_{m,n})^T \in \mathbb{C}^2, \qquad \psi = (\psi_{m,n}) \in l^2(\mathbb{Z}^2; \mathbb{C}^2) \cong \ell^2(\mathbb{N}).$$

Here, $(m,n) \in \mathbb{Z}^2$ labels a position on the sub-lattices and $\ell^2(\mathbb{Z}^2;\mathbb{C}^2)$ denotes the space of square summable $\mathbb{C}^2$-valued sequences indexed by $\mathbb{Z}^2$. To define the Hamiltonian, consider the following three magnetic hopping operators $T_1, T_2, T_3 : \ell^2(\mathbb{Z}^2;\mathbb{C}^2) \rightarrow \ell^2(\mathbb{Z}^2;\mathbb{C}^2)$ for a given magnetic flux per unit cell Φ (in dimensionless units):

$$(T_1\psi)_{m,n} = \begin{pmatrix} \psi^{[2]}_{m,n} \\ \psi^{[1]}_{m,n} \end{pmatrix}, \quad (T_2\psi)_{m,n} = \begin{pmatrix} \psi^{[2]}_{m+1,n} \\ \psi^{[1]}_{m-1,n} \end{pmatrix}, \quad (T_3\psi)_{m,n} = \begin{pmatrix} e^{-2\pi i \Phi m}\psi^{[2]}_{m,n+1} \\ e^{2\pi i \Phi m}\psi^{[1]}_{m,n-1} \end{pmatrix}.$$

After a suitable gauge transformation, the free Hamiltonian can be expressed as $H_0 = T_1 + T_2 + T_3$ and has $\mathrm{Sp}(H_0) \subset [-3,3]$. A suitable ordering of lattice points leads to a sparse discretisation of H_0, where the kth row contains $\mathcal{O}(\sqrt{k})$ non-zero entries (see Figure 4.6 (right)). Therefore, for an approximation using N basis sites, the action of the resolvent can be computed in $\mathcal{O}(N^{3/2})$ operations [TBI97].

Figure 4.7 shows how the spectral measure of H_0, taken with respect to the vector e_1 (the labelling does not matter due to the translational invariance of the lattice), varies with Φ. For $\Phi \in \mathbb{Q}$, the spectrum

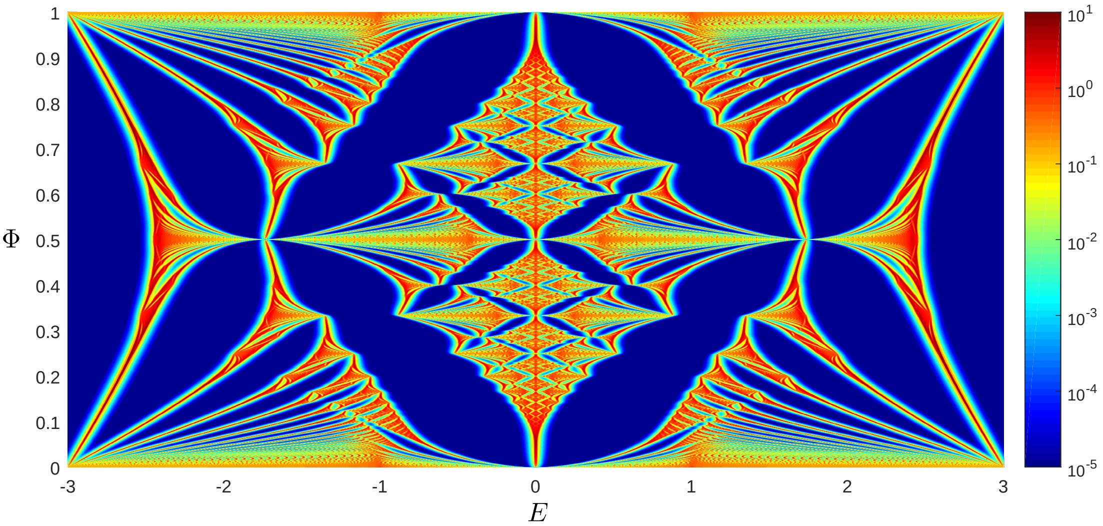


Figure 4.7: Radon–Nikodym derivative $\rho_{e_1,e_1}^{H_0}$ (`log10` scale) of the measure for various magnetic field strengths Φ. The axis label E (energy) stands for the spectral parameter. The Radon–Nikodym derivative is computed to high precision using $\epsilon = 0.01$ and a fourth-order kernel with poles corresponding to (4.5.21). The spectrum is fractal for irrational Φ, which is approximated by rational Φ. The small gaps in the spectrum are clearly visible (corresponding to the blue shaded regions) and the logarithmic scale shows the sharpness of the approximation to $\rho_{e_1,e_1}^{H_0}$ (which vanishes in these gaps).

is absolutely continuous, and hence we have plotted the Radon–Nikodym derivative of the measure $\mu_{e_1,e_1}^{H_0}$. The calculations, performed with a fourth-order kernel and $\epsilon = 0.01$, show a sharp Hofstadter-type butterfly, but now with the additional information of the spectral measure.

Figure 4.8 (left) shows an approximation of $\rho_{e_1,e_1}^{H_0}$ when $\Phi = 1/4$ using a fourth-order kernel and $\epsilon = 0.01$. We also show, as shaded vertical strips, the output of the algorithm in Chapter 3 [CRH19] which computes the spectrum with error control (we used an error bound of 10^{-3}) and without spectral pollution.[5] The support of $K_\epsilon * \mu_{e_1,e_1}^{H_0}$ is the whole real line due to the non-compact support of the kernel K. However, if $\lambda \notin \mathrm{Sp}(H_0)$, then $|[K_\epsilon * \mu_{e_1,e_1}^{H_0}](\lambda)| \leq C_K \epsilon^m (\epsilon + \mathrm{dist}(\lambda, \mathrm{Sp}(H_0)))^{-(m+1)}$, where C_K is the constant in (4.5.2) and m is the order of the kernel, so $|[K_\epsilon * \mu_{e_1,e_1}^{H_0}](\lambda)|$ decays rapidly off of the spectrum. We also consider a multiplication operator (potential) perturbation, modeling a defect, of the form

$$V(\mathbf{x}) = \frac{\cos(\|\mathbf{x}\|_2 \pi)}{(\|\mathbf{x}\|_2 + 1)^2}, \tag{4.6.1}$$

where $\mathbf{x}$ denotes the position of a vertex normalised so each edge has length 1. The perturbed operator is then $H_0 + V$. Since the perturbation is trace class, the absolutely continuous part of the spectrum remains the same (though the measure changes) and the potential induces additional eigenvalues (see Figure 4.8 (right)). Again, we see that $|[K_\epsilon * \mu_{e_1,e_1}^{H_0+V}](\lambda)|$ decays rapidly off of the spectrum. In particular, the measure is not corrupted by spikes in the gaps in the essential spectrum or similar artefacts caused by spectral pollution.

4.6.2 Hunting eigenvalues of the Dirac operator

In this example, we show how the results of this chapter can be used as an effective tool to find eigenvalues in gaps of the essential spectrum, whilst avoiding spectral pollution. This example also demonstrates that the methods of this chapter apply to partial differential operators.

[5] With a non-periodic potential (4.6.1), this is a highly non-trivial problem since finite truncation methods typically suffer from spectral pollution inside the convex hull in the essential spectrum.

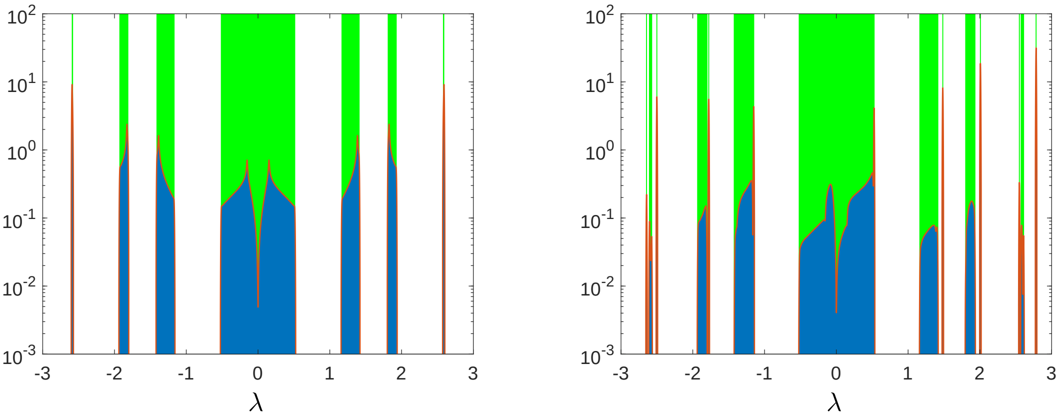


Figure 4.8: Left: Smoothed measure with no potential. We show the algorithm from Chapter 3 as shaded strips (green) for comparison. Right: The same computation but with the added potential in (4.6.1). The additional eigenvalues correspond to spikes in the smoothed measure.

We consider the Dirac operator (defined below) which often has discrete spectrum in the interval $(-1,1)$. This interval forms a gap of the essential spectrum. It follows that standard finite section methods used to compute the discrete spectrum will suffer from spectral pollution within the gap $(-1,1)$ - i.e. there exist accumulation points of the approximations which do not belong to the spectrum. There is a rich literature on how to avoid this [DG81, Kut84, Tal86, Kut97, STY$^+$04, LS14]. The majority of existing approaches work for certain classes of potentials and avoid spectral pollution on particular subsets of $(-1,1)$. Even for simple Coulomb-type potentials, spectral pollution can be a difficult issue to overcome, and computations typically achieve a few digits of precision for the ground state and a handful of the first few excited states. A popular approach is the so-called kinetic balance condition, which does not always work for Coulomb potentials [SH84, DFJ90, LS09]. Our approach does not suffer from spectral pollution and can compute the first thousand eigenvalues to near machine precision accuracy. The problem of spectral pollution is discussed further in §7.1 and Chapter 7.

The Dirac operator acts on $L^2(\mathbb{R}^3;\mathbb{C}^4)$ as [ELS08] $D_0 := -i\sum_{k=1}^3 \alpha_k\partial_k + \beta$, where

$$\alpha_j = \begin{pmatrix} 0 & \sigma_j \\ \sigma_j & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} I_{\mathbb{C}^2} & 0 \\ 0 & -I_{\mathbb{C}^2} \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

are the so-called Pauli matrices [Tha92]. For simplicity we have chosen units corresponding to $m = c = \hbar = 1$. The spectrum of D_0 is equal to $(-\infty,-1]\cup[1,\infty)$ and an important problem in quantum chemistry/physics is the computation of the spectrum of

$$D_V := D_0 + V,$$

where V is some (real-valued) potential. The addition of the potential can cause the appearance of eigenvalues in the gap $(-1,1)$, where, roughly speaking, positive eigenvalues correspond to bound states of a relativistic quantum electron in the external field V and negative eigenvalues correspond to bound states of a positron, the anti-particle of the electron. If V satisfies suitable conditions (precisely which conditions is a broad topic - see [Tha92] for many potentials of physical interest), then D_V is self-adjoint with essential spectrum $\mathrm{Sp}(D_0) = (-\infty,-1]\cup[1,\infty)$.

We consider radially symmetric potentials $V = V(r)I_{\mathbb{C}^4}$. In this case, we can decompose our Hilbert space as a sum of two-dimensional angular momentum subspaces $\mathcal{H}_{m_j,k_j}$ [Tha92] for $m_j \in \{-j,...,j\}$

and $k_j \in \{\pm(j+1/2)\}$ for $j \in \{(2l+1)/2 : l \in \mathbb{Z}_{\geq 0}\}$. The operator $D_V|_{C_0^\infty(0,\infty)\otimes\mathcal{H}_{m_j,k_j}}$ is then unitarily equivalent to

$$D_V^{k_j} := \begin{pmatrix} 1+V(r) & -\frac{d}{dr}+\frac{k_j}{r} \\ \frac{d}{dr}+\frac{k_j}{r} & -1+V(r) \end{pmatrix}.$$

Again, under suitable conditions on the potential V, we have that $D_V^{k_j}|_{C_0^\infty(0,\infty)^2}$ are essentially self-adjoint and the full spectrum and discrete spectrum can be recovered from

$$\mathrm{Sp}(D_V) = \mathrm{cl}\left(\bigcup \mathrm{Sp}\left(D_V^{k_j}\right)\right), \quad \mathrm{Sp}_d(D_V) = \bigcup \mathrm{Sp}_d\left(D_V^{k_j}\right).$$

We treat the case of $k_j = -1$ for simplicity and, with an abuse of notation, write $D_V^{k_j}$ as simply D_V.

To compute the spectral measure of D_V, we must be able to compute the resolvent and the corresponding inner products to compute the scalar measures $\mu_{f,g}^{D_V}$. This involves solving near singular PDEs corresponding to the computation of the resolvent near the real axis. Letting r denote the variable on the half-line, we first map to the interval $(-1,1)$ via

$$x = \frac{r-L}{r+L}, \qquad r = L\left(\frac{1+x}{1-x}\right).$$

The resolvent then gives rise to a singular variable coefficient ODE via the relations

$$\frac{d}{dr} = \frac{(1-x)^2}{2L}\frac{d}{dx}, \quad \frac{1}{r} = \frac{1}{L}\frac{1-x}{1+x}.$$

To solve these ODEs, we use the ultraspherical method [OT13], which is based on representations of the solution in different ultraspherical polynomial bases. A full discussion of the ultraspherical method is beyond the scope of this course. For us, the key point is that the ultraspherical method leads to a sparse and well-conditioned linear system that can be solved in linear time up to log factors (and will compute the correct solution bounded at infinity and zero). To compute inner products, we map the inner product over the half-line to the interval (with a suitable Jacobian weight) and then use Clenshaw–Curtis quadrature. In the method, L is a scaling parameter, which for our experiments we set to $L = 10$.

As mentioned above, the Dirac operator poses a serious challenge in terms of spectral computations, owing to the gap in the essential spectrum. Let $f \in L^2(0,\infty)\oplus L^2(0,\infty)$ and define $\nu_f^\epsilon(\lambda) := \epsilon\pi\langle K_H(\lambda + i\epsilon; D_V f), f\rangle$. Then, denoting the orthogonal projection onto the eigenspace corresponding to eigenvalue E_j by P_{E_j}, we have[6]

$$\lim_{\epsilon\downarrow 0}\nu_f^\epsilon(\lambda) = \begin{cases} \|P_{E_j}f\|^2, & \text{if } \lambda = E_j \\ 0, & \text{otherwise} \end{cases}.$$

If f is not orthogonal to any of the eigenspaces, we expect the positions of the peaks of ν_f^ϵ to correspond to the eigenvalues. To test this, we consider the case of the Coulombic potential

$$V(r) = \frac{\gamma}{r}, \quad -\sqrt{3}/2 < \gamma < 0$$

for which the eigenvalues are known analytically and given by

$$E_j = \left(1 + \frac{\gamma^2}{\left(j+\sqrt{1-\gamma^2}\right)^2}\right)^{-1/2}, \quad j \in \mathbb{Z}_{\geq 0}.$$

[6] One can show that if there is no singular continuous spectra in a neighbourhood of λ and if λ is not an accumulation point of the point spectrum then the difference between the values for positive ϵ and the limit are $\mathcal{O}(\epsilon)$.

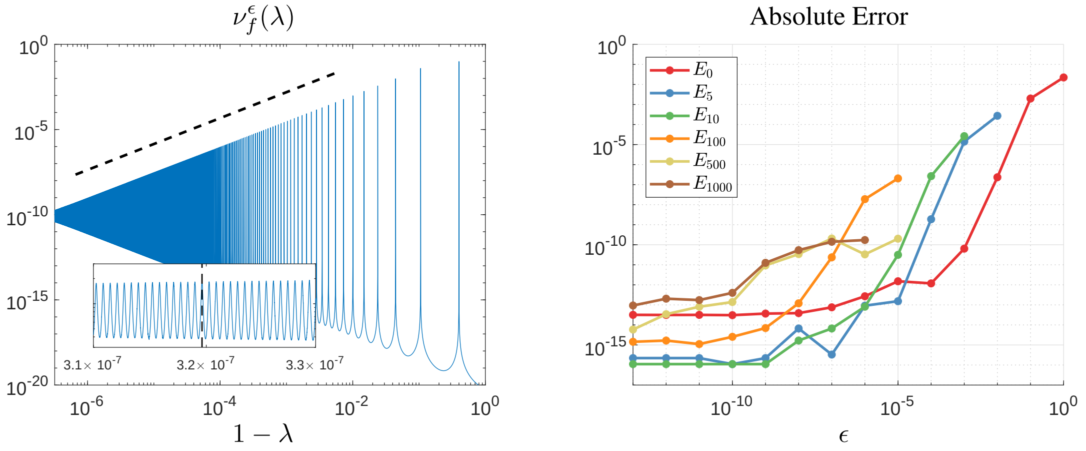


Figure 4.9: Left: The function $\nu_f^\epsilon(x)$ for λ near 1. We have plotted the function against $1-\lambda$ to aid visability of the accumulation at $\lambda = 1$. The sloped dashed line shows the algebraic decay of $\|P_{E_j} f\|^2$ ($\mathcal{O}(j^{-3})$). The magnified region shows the extreme clustering, where the vertical dashed line corresponding to E_{1000}. Right: The absolute error in the computed eigenvalues E_j for $j = 0, 5, 10, 100, 500, 1000$ as $\epsilon \downarrow 0$.

The eigenvalues accumulate at 1, meaning that, even ignoring the problem of spectral pollution, they are very hard to compute for large j.

Figure 4.9 (left) shows ν_f^ϵ with $\epsilon = 10^{-10}$, $f(r) = (\sqrt{2}re^{-r}, \sqrt{2}re^{-r})$, and $\gamma = -0.8$. One can robustly compute ν_f^ϵ for a fixed $\epsilon > 0$ by using the ultraspherical method and adaptively selecting the discretisation size. For $\epsilon = 10^{-10}$, we can accurately compute $E_1, \ldots, E_{1000}$ by the location of the local maxima of ν_f^ϵ. We can obtain a coarse estimate first using a few λ values and then refine our search as we converge to an eigenvalue. Moreover, the size of the peaks correspond to $\|P_{E_j} f\|^2$, and the figure shows that these decrease at an algebraic rate as $j \to \infty$. If one is not satisfied with the accuracy of the computed eigenvalues, then one can decrease ϵ at the expense of an increased computational cost. In Figure 4.9 (right), we show the absolute error in the computed eigenvalues E_j as $\epsilon \downarrow 0$. We can resolve hundreds of eigenvalues, even when highly clustered, to an accuracy of essentially machine precision.

4.6.3 Matlab demo for radial Schrödinger operator

Consider the radial Schrödinger operator with a Hellmann potential and angular momentum ℓ,

$$\mathcal{L}u(r) = -\frac{d^2u(r)}{dr^2} + \left(\frac{\ell(\ell+1)}{r^2} + \frac{1}{r}(e^{-r} - 1)\right) u(r), \qquad r > 0. \tag{4.6.2}$$

The spectral properties of $\mathcal{L}$ are of interest in quantum chemistry, where the Hellman potential models atomic and molecular ionisation processes. Ionisation rates and related transition probabilities are usually studied by computing bound and resonant states of $\mathcal{L}$; however, we compute this information directly from the spectral measure.

For example, if $f(r) = Ce^{-(r-r_0)^2}$ (where C is chosen so that $\|f\|_{L^2(\mathbb{R}_+)} = 1$) is the radial component of the wave function of an electron interacting with an atomic core via the Hellmann potential in (4.6.2), then we can calculate the probability that the electron escapes from the atomic core with energy $E \in [a, b]$ (with $0 < a < b$) via

$$\mathbb{P}(a \leq E \leq b) = \mu_{f,f}^{\mathcal{L}}([a, b]) \approx \int_a^b [K_\epsilon * \mu_{f,f}^{\mathcal{L}}](y)\, dy, \qquad \epsilon \ll 1. \tag{4.6.3}$$

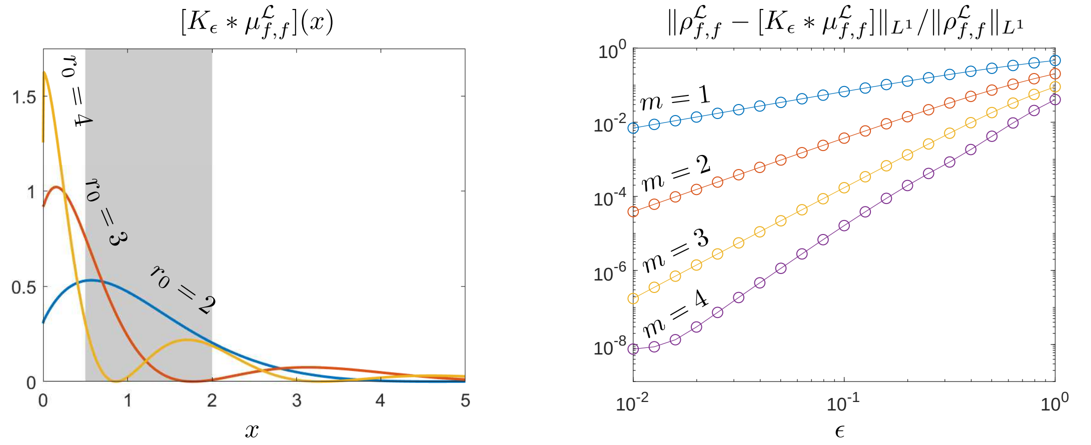


Figure 4.10: Left: The smoothed approximation to the density on the absolutely continuous spectrum of $\mathcal{L}$ in (4.6.2), with $f_{r_0}(r) = C_{r_0} e^{-(r-r_0)^2}$, for $r_0 = 2$, $r_0 = 3$, and $r_0 = 4$ (C_{r_0} is a normalisation constant so that $\|f_u\|_{L^2(\mathbb{R}_+)} = 1$). The shaded area under each curve corresponds to $\mathbb{P}(1/2 \leq E \leq 2)$ in (4.6.3) for the particle with wave function $f_{r_0}(r)$. Right: The $L^1((1/2, 2))$ relative error in smoothed measures for the radial Schrödinger operator in (4.6.2). The relative error is computed by comparing with a numerical solution that is resolved to machine precision.

The error in this approximation is bounded via

$$\left| \mu_{f,f}^{\mathcal{L}}([a,b]) - \int_a^b [K_\epsilon * \mu_{f,f}^{\mathcal{L}}](y)\, dy \right| \leq \int_a^b |\rho_{f,f}^{\mathcal{L}}(y) - [K_\epsilon * \mu_{f,f}^{\mathcal{L}}](y)|\, dy = \|\rho_{f,f}^{\mathcal{L}} - K_\epsilon * \mu_{f,f}^{\mathcal{L}}\|_{L^1([a,b])}.$$

We can compute $\mathbb{P}(1/2 \leq E \leq 2)$ for $\ell = 1$ with a few lines of code calling `Specfun`:

```
normf = sqrt(pi/8)*(2-igamma(1/2,8)/gamma(1/2)); % Normalisation
f = @(r) exp(-(r-2).^2)/sqrt(normf);              % Measure wrt f(r)
v = {@(r) 3, @(r) (exp(-r)-1), @(r) 0};           % Radial potential
[xi, wi] = chebpts(50,  [1/2 2]);                 % Quadrature rule
smooth_meas = rsMeas(v, f, xi, 0.01)              % Smoothed measure
ion_prob = wi * smooth_meas;                      % Ionisation prob
```

This makes it easy to explore how the probability of ionisation changes as we adjust the problem parameters. We can explore the effect of changing the angular momentum number, ℓ, or the initial wave function, f (see Figure 4.10 (left)). The L^1 convergence for the approximation to the probabilities in (4.6.3) is shown in Figure 4.10 (right), which agrees with the asymptotic rates implied by Theorem 4.5.5.

Chapter 5

Computing Spectral Type

This chapter, based on [Col21], complements Chapter 4 and classifies the computation of $\mathrm{Sp}_{\mathrm{ac}}(A)$, $\mathrm{Sp}_{\mathrm{sc}}(A)$ and $\mathrm{Sp}_{\mathrm{pp}}(A)$ in the SCI hierarchy. These different sets often characterise different physical properties in quantum mechanics (such as the famous RAGE theorem [Rue69, AG74, Ens78]), where a system is modelled by some Hamiltonian $A \in \Omega_{\mathrm{SA}}$ [CFKS87, Com93, GKP91, Las96]. For example, pure point spectrum implies the absence of ballistic motion for many Schrödinger operators [Sim90].

5.1 Computing Spectral Types as Sets - the Main Result

Define the problem functions $\Xi_{\mathcal{I}}^{\mathbb{C}}(A) = \mathrm{Sp}_{\mathcal{I}}(A)$ for $\mathcal{I} = \mathrm{ac}, \mathrm{sc}$ or pp. Note also that $\mathrm{Sp}_{\mathrm{pp}}(A) = \mathrm{cl}(\mathrm{Sp}_{\mathrm{p}}(A))$, the closure of the set of eigenvalues. Since we are dealing with unbounded operators, we use the Attouch–Wets metric, which we recall for the benefit of the reader,

$$d_{\mathrm{AW}}(C_1, C_2) = \sum_{n=1}^{\infty} 2^{-n} \min\left\{1, \sup_{|x|\leq n} |\mathrm{dist}(x, C_1) - \mathrm{dist}(x, C_2)|\right\},$$

for $C_1, C_2 \in \mathrm{Cl}(\mathbb{C})$, where $\mathrm{Cl}(\mathbb{C})$ denotes the set of closed non-empty subsets of $\mathbb{C}$. When considering bounded A, we let $(\mathcal{M}, d)$ be the set of all non-empty compact subsets of $\mathbb{C}$ provided with the Hausdorff metric $d = d_{\mathrm{H}}$:

$$d_{\mathrm{H}}(X, Y) = \max\left\{\sup_{x\in X}\inf_{y\in Y} d(x, y), \sup_{y\in Y}\inf_{x\in X} d(x, y)\right\},$$

where $d(x, y) = |x - y|$ is the usual Euclidean distance. Recall that for compact sets, the topological notions of convergence according to d_{H} and d_{AW} coincide. To allow the possibility that the spectral sets are empty, we add the empty set to our metric space as a separated point (the space remains metrisable). This simply means that $F_n \to \emptyset$ if and only if $F_n = \emptyset$ eventually.

The main theorem of this chapter is the following:

Theorem 5.1.1. *Given the above set-up (see also §4.1), it holds that*

$$\Delta_2^G \not\ni \{\Xi_{\mathrm{ac}}^{\mathbb{C}}, \Omega_{f,\alpha}, \Lambda_1\} \in \Sigma_2^A, \quad \Delta_2^G \not\ni \{\Xi_{\mathrm{sc}}^{\mathbb{C}}, \Omega_{f,\alpha}, \Lambda_1\} \in \Sigma_3^A, \quad \Delta_2^G \not\ni \{\Xi_{\mathrm{pp}}^{\mathbb{C}}, \Omega_{f,\alpha}, \Lambda_1\} \in \Sigma_2^A.$$

If $f(n) - n \geq \sqrt{2n} + \frac{1}{2}$, then the sharp lower bound $\{\Xi_{\mathrm{sc}}^{\mathbb{C}}, \Omega_{f,0}, \Lambda_1\} \notin \Delta_3^G$ also holds.

5.2 Anderson Localisation and the Fractional Moment Method

One of the tools we will use to prove the lower bounds in Theorem 5.1.1 is the Anderson model. We refer the reader to [CL90, CFKS87, Kir07] for broader surveys of Anderson localisation.

We consider a connected, undirected graph G, such that the degree of each vertex is bounded by some constant C_G and such that the set of vertices $V(G)$ is countably infinite. We assume that $v = v_\omega$ is a random potential, where $\omega = \{v_x\}_{x\in V(G)}$ is a collection of independent identically distributed random variables and the single-site probability distribution has a density $\rho \in L^1(\mathbb{R})$ with $\|\rho\|_1 = 1$ (with respect to the standard Lebesgue measure). For such a potential, a measure of disorder is given by the quantity $\|\rho\|_\infty^{-1}$. The following theorem, proven in [Col20a], generalises the results of [Gra94] to certain finite rank perturbations and more general graphs, and is used in the proof of Theorem 5.1.1.

Theorem 5.2.1 (Anderson Localisation for Perturbed Operator)**.** *There exists a constant $\delta(C_G) > 0$ such that if $\|\rho\|_\infty \leq \delta(C_G)$ and ρ has compact support, then the operator $H_v + W$ has only pure point spectrum with probability 1 for any fixed self-adjoint operator W of the form*

$$W = \sum_{j=1}^{M} \alpha_j \left|x_{m_j}\right\rangle \left\langle x_{n_j}\right| . \tag{5.2.1}$$

5.3 Proof of Theorem 5.1.1

5.3.1 Point spectra

Proof that $\{\Xi_{\mathrm{pp}}^{\mathbb{C}}, \Omega_{f,\alpha}, \Lambda_1\} \notin \Delta_2^G$. To prove this, it is enough to consider bounded Schrödinger operators acting on $l^2(\mathbb{N})$, which are clearly a subclass of $\Omega_{f,0}$ for $f(n) = n + 1$. Suppose for a contradiction that there does exist a sequence of general algorithms, Γ_n, with

$$\lim_{n\to\infty} \Gamma_n(H_v) = \Xi_{\mathrm{pp}}^{\mathbb{C}}(H_v).$$

We will construct a potential v such that $\Gamma_n(H_v)$ does not converge. To do this, choose $\rho = \chi_{[-c,c]}/(2c)$ for some constant c such that the conditions of Theorem 5.2.1 hold. We will use Theorem 5.2.1 and the following well-known facts:

1. If v has compact support then $\mathrm{Sp}_{\mathrm{pp}}(H_v) \cap (0, 4) = \emptyset$ [Rem98], but $[0, 4] \subset \mathrm{Sp}(H_v)$ (the potential acts as a compact perturbation so the essential spectrum is $[0, 4]$).
2. If we are in the setting of Theorem 5.2.1 with $W = 0$ then $\mathrm{Sp}(H_v) = [-c, 4 + c]$ almost surely (see for example [KM82]). If $W \neq 0$ then since compact perturbations preserve the essential spectrum, we still have $[-c, 4 + c] \subset \mathrm{Sp}(H_v + W)$ almost surely.

We will define the potential v inductively as follows. Let v_1 be a potential of the form v_ω (with density ρ) such that $[-c, 4 + c] \subset \mathrm{Sp}(H_{v_1})$ and $\mathrm{Sp}(H_{v_1})$ is pure point. Such a v_1 exists by Theorem 5.2.1 and fact (2) above. Then for large enough n there exists $z_n \in \Gamma_n(H_{v_1})$ such that $|z_n - 2| \leq 1$. Fix n_1 such that this holds. Then $\Gamma_{n_1}(H_{v_1})$ only depends on $\{v_1(j) : j \leq N_1\}$ for some integer N_1 by (i) of Definition 2.1.1. Define the potential v_2 by $v_2(j) = v_1(j)$ for all $j \leq N_1$ and $v_2(j) = 0$ otherwise. Then by fact (1) above $\Gamma_n(H_{v_2}) \cap [1/2, 7/2] = \emptyset$ for large n, say for n_2. But then $\Gamma_{n_2}(H_{v_2})$ only depends on $\{v_2(j) : j \leq N_2\}$ for some integer N_2.

We repeat this process inductively switching between potentials which induce $\Gamma_{n_k}(H_{v_k})\cap[1/2,7/2]=\emptyset$ for k even and potentials which induce $\Gamma_{n_k}(H_{v_k})\cap[1,3]\neq\emptyset$ for k odd. Explicitly, if k is even then define a potential v_{k+1} by $v_{k+1}(j)=v_k(j)$ for all $j\leq N_k$ and $v_{k+1}(j)=v_\omega(j)$ (with the density ρ) otherwise such that $[-c,4+c]\subset\mathrm{Sp}(H_{v_{k+1}})$ and $\mathrm{Sp}(H_{v_{k+1}})$ is pure point. Such a ω exists from Theorem 5.2.1 and fact (2) above applied with the perturbation W to match the potential for $j\leq N_k$. If k is odd then we define v_{k+1} by $v_{k+1}(j)=v_k(j)$ for all $j\leq N_k$ and $v_{k+1}(j)=0$ otherwise. We can then choose n_{k+1} such that the above intersections hold and N_{k+1} such that $\Gamma_{n_{k+1}}(H_{v_{k+1}})$ only depends on $\{v_{k+1}(j):j\leq N_{k+1}\}$. Finally set $v(j)=v_k(j)$ for $j\leq N_k$. It is clear from (iii) of Definition 2.1.1, that $\Gamma_{n_k}(H_v)=\Gamma_{n_k}(H_{v_k})$. But then this implies that $\Gamma_{n_k}(H_v)$ cannot converge, the required contradiction. □

Remark 5.3.1. *The result can be extended to Schrödinger operators on $\mathbb{Z}^d$ or much more general lattices. It can also be extended to Schrödinger operators acting on $L^2(\mathbb{R}^d)$ via Kato's famous theorem regarding potentials decaying faster than $\mathcal{O}(1/|x|)$ (see for example [RS78]) and recent results on Anderson localisation for Bernoulli random variables [BK05].*

We now shift our attention to proving that $\Xi_{\mathrm{pp}}^{\mathbb{C}}$ can be computed using a Σ_2^A tower. The first step is the following technical lemma, whose proof will also be used later when considering $\Xi_{\mathrm{ac}}^{\mathbb{C}}$.

Lemma 5.3.2. *Let $a<b$ with $a,b\in\mathbb{R}$ and consider the decision problem*

$$\Xi_{a,b,\mathrm{pp}}:\Omega_{f,\alpha}\to\{0,1\}$$

$$A\to\begin{cases}1, & \textit{if}\ \mathrm{Sp}_{\mathrm{pp}}(A)\cap[a,b]\neq\emptyset\\ 0, & \textit{otherwise.}\end{cases}$$

Then there exists a height two arithmetical tower Γ_{n_2,n_1} (with evaluation functions Λ_1) for $\Xi_{a,b,\mathrm{pp}}$. Furthermore, the final limit is from below in the sense that $\Gamma_{n_2}(A):=\lim_{n_1\to\infty}\Gamma_{n_2,n_1}(A)\leq\Xi_{a,b,\mathrm{pp}}(A)$.

Proof. Step 1 of the proof of Theorem 4.3.2 yields a height two arithmetical tower $\widehat{\Gamma}^j_{n_2,n_1}(A)$ for the computation of $\mu^A_{e_j,e_j,\mathrm{c}}((a,b))$. Note that the final limit is from above and using the fact that $\mu^A_{e_j,e_j,\mathrm{c}}(\{a,b\})=0$ we obtain a height two tower for $\mu^A_{e_j,e_j,\mathrm{c}}([a,b])$. We can then use the height one tower for $\mu^A_{e_j,e_j}([a,b])$, denoted by $\widetilde{\Gamma}^j_{n_1}(A)$, and define

$$a_{j,n_2,n_1}(A)=\widetilde{\Gamma}^j_{n_1}(A)-\widehat{\Gamma}^j_{n_2,n_1}(A).$$

This provides a height two arithmetical tower for $\mu^A_{e_j,e_j,\mathrm{pp}}([a,b])$ with the final limit from below. Without loss of generality (by taking successive maxima) we can assume that these towers are non-decreasing in n_2. Now set

$$\Upsilon_{n_2,n_1}(A)=\max_{1\leq j\leq n_2}a_{j,n_2,n_1}(A).$$

Then it is clear that the limit $\lim_{n_1\to\infty}\Upsilon_{n_2,n_1}(A)=\Upsilon_{n_2}(A)$ exists. Furthermore, the monotonicity of $a_{j,n_2,n_1}(A)$ in n_2 implies that

$$\lim_{n_2\to\infty}\Upsilon_{n_2}(A)=\sup_{n\in\mathbb{N}}\mu^A_{e_n,e_n,\mathrm{pp}}([a,b]),$$

with monotonic convergence from below. This limiting value is zero if $\Xi_{a,b,\mathrm{pp}}(A)=0$, otherwise it is a positive finite number.

To convert this to a height two tower for the decision problem $\Xi_{a,b,\mathrm{pp}}$, that maps to the discrete space $\{0,1\}$, we use the following trick. Consider the intervals $J_1^{n_2} = [0,1/n_2]$, and $J_2^{n_2} = [2/n_2,\infty)$. Let $k(n_2,n_1) \leq n_1$ be maximal such that $\Upsilon_{n_2,n_1}(A) \in J_1^{n_2} \cup J_2^{n_2}$. If no such k exists or $\Upsilon_{n_2,k}(A) \in J_1^{n_2}$ then set $\Gamma_{n_2,n_1}(A) = 0$. Otherwise set $\Gamma_{n_2,n_1}(A) = 1$. These can be computed using finitely many arithmetic operations and comparisons using Λ_1. The point of the intervals $J_1^{n_2}$ and $J_2^{n_2}$ is that we can show $\lim_{n_1\to\infty}\Gamma_{n_2,n_1}(A) = \Gamma_{n_2}(A)$ exists. This is because $\lim_{n_1\to\infty}\Upsilon_{n_2,n_1}(A) = \Upsilon_{n_2}(A)$ exists and hence we cannot oscillate infinitely often between the separated intervals $J_1^{n_2}$ and $J_2^{n_2}$. Now suppose that $\Xi_{a,b,\mathrm{pp}}(A) = 0$, then $\lim_{n_1\to\infty}\widehat{\Gamma}_{n_2,n_1}(A) = 0$ and hence $\lim_{n_1\to\infty}\Gamma_{n_2,n_1}(A) = 0$ for all n_2. Now suppose that $\Xi_{a,b,\mathrm{pp}}(A) = 1$, then for large enough n_2 we must have that $\Upsilon_{n_2}(A) > 2/n_2$ and hence $\Gamma_{n_2}(A) = 1$. Together, these prove the convergence and that $\Gamma_{n_2}(A) \leq \Xi_{a,b,\mathrm{pp}}(A)$. □

Proof that $\{\Xi_{\mathrm{pp}}^{\mathbb{C}},\Omega_{f,\alpha},\Lambda_1\} \in \Sigma_2^A$. **Step 1:** Construction of height two tower. To construct a height two arithmetical tower for $\Xi_{\mathrm{pp}}^{\mathbb{C}}$ we will use Lemma 5.3.2 repeatedly. Let $\widehat{\Gamma}_{n_2,n_1}(\cdot,I)$ denote the height two tower constructed in the proof of Lemma 5.3.2 for the closed interval I ($I=[a,b]$), where without loss of generality by taking successive maxima in n_2, we can assume that this tower is non-decreasing in n_2 (this is where we use convergence from below in the final limit in the statement of the lemma). For a given n_1 and n_2, we construct $\Gamma_{n_2,n_1}(A)$ as follows (we will use some basic terminology from graph theory).

Define the intervals $I_{n_2,n_1,j}^0 = [j,j+1]$ for $j = -n_2,...,n_2-1$ so that these form a cover of the interval $[-n_2,n_2]$. Now suppose that $I_{n_2,n_1,j}^k$ are defined for $j = 1,...,r_k(n_2,n_1,A)$. Compute each $\widehat{\Gamma}_{n_2,n_1}(A,I_{n_2,n_1,j}^k)$ and if this is 1, bisect $I_{n_2,n_1,j}^k$ via its midpoint into two equal halves consisting of closed intervals. We then take all these bisected intervals and label them as $I_{n_2,n_1,j}^{k+1}$ for $j = 1,...,r_{k+1}(n_2,n_1,k,A)$. This is repeated until we have no further bisections or the intervals $I_{n_2,n_1,j}^{n_2}$ have been computed. By adding the interval $[-n_2,n_2]$ as a root with children $I_{n_2,n_1,j}^0$, this creates a finite binary tree structure where a non-root interval I is a parent of two intervals precisely if those two intervals are formed from its bisection and $\widehat{\Gamma}_{n_2,n_1}(A,I) = 1$. We then prune this tree by discarding all leaves I which have $\widehat{\Gamma}_{n_2,n_1}(A,I) = 0$ to form the tree $\mathcal{T}_{n_2,n_1}(A)$. Finally, we let $\Gamma_{n_2,n_1}(A)$ be the union of all the leaves of $\mathcal{T}_{n_2,n_1}(A)$. Clearly this can be computed using finitely many arithmetic operations and comparisons using Λ_1. The construction is shown visually in Figure 5.1.

In the above construction, the number of intervals considered (including those not in the tree $\mathcal{T}_{n_2,n_1}(A)$) for a fixed n_2 is $n_2 2^{n_2+1}+1$ and hence independent of n_1. It follows that $\mathcal{T}_{n_2,n_1}(A)$ and $\Gamma_{n_2,n_1}(A)$ are constant for large n_1 (due to the convergence of the $\widehat{\Gamma}_{n_2,n_1}(A,I)$ in $\{0,1\}$). We denote these limiting values by $\mathcal{T}_{n_2}(A)$ and $\Gamma_{n_2}(A)$ respectively and also denote the corresponding intervals in the construction at the $m-$th level of this limit by $I_{n_2,j}^m$. Note also that if $\Xi_{\mathrm{pp}}^{\mathbb{C}}(A) = \emptyset$ then $\Gamma_{n_2}(A) = \emptyset$.

Now suppose that $z \in \Xi_{\mathrm{pp}}^{\mathbb{C}}(A)$, then there exists a sequence of nested intervals $I_m = I_{n_2,a_{m,n_2}}^m$ containing z for $m = 0,...,n_2$ (where the notation means that these intervals are independent of n_2). Fix m, then for large n_2 we must have that $\widehat{\Gamma}_{n_2}(A,I_j) = 1$ for $j = 1,...,m$. It follows that I_m has a descendent interval $I_{n_2,m}$ contained in $\Gamma_{n_2}(A)$ and hence we must have $\mathrm{dist}(z,\Gamma_{n_2}(A)) \leq 2^{-m}$. Since m was arbitrary it follows that $\mathrm{dist}(z,\Gamma_{n_2}(A))$ converges to 0 as $n_2 \to \infty$.

Conversely, suppose that $z_{m_j} \in \Gamma_{m_j}(A)$ with $m_j \to \infty$, then we must show that all limit points of $\{z_{m_j}\}$ lie in $\Xi_{\mathrm{pp}}^{\mathbb{C}}(A)$. Suppose this were false, then by taking a subsequence if necessary, we can assume that $z_{m_j} \to z$ and $\mathrm{dist}(z_{m_j},\Xi_{\mathrm{pp}}^{\mathbb{C}}(A)) \geq \delta$ for some $\delta > 0$. We claim that it is sufficient to prove that the maximum length of the leaves of $\mathcal{T}_{n_2}(A)$ intersecting a fixed compact subset of $\mathbb{R}$, converges to zero as

$n_2 \to \infty$. Suppose this has been shown, then $z_{m_j} \in I_{m_j}$ for some leaf I_{m_j} of $\mathcal{T}_{m_j}(A)$. It follows that $I_{m_j} \cap \Xi_{\mathrm{pp}}^{\mathbb{C}}(A) \neq \emptyset$ and $|I_{m_j}| \to 0$. But this contradicts z_{m_j} being positively separated from $\Xi_{\mathrm{pp}}^{\mathbb{C}}(A)$.

To prove convergence, we are thus left with proving the claim regarding the lengths of leaves. Suppose this were false, then there exists a compact set $K \subset \mathbb{R}$ and leaves I_j in $\mathcal{T}_{b_j}(A)$ such that the lengths of I_j do not converge to zero and I_j intersect K. By taking subsequences if necessary, we can assume that the lengths of each I_j are constant. Then by the compactness of K and taking subsequences if necessary again, we can assume that each of the I_j are equal to a common interval I. It follows that $\widehat{\Gamma}_{b_j}(A, I) = 1$ but that $\widehat{\Gamma}_{b_j}(A, I_1) = \widehat{\Gamma}_{b_j}(A, I_2) = 0$ since I is a leaf, where I_1 and I_2 form the bisection of I. Taking $b_j \to \infty$, this implies that $I \cap \Xi_{\mathrm{pp}}^{\mathbb{C}}(A) \neq \emptyset$ but $I_1 \cap \Xi_{\mathrm{pp}}^{\mathbb{C}}(A) = I_2 \cap \Xi_{\mathrm{pp}}^{\mathbb{C}}(A) = \emptyset$ which is absurd. Hence we have shown the required contradiction, and proven convergence.

Step 2: Adaptation to achieve a Σ_2^A tower. Let

$$\widetilde{\Gamma}_{n_2,n_1}(A) = \mathrm{Sp}_{\mathrm{pp}}(A) \cup \Gamma_{n_2,n_1}(A), \quad \widetilde{\Gamma}_{n_2}(A) = \lim_{n_1\to\infty} \widetilde{\Gamma}_{n_2,n_1}(A),$$

where we remark that the limit is guaranteed to exist. For $m = 1, ..., n_2$ we define $\hat{\delta}_m(n_1, n_2)$ via the following procedure. If $\Gamma_{n_2,n_1}(A) \cap B_m(0) \neq \emptyset$, then we let $\hat{\delta}_m(n_1, n_2) \leq 1$ be the length of the longest leaf in $\mathcal{T}_{n_2,n_1}(A)$ that intersects $B_{2m}(0)$. If $\Gamma_{n_2,n_1}(A) \cap B_m(0) = \emptyset$, then we let $\hat{\delta}_m(n_1, n_2) = 1$. We then set $\delta_m(n_1, n_2) = \min\{\hat{\delta}_k(n_1, n_2) : m \leq k \leq n_2\}$ and, if $\Gamma_{n_2,n_1}(A) \neq \emptyset$, define

$$E_{n_2,n_1}(A) = 2^{-n_2} + \sum_{m=1}^{n_2} 2^{-m} \cdot \delta_m(n_1, n_2).$$

Otherwise we set $E_{n_2,n_1}(A) = 0$. Note that this can be computed using finitely many arithmetic operations and comparisons. We also define

$$\delta_m(n_2) = \lim_{n_1\to\infty} \delta_m(n_1, n_2), \quad E_{n_2}(A) = \lim_{n_1\to\infty} E_{n_2,n_1}(A),$$

where, again, both limits exist (in fact the sequences are eventually constant) since the finite number of decision problems deciding $\Gamma_{n_2,n_1}(A)$ and $\mathcal{T}_{n_2,n_1}(A)$ are eventually constant.

If $m \in \{1, 2, ..., n_2\}$ and $x \in B_m(0)$, then the closest point to x that lies in $\widetilde{\Gamma}_{n_2}(A)$ either lies in $\mathrm{Sp}_{\mathrm{pp}}(A)$, in which case the inclusion $\mathrm{Sp}_{\mathrm{pp}}(A) \subset \widetilde{\Gamma}_{n_2}(A)$ implies that

$$\min\left\{1, \left|\mathrm{dist}(x, \widetilde{\Gamma}_{n_2}(A)) - \mathrm{dist}(x, \mathrm{Sp}_{\mathrm{pp}}(A))\right|\right\} = 0 \leq \hat{\delta}_m(n_2),$$

or it lies in $\Gamma_{n_2}(A)$. In the latter case, if $\Gamma_{n_2}(A) \cap B_m(0) \neq \emptyset$ then the closest point must also lie in $B_{2m}(0)$ and hence

$$\min\left\{1, \left|\mathrm{dist}(x, \widetilde{\Gamma}_{n_2}(A)) - \mathrm{dist}(x, \mathrm{Sp}_{\mathrm{pp}}(A))\right|\right\} \leq \hat{\delta}_m(n_2),$$

since the final limit of the algorithm from Lemma 5.3.2 is from below. This implies that

$$\min\left\{1, \sup_{|x|\leq m} \left|\mathrm{dist}(x, \widetilde{\Gamma}_{n_2}(A)) - \mathrm{dist}(x, \mathrm{Sp}_{\mathrm{pp}}(A))\right|\right\}$$
$$\leq \min_{m\leq k\leq n_2} \left\{1, \sup_{|x|\leq k} \left|\mathrm{dist}(x, \widetilde{\Gamma}_{n_2}(A)) - \mathrm{dist}(x, \mathrm{Sp}_{\mathrm{pp}}(A))\right|\right\} \leq \delta_m(n_2).$$

It follows that we must have

$$d_{\mathrm{AW}}(\widetilde{\Gamma}_{n_2}(A), \mathrm{Sp}_{\mathrm{pp}}(A)) \leq E_{n_2}(A), \tag{5.3.1}$$

with this bound being trivial in the case that $\Gamma_{n_2}(A) = \emptyset$. Now if m is such that $\Gamma_{n_2}(A) \cap B_m(0) \neq \emptyset$ for large n_2, then since the maximum length of the leaves of $\mathcal{T}_{n_2}(A)$ over any compact set converges to zero,

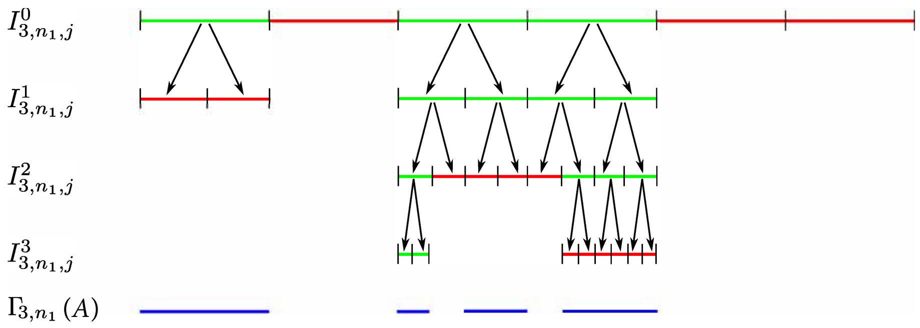


Figure 5.1: Example of tree structure used to compute the point spectrum for $n_2 = 3$. Each tested interval is shown in green ($\widehat{\Gamma}_{n_2,n_1}(A, I) = 1$) or red ($\widehat{\Gamma}_{n_2,n_1}(A, I) = 0$). The arrows show the bisections and the final output is shown in blue.

we must have that $\lim_{n_2\to\infty} \hat{\delta}_m(n_2) = 0$. It follows that if $\mathrm{Sp}_{\mathrm{pp}}(A) \neq \emptyset$ then $\lim_{n_2\to\infty} \delta_m(n_2) = 0$ for each m and hence $\lim_{n_2\to\infty} E_{n_2}(A) = 0$. Clearly this convergence also holds if $\mathrm{Sp}_{\mathrm{pp}}(A) = \emptyset$ since, in this case, $\Gamma_{n_2}(A) = \emptyset$ for large n_2.

To construct a Σ^A_2 tower, it is enough (by taking subsequences) to show that given $\epsilon \in \mathbb{Q}_{>0}$, we can choose $n_2(\epsilon, n_1) \geq \epsilon^{-1}$ such that $\lim_{n_1\to\infty} n_2(\epsilon, n_1) = n^\epsilon_2 \in \mathbb{N}$ exists and

$$d_{\mathrm{AW}}(\widetilde{\Gamma}_{n^\epsilon_2}(A), \mathrm{Sp}_{\mathrm{pp}}(A)) \leq \epsilon.$$

To do this, fix ϵ and consider $\mathcal{S}(\epsilon, n_1) = \mathbb{N} \cap [\epsilon^{-1}, n_1]$. If $n_1 < \epsilon^{-1}$ then set $n_2(\epsilon, n_1) = \lceil \epsilon^{-1} \rceil$. Otherwise, let $\mathcal{S}'(\epsilon, n_1)$ be the set of all $k \in \mathcal{S}(\epsilon, n_1)$ such $E_{k,n_1}(A) \leq \epsilon$. If $\mathcal{S}'(\epsilon, n_1) = \emptyset$ then we set $n_2(\epsilon, n_1) = \lceil \epsilon^{-1} \rceil$, otherwise we set $n_2(\epsilon, n_1)$ to be the minimal element of $\mathcal{S}'(\epsilon, n_1)$. For large n_1, since each $E_{n_2,n_1}(A)$ is eventually constant and the $E_{n_2}(A)$ converge to 0, we must have that $\mathcal{S}'(\epsilon, n_1) \neq \emptyset$. In fact, we have that

$$n^\epsilon_2 = \lim_{n_1\to\infty} n_2(\epsilon, n_1) = \min\{k : k \geq \lceil \epsilon^{-1} \rceil, E_k(A) \leq \epsilon\}.$$

The bound (5.3.1) now finishes the proof. □

5.3.2 Absolutely continuous spectra

We will first prove the lower bound and recall the following result which will be crucial for the proof.

Theorem 5.3.3 (Krutikov and Remling [KR01]). *Consider discrete Schrödinger operators acting on $l^2(\mathbb{N})$. Let v be a (real-valued and bounded) potential of the following form:*

$$v(n) = \sum_{j=1}^{\infty} g_j \delta_{n,m_j}, \quad m_{j-1}/m_j \to 0.$$

Then $[0, 4] \subset \mathrm{Sp}_{\mathrm{ess}}(H_0 + v)$ and the following dichotomy holds:

(a) If $\sum_{j\in\mathbb{N}} g^2_j < \infty$ then $H_0 + v$ is purely absolutely continuous on $(0, 4)$.

(b) If $\sum_{j\in\mathbb{N}} g^2_j = \infty$ then $H_0 + v$ is purely singular continuous on $(0, 4)$.

To prove the lower bound (that one limit will not suffice) our strategy will be to reduce a certain decision problem to the computation of $\Xi^{\mathbb{C}}_{\mathrm{ac}}$. Let $(\mathcal{M}', d')$ be the discrete space $\{0, 1\}$, let Ω' denote the collection of all infinite sequence $\{a_j\}_{j\in\mathbb{N}}$ with entries $a_j \in \{0, 1\}$ and consider the problem function

$$\Xi'(\{a_j\}) : \text{ Does } \{a_j\} \text{ have infinitely many non-zero entries?}$$

In [Colns], it was shown that $\mathrm{SCI}(\Xi', \Omega')_G = 2$ (where the evaluation functions consist in component-wise evaluation of the array $\{a_j\}$).

Proof that $\{\Xi_{\mathrm{ac}}^{\mathbb{C}}, \Omega_{f,\alpha}, \Lambda_1\} \notin \Delta_2^G$. We are done if we prove the result for $f(n) = n+1$ and $\alpha = 0$. Suppose for a contradiction that Γ_n is a height one tower of general algorithms solving $\{\Xi_{\mathrm{ac}}^{\mathbb{C}}, \Omega_{f,0}, \Lambda_1\}$. We will gain a contradiction by using the supposed tower to solve $\{\Xi', \Omega'\}$.

Given $\{a_j\} \in \Omega'$, consider the operator $H = H_0 + v$ where the potential is of the following form:

$$v(m) = \sum_{k=1}^{\infty} a_k \delta_{m,k!}.$$

Then by Theorem 5.3.3, $[0,4] \subset \mathrm{Sp}_{\mathrm{ac}}(H)$ if $\sum_k a_k < \infty$ (that is, if $\Xi'(\{a_j\}) = 0$) and $\mathrm{Sp}_{\mathrm{ac}}(H) \cap (0,4) = \emptyset$ otherwise. Given N we can evaluate any matrix value of H using only finitely many evaluations of $\{a_j\}$ and hence the evaluation functions Λ_1 can be computed using component-wise evaluations of the sequence $\{a_j\}$. We now set

$$\widehat{\Gamma}_n(\{a_j\}) = \begin{cases} 0, & \text{if } \mathrm{dist}(2, \Gamma_n(H)) < 1 \\ 1, & \text{otherwise.} \end{cases}$$

The above comments show that each of these is a general algorithm and it is clear that it converges to $\Xi'(\{a_j\})$ as $n \to \infty$, the required contradiction. □

To construct the Σ_2^A tower for $\Xi_{\mathrm{ac}}^{\mathbb{C}}$ we will need the following lemma.

Lemma 5.3.4. *Let $a < b$ with $a, b \in \mathbb{R}$ and consider the decision problem*

$$\begin{aligned} \Xi_{a,b,\mathrm{ac}} &: \Omega_{f,\alpha} \to \{0,1\} \\ A &\to \begin{cases} 1, & \text{if } \mathrm{Sp}_{\mathrm{ac}}(A) \cap [a,b] \neq \emptyset \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Then there exists a height two arithmetical tower Γ_{n_2,n_1} (with evaluation functions Λ_1) for $\Xi_{a,b,\mathrm{ac}}$. Furthermore, the final limit is from below in the sense that $\Gamma_{n_2}(A) := \lim_{n_1 \to \infty} \Gamma_{n_2,n_1}(A) \leq \Xi_{a,b,\mathrm{ac}}(A)$.

Proof. Fix such an a and b and let χ_n be a sequence of non-negative, continuous piecewise linear functions on $\mathbb{R}$, bounded by 1 and of compact support such that χ_n converge pointwise monotonically up to the constant function 1. Define also the function

$$\upsilon_{m,n}(u, A) = \langle K_H(u + i/n, A, e_m), e_m\rangle$$

and set

$$a_{m,n_2,n_1}(A) = \int_a^b \upsilon_{m,n_1}(u,A)\chi_{n_2}(|\upsilon_{m,n_1}(u,A)|)du.$$

Since each χ_n is continuous and has compact support, and since $\upsilon_{m,n}(u,A)$ converges almost everywhere to $\rho_{e_m,e_m}^A(u)$ (the Radon–Nikodym derivative of the absolutely continuous part of the measure μ_{e_m,e_m}^A), it follows by the dominated convergence theorem that

$$\lim_{n_1 \to \infty} a_{m,n_2,n_1}(A) =: a_{m,n_2}(A) = \int_a^b \rho_{e_m,e_m}^A(u)\chi_{n_2}(\rho_{e_m,e_m}^A(u))du.$$

We now use the fact that the χ_n are increasing and the dominated convergence theorem again to deduce that

$$\lim_{n_2 \to \infty} a_{m,n_2}(A) = \mu_{e_m,e_m,\mathrm{ac}}^A([a,b]),$$

with monotonic convergence from below.

Using Corollary 4.2.2 (and the now standard argument of Lipschitz continuity of the resolvent), we can compute approximations of $a_{m,n_2,n_1}(A)$ to accuracy $1/n_1$ in finitely many arithmetic operations and comparisons. Call these approximations $\widetilde{a}_{m,n_2,n_1}(A)$ and set

$$\Upsilon_{n_2,n_1}(A) = \max_{1\leq j\leq n_2} \widetilde{a}_{j,n_2,n_1}(A).$$

The proof now follows that of Lemma 5.3.2 exactly. □

Proof that $\{\Xi_{\mathrm{ac}}^{\mathbb{C}}, \Omega_{f,\alpha}, \Lambda_1\} \in \Sigma_2^A$. This is exactly the same construction as in the above proof of the inclusion $\{\Xi_{\mathrm{pp}}^{\mathbb{C}}, \Omega_{f,\alpha}, \Lambda_1\} \in \Sigma_2^A$. We simply replace the tower constructed in the proof of Lemma 5.3.2 by the tower constructed in the proof of Lemma 5.3.4. □

5.3.3 Singular continuous spectra

We will first prove the lower bound for the singular continuous spectrum via Theorem 5.3.3. Note that the impossibility result $\{\Xi_{\mathrm{sc}}^{\mathbb{C}}, \Omega_{f,\alpha}, \Lambda_1\} \notin \Delta_2^G$ follows from the same argument that was used to show $\{\Xi_{\mathrm{ac}}^{\mathbb{C}}, \Omega_{f,\alpha}, \Lambda_1\} \notin \Delta_2^G$. To show that two limits will not suffice for $f(n) - n \geq \sqrt{2n} + 1/2$, our strategy will be to reduce a certain decision problem to the computation of $\Xi_{\mathrm{sc}}^{\mathbb{C}}$. Let $(\mathcal{M}', d')$ be the space $[0, 1]$ with the usual topology and $\tilde{\Omega}$ denote the collection of all infinite matrices $\{a_{i,j}\}_{i,j\in\mathbb{N}}$ with entries $a_{i,j} \in \{0, 1\}$ and consider the problem function

$$\tilde{\Xi}_1(\{a_{i,j}\}) : \text{ Does } \{a_{i,j}\} \text{ have a column containing infinitely many non-zero entries?}$$

Recall that it was shown in Theorem 2.3.7 in Chapter 2 §2.3 that $\mathrm{SCI}(\tilde{\Xi}_1, \tilde{\Omega})_G = 3$ (where the evaluation functions consist in component-wise evaluation of the array $\{a_{i,j}\}$). We will gain a contradiction by using the supposed height two tower to solve $\{\tilde{\Xi}_1, \tilde{\Omega}\}$.

Proof that $\{\Xi_{\mathrm{sc}}^{\mathbb{C}}, \Omega_{f,\alpha}, \Lambda_1\} \notin \Delta_3^G$ *if* $f(n) - n \geq \sqrt{2n} + 1/2$. Assume that the function f satisfies $f(n) - n \geq \sqrt{2n} + 1/2$. The proof will use a direct sum construction. Given $\{a_{i,j}\} \in \tilde{\Omega}$, consider the operators $H_j = H_0 + v_{(j)}$ where the potential is of the following form:

$$v_{(j)}(n) = \sum_{k=1}^{\infty} a_{k,j}\delta_{n,k!}.$$

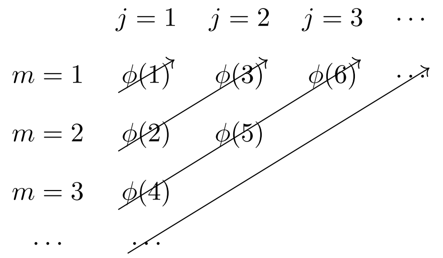
Using Theorem 5.3.3, $[0, 4] \subset \mathrm{Sp}_{\mathrm{sc}}(H_j)$ if $\sum_k a_{k,j} = \infty$ (that is, if the j-th column has infinitely many 1s) and $\mathrm{Sp}_{\mathrm{sc}}(H_j) \cap (0, 4) = \emptyset$ otherwise. Now consider an effective bijection (with effective inverse) between the canonical bases of $l^2(\mathbb{N})$ and $\oplus_{j=1}^{\infty} l^2(\mathbb{N})$:

$$\phi : \{e_n : n \in \mathbb{N}\} \to \{e_{\mathbf{k}} : \mathbf{k} \in \mathbb{N}^{\mathbb{N}}, \|\mathbf{k}\|_0 = 1\}.$$

Set $H(\{a_{i,j}\}) = \bigoplus_{j=1}^{\infty} H_j$. Then through ϕ, we view $H = H(\{a_{i,j}\})$ as a self-adjoint operator acting on $l^2(\mathbb{N})$. Explicitly, we consider the matrix

$$H_{m,n} = \langle He_{\phi(n)}, e_{\phi(m)}\rangle.$$

We choose the following bijection (where m lists the canonical basis in each Hilbert space):



A straightforward computation shows that $H \in \Omega_{f,0}$. We also observe that if $\tilde{\Xi}_1(\{a_{i,j}\}) = 1$ then $[0,4] \subset \mathrm{Sp}_{\mathrm{sc}}(H)$, otherwise $\mathrm{Sp}_{\mathrm{sc}}(H) \cap (0,4) = \emptyset$.

Suppose for a contradiction that Γ_{n_2,n_1} is a height two tower of general algorithms solving the problem $\{\Xi_{\mathrm{sc}}^{\mathbb{C}}, \Omega_{f,0}, \Lambda_1\}$. We will gain a contradiction by using the supposed height two tower to solve $\{\tilde{\Xi}_1, \tilde{\Omega}\}$. We now set

$$\widehat{\Gamma}_{n_2,n_1}(\{a_{i,j}\}) = 1 - \min\{1, \mathrm{dist}(3, \Gamma_{n_2,n_1}(A(\{a_{i,j}\})))\},$$

where we use the convention $\mathrm{dist}(3,\emptyset) = 1$. The comments above show that each of these is a general algorithm. Furthermore, the convergence of Γ_{n_2,n_1} implies that

$$\lim_{n_2\to\infty}\lim_{n_1\to\infty} \widehat{\Gamma}_{n_2,n_1}(\{a_{i,j}\}) = 1 - \min\{1, \mathrm{dist}(3, \mathrm{Sp}_{\mathrm{sc}}(H(\{a_{i,j}\})))\} = \tilde{\Xi}_1(\{a_{i,j}\}).$$

Hence $\widehat{\Gamma}_{n_2,n_1}$ is a height two tower of general algorithms solving $\{\tilde{\Xi}_1, \tilde{\Omega}\}$, a contradiction. □

Finally, we will use the following lemma to prove that the singular continuous spectrum can be computed in three limits.

Lemma 5.3.5. *Let $a < b$ with $a, b \in \mathbb{R}$ and consider the decision problem*

$$\Xi_{a,b,\mathrm{sc}} : \Omega_{f,\alpha} \to \{0,1\}$$
$$A \to \begin{cases} 1, & \text{if } \mathrm{Sp}_{\mathrm{sc}}(A) \cap [a,b] \neq \emptyset \\ 0, & \text{otherwise.} \end{cases}$$

Then there exists a height three arithmetical tower Γ_{n_3,n_2,n_1} (with evaluation functions Λ_1) for $\Xi_{a,b,\mathrm{sc}}$. Furthermore, the final limit is from below in the sense that $\Gamma_{n_3}(A) := \lim_{n_2\to\infty}\lim_{n_1\to\infty}\Gamma_{n_3,n_2,n_1}(A) \leq \Xi_{a,b,\mathrm{sc}}(A)$.

Once this is proven, we use the same construction that was used for $\{\Xi_{\mathrm{pp}}^{\mathbb{C}}, \Omega_{f,\alpha}, \Lambda_1\}, \{\Xi_{\mathrm{ac}}^{\mathbb{C}}, \Omega_{f,\alpha}, \Lambda_1\} \in \Sigma_2^A$ to show that $\{\Xi_{\mathrm{sc}}^{\mathbb{C}}, \Omega_{f,\alpha}, \Lambda_1\} \in \Sigma_3^A$, but with an additional limit. Namely, we replace (n_2, n_1) by (n_3, n_2) in the proof and use the tower constructed in the proof of Lemma 5.3.4 instead of $\widehat{\Gamma}_{n_2,n_1}(A, I)$ for an interval I. We still gain the required convergence since the only change is an additional limit in the finite number of decision problems that decide the appropriate tree.

Proof of Lemma 5.3.5. Note that we can write

$$\mu_{e_m,e_m,\mathrm{sc}}^A([a,b]) = \mu_{e_m,e_m}^A([a,b]) - \mu_{e_m,e_m,\mathrm{pp}}^A([a,b]) - \mu_{e_m,e_m,\mathrm{ac}}^A([a,b]).$$

From this and the proofs of Lemmas 5.3.2 and 5.3.4, it is clear that we can construct a height two arithmetical tower, $a_{m,n_2,n_1}(A)$, for $\mu_{e_m,e_m,\mathrm{sc}}^A([a,b])$ where the final limit is from above. Now set

$$\Upsilon_{n_3,n_2,n_1}(A) = \max_{1\leq j\leq n_3} a_{j,n_2,n_1}(A).$$

We see that each successive limit converges, with the second from above and the final from below. By taking successive maxima, minima of our base algorithms, we can assume that the second and final limits are monotonic and that $\Upsilon_{n_3,n_2,n_1}(A)$ is monotonic in both n_2 and n_3. Define the limiting sets $\Upsilon_{n_3,n_2}(A) = \lim_{n_1\to\infty}\Upsilon_{n_3,n_2,n_1}(A)$, $\Upsilon_{n_3}(A) = \lim_{n_2\to\infty}\Upsilon_{n_3,n_2}(A)$ and $\Upsilon(A) = \lim_{n_3\to\infty}\Upsilon_{n_3}(A)$. Then $\Upsilon(A)$ is zero if $\Xi_{a,b,\mathrm{sc}}(A) = 0$, otherwise it is a positive finite number.

With a slight change to the previous argument (the monotonicity in n_2 and n_3 is crucial for this to work), consider the intervals $J_1^m = [0, 1/m]$, and $J_2^m = [2/m, \infty)$. Let $k(m, n, n_1) \leq n_1$ be maximal such that $\Upsilon_{m,n,n_1}(A) \in J_1^m \cup J_2^m$. If no such k exists or $\Upsilon_{m,n,k}(A) \in J_1^m$ then set $\widehat{\Gamma}_{m,n,n_1}(A) = 0$. Otherwise set $\widehat{\Gamma}_{m,n,n_1}(A) = 1$. We then define

$$\Gamma_{n_3,n_2,n_1}(A) = \max_{1\leq m\leq n_3}\min_{1\leq n\leq n_2}\widehat{\Gamma}_{m,n,n_1}(A).$$

These can be computed using finitely many arithmetic operations and comparisons using Λ_1, and, as before, the first limit exists with

$$\Gamma_{n_3,n_2}(A) = \lim_{n_1\to\infty}\Gamma_{n_3,n_2,n_1}(A) = \max_{1\leq m\leq n_3}\min_{1\leq n\leq n_2}\widehat{\Gamma}_{m,n}(A).$$

Note that the second and third sequential limits exist through the use of maxima and minima.

Now suppose that $\Xi_{a,b,\mathrm{sc}}(A) = 0$ and fix n_3. Then for large n_2, we must have that $\Upsilon_{m,n_2}(A) < 1/(2n_3)$ for all $m \leq n_3$ due to the monotonic convergence of Υ_p as $p \to \infty$. It follows in this case that

$$\lim_{n_2\to\infty}\Gamma_{n_3,n_2}(A) = 0, \quad \text{for all } n_3.$$

Now suppose that $\Xi_{a,b,\mathrm{sc}}(A) = 1$. It follows in this case that there exists $M \in \mathbb{N}$ such that if $m \geq M$ then $\Upsilon_m(A) > 3/m$. Due to the monotonic convergence of $\Upsilon_{m,p}$ as $p \to \infty$ it follows that for all p we must have $\Upsilon_{m,p} > 3/m$ and hence there exists $N(m,p) \in \mathbb{N}$ such that if $n_1 \geq N(m,p)$ then we must have $\Upsilon_{m,p,n_1} \geq 2/m$. It follows that if $n_3 \geq M$ then we must have $\widehat{\Gamma}_{n_3,p}(A) = 1$ for all p and hence that

$$\lim_{n_3\to\infty}\Gamma_{n_3}(A) = 1.$$

The conclusion of the lemma now follows. □

Chapter 6

Discrete Spectra and Spectral Gap

Computing discrete spectra of normal operators is a problem encountered in many areas of applied mathematics and theoretical physics, as well as being of purely theoretical interest. We provide an algorithm that converges to the discrete spectrum and separates it from the essential spectrum. The method yields a sharp classification in the SCI hierarchy. This problem is subtly different to that of computing the point spectrum (eigenvalues) discussed in Chapter 5, since the discrete spectrum does not include eigenvalues of infinite multiplicity or eigenvalues embedded in the essential spectrum.

A second problem considered in this chapter is the spectral gap problem, which is related to the dichotomy between the discrete and essential spectrum. The spectral gap problem has a long tradition and is linked to many important conjectures and problems such as the Haldane conjecture [Hal83, GJL94] or the Yang–Mills mass gap problem in quantum field theory [BCD$^+$06]. In the seminal paper [CPGW15], it was shown that the spectral gap problem is undecidable (i.e., the problem $\notin \Delta_1^A$) when considering the thermodynamic limit of finite-dimensional Hamiltonians. We consider the infinite-dimensional statement of the problem and provide classifications in the SCI, as well as an extension to classifying the geometric/algebraic properties of the bottom of the spectrum.

This chapter is based on [CHns].

6.1 Main Results

Throughout this chapter, we consider various operators acting on $l^2(\mathbb{N})$. The information given to us through the functions Λ is the collection of matrix values of an operator A with respect to the canonical basis.

6.1.1 Computing discrete spectra

Let Ω_{N}^d denote the class of bounded normal operators on $l^2(\mathbb{N})$ with (known) bounded dispersion (recall (3.1.1) and this concept from §3.1.1) and with non-empty discrete spectrum (this condition can be dropped - see below), and denote by Ω_{D}^d the class of bounded diagonal self-adjoint operators in Ω_{N}^d. For a normal operator A, there is a simple decomposition of $\mathrm{Sp}(A)$ into the discrete spectrum and the essential spectrum, denoted by $\mathrm{Sp}_d(A)$ and $\mathrm{Sp}_{\mathrm{ess}}(A)$ respectively. The discrete spectrum consists of isolated points of the spectrum that are eigenvalues of finite multiplicity. The essential spectrum has numerous definitions in the non-normal case, but for the normal case is defined as the set of z such that $A - zI$ is not a Fredholm operator. Define the problem function

$$\Xi_1^d : \Omega_{\mathrm{N}}^d, \Omega_{\mathrm{D}}^d \ni A \mapsto \mathrm{cl}\left(\mathrm{Sp}_d(A)\right).$$

We have taken the closure and restricted to operators with non-empty discrete spectrum, since we want convergence with respect to the Hausdorff metric. However, the algorithm we build, Γ_{n_2,n_1}, has the property that $\lim_{n_1\to\infty}\Gamma_{n_2,n_1}(A) \subset \mathrm{Sp}_d(A)$, so this is not restrictive in practice.

Theorem 6.1.1. *Let Ξ_1^d, Ω_{N}^d and Ω_{D}^d be as above. Then,*

$$\Delta_2^G \not\ni \{\Xi_1^d, \Omega_{\mathrm{N}}^d\} \in \Sigma_2^A, \quad \Delta_2^G \not\ni \{\Xi_1^d, \Omega_{\mathrm{D}}^d\} \in \Sigma_2^A.$$

The constructed algorithm Γ_{n_2,n_1} has the property that given $A \in \Omega_{\mathrm{N}}^d$ and $z \in \mathrm{Sp}_d(A)$, the following holds. If $\epsilon > 0$ is such that $\mathrm{Sp}(A) \cap B_{2\epsilon}(z) = \{z\}$, then there is at most one point in $\Gamma_{n_2,n_1}(A)$ that also lies in $B_\epsilon(z)$. Furthermore, the limit $\lim_{n_1\to\infty}\Gamma_{n_2,n_1}(A) = \Gamma_{n_2}(A)$ is contained in the discrete spectrum and increases to $\mathrm{cl}\left(\mathrm{Sp}_d(A)\right)$ in the Hausdorff metric as $n_2 \to \infty$. In other words, a given point of $\mathrm{Sp}_d(A)$ has at most one point in $\Gamma_{n_2,n_1}(A)$ approximating it.

Let Ω_{N}^f denote the class of bounded normal operators with (known) bounded dispersion with respect to the function f. Let Ω_{D} denote the class of bounded self-adjoint diagonal operators and consider the following discrete problem (mapping into the discrete space $\{0, 1\}$)

$$\Xi_2^d : \Omega_{\mathrm{N}}^f, \Omega_{\mathrm{D}} \ni A \mapsto \text{ Is } \mathrm{Sp}_d(A) \neq \emptyset?$$

An easy corollary of the proof of Theorem 6.1.1 is as follows.

Corollary 6.1.2. *Let Ξ_2^d, Ω_{N}^f and Ω_{D} be as above. Then,*

$$\Delta_2^G \not\ni \{\Xi_2^d, \Omega_{\mathrm{N}}^f\} \in \Sigma_2^A, \quad \Delta_2^G \not\ni \{\Xi_2^d, \Omega_{\mathrm{D}}\} \in \Sigma_2^A.$$

What happens when we cannot bound the dispersion?

The algorithm constructed for Theorem 6.1.1 has $\lim_{n_1\to\infty}\Gamma_{n_2,n_1}(A) \subset \mathrm{Sp}_d(A)$. But what happens if we do not know a dispersion function f as in (3.1.1) such that we may not have known bounded dispersion? To investigate this case, let Ω_1^d denote the class of bounded normal operators with non-empty discrete spectrum and Ω_2^d the class of bounded normal operators. As the next theorem reveals, we get a jump in the SCI hierarchy.

Theorem 6.1.3. *Let Ξ_i^d and Ω_i^d be as above. Then,*

$$\Delta_3^G \not\ni \{\Xi_1^d, \Omega_1^d\} \in \Sigma_3^A, \quad \Delta_3^G \not\ni \{\Xi_2^d, \Omega_2^d\} \in \Sigma_3^A.$$

The proof shows that, without additional structure, it requires three limits to compute the discrete spectrum of self-adjoint matrices or to check if there are any isolated eigenvalues of finite multiplicity.

6.1.2 The spectral gap problem

The question can be formulated in the following way. Let $\widehat{\Omega}_{\mathrm{SA}}$ be the set of all bounded below, self-adjoint operators A on $l^2(\mathbb{N})$, for which the linear span of the canonical basis form a core of A (we do not assume A is bounded above) and such that one of the two following cases occur:

(1) The minimum of the spectrum, a, is an isolated eigenvalue with multiplicity one.

(2) There is some $\epsilon > 0$ such that $[a, a+\epsilon] \subset \mathrm{Sp}(A)$.

[DRAW PICTURE ON BOARD]

In the former case, we say the spectrum is gapped, whereas in the latter we say it is gapless. Note that, because we have restricted ourselves to the class where either (1) or (2) must hold, our problem is well-defined as a decision problem. Moreover, this definition is in line with the definitions in [CPGW15] and the physics literature. We also let $\widehat{\Omega}_{\mathrm{D}}$ denote the operators in $\widehat{\Omega}_{\mathrm{SA}}$ that are diagonal and define the decision problem (mapping into the discrete space $\{0,1\}$)

$$\Xi_{\text{gap}} : \widehat{\Omega}_{\mathrm{SA}}, \widehat{\Omega}_{\mathrm{D}} \ni A \mapsto \text{ Is the spectrum of } A \text{ gapped?} \tag{6.1.1}$$

Theorem 6.1.4 (Spectral gap)**.** *Let Ξ_{gap} be as in (6.1.1) and $\widehat{\Omega}_{\mathrm{SA}}, \widehat{\Omega}_{\mathrm{D}}$ as above. Then*

$$\Delta_2^G \not\ni \{\Xi_{\text{gap}}, \widehat{\Omega}_{\mathrm{SA}}\} \in \Sigma_2^A, \quad \Delta_2^G \not\ni \{\Xi_{\text{gap}}, \widehat{\Omega}_{\mathrm{D}}\} \in \Sigma_2^A.$$

Remark 6.1.5 (Diagonal vs. full matrix)**.** *It is worth noting that Theorem 6.1.4 shows that there is no difference in the classification of the spectral gap problem between the set of diagonal matrices and the collection of full matrices.*

The above spectral gap problem can also be extended as follows. Let $\widetilde{\Omega}_{\mathrm{SA}}^f$ denote the class of operators that are bounded below, self-adjoint, for which the linear span of the canonical basis form a core, and that have (known) bounded dispersion with respect to the function f. Let $a(A) = \inf\{x : x \in \mathrm{Sp}(A)\}$ and consider the following four cases

1. $a(A)$ lies in the discrete spectrum and has multiplicity 1,
2. $a(A)$ lies in the discrete spectrum and has multiplicity ≥ 2,
3. $a(A)$ lies in the essential spectrum but is an isolated point of the spectrum,
4. $a(A)$ is a cluster point of $\mathrm{Sp}(A)$.

[DRAW PICTURE ON BOARD]

We consider the classification problem Ξ_{class} which maps $\widetilde{\Omega}_{\mathrm{SA}}^f$ (or relevant subclasses) to the discrete space $\{1,2,3,4\}$ (with the natural order). We denote by $\widetilde{\Omega}_{\mathrm{D}}$ the class of diagonal operators in $\widetilde{\Omega}_{\mathrm{SA}}^f$.

Theorem 6.1.6 (Spectral Classification)**.** *Let Ξ_{class}, $\widetilde{\Omega}_{\mathrm{SA}}^f$ and $\widetilde{\Omega}_{\mathrm{D}}$ be as above. Then*

$$\Delta_2^G \not\ni \{\Xi_{\text{class}}, \widetilde{\Omega}_{\mathrm{SA}}^f\} \in \Pi_2^A, \quad \Delta_2^G \not\ni \{\Xi_{\text{class}}, \widetilde{\Omega}_{\mathrm{D}}\} \in \Pi_2^A.$$

6.2 Proofs of Theorems on Discrete Spectra

The following are well-known and follow from the ‘min-max’ theorem characterising eigenvalues.

Lemma 6.2.1. *Let $B \in \mathcal{B}(l^2(\mathbb{N}))$ be self-adjoint with eigenvalues $\lambda_1 \leq \lambda_2 \leq ...$ (infinitely many, counted according to multiplicity) below the essential spectrum. Consider the finite section approximates $B_n = P_n B P_n \in \mathbb{C}^n$ and list the eigenvalues of B_n as $\mu_1^n \leq \mu_2^n \leq ... \leq \mu_n^n$. Then the following hold:*

1. *$\lambda_j \leq \mu_j^n$ for $j = 1, ..., n$,*

2. *for any $j \in \mathbb{N}$, $\mu_j^n \downarrow \lambda_j$ as $n \to \infty$ ($n \geq j$ so that μ_j^n makes sense).*

Lemma 6.2.2. *Let $B \in \mathcal{B}(l^2(\mathbb{N}))$ be self-adjoint with finitely many eigenvalues $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_m$ (counted according to multiplicity) below the essential spectrum and let $a = \inf\{x : x \in \mathrm{Sp}_{\mathrm{ess}}(B)\}$. For $j > m$ we set $\lambda_j = a$. Consider the finite section approximates $B_n = P_nBP_n \in \mathbb{C}^n$ and list the eigenvalues of B_n as $\mu_1^n \leq \mu_2^n \leq ... \leq \mu_n^n$. Then the following hold:*

1. *$\lambda_j \leq \mu_j^n$ for $j = 1, ..., n$,*

2. *for any $j \leq m$, $\mu_j^n \downarrow \lambda_j$ as $n \to \infty$ ($n \geq j$ so that μ_j^n makes sense),*

3. *given $\epsilon > 0$ and $k \in \mathbb{N}$, there exists N such that for all $n \geq N$, $\mu_k^n \leq a + \epsilon$.*

Exercise: Prove these two lemmas.

Proof of Theorem 6.1.1. **Step 1**: $\{\Xi_1^d, \Omega_{\mathrm{D}}^d\} \notin \Delta_2^G$. Suppose this were false and that there exists some height one tower Γ_n solving the problem. Consider the matrix operators $A_m = \mathrm{diag}\{0, 0, ..., 0, 2\} \in \mathbb{C}^{m\times m}$ and $C = \mathrm{diag}\{0, 0, ...\}$ and set

$$A = \mathrm{diag}\{1, 2\} \oplus \bigoplus_{m=1}^{\infty} A_{k_m},$$

where we choose an increasing sequence k_m inductively as follows. Set $k_1 = 1$ and suppose that $k_1, ..., k_m$ have been chosen. $\mathrm{Sp}_d(\mathrm{diag}\{1, 2\} \oplus A_{k_1} \oplus A_{k_2} \oplus ... \oplus A_{k_m} \oplus C) = \{1, 2\}$ is closed and so there exists some $n_m \geq m$ such that if $n \geq n_m$ then

$$\mathrm{dist}(2, \Gamma_n(\mathrm{diag}\{1, 2\} \oplus A_{k_1} \oplus ... \oplus A_{k_m} \oplus C) \leq \frac{1}{4}. \tag{6.2.1}$$

Now let $k_{m+1} \geq \max\{N(\mathrm{diag}\{1, 2\} \oplus A_{k_1} \oplus ... \oplus A_{k_m} \oplus C, n_m), k_m + 1\}$. Arguing as in the proof of Theorem 3.1.6, it follows that $\Gamma_{n_m}(A) = \Gamma_{n_m}(\mathrm{diag}\{1, 2\} \oplus A_{k_1} \oplus ... \oplus A_{k_m} \oplus C)$. But $\Gamma_{n_m}(A)$ converges to $\mathrm{Sp}_d(A) = \{1\}$, contradicting (6.2.1).

Step 2: $\{\Xi_1^d, \Omega_{\mathrm{N}}^d\} \in \Sigma_2^A$. We now construct an arithmetic height two tower for Ξ_1^d and the class Ω_{N}^d. To do this, we recall that a height two tower $\tilde{\Gamma}_{n_2,n_1}$ for the essential spectrum of operators in Ω_{N}^d was constructed in [BACH$^+$20]. For completeness, we write out the algorithm here. Let P_n be the usual projection onto the first n basis elements and set $Q_n = I - P_n$. Define

$$\begin{aligned}
\mu_{m,n}(A) &:= \min\{\sigma_{\inf}(P_{f(n)}(A - zI)|_{Q_mP_n(l^2(\mathbb{N}))}), \sigma_{\inf}(P_{f(n)}(A - zI)^*|_{Q_mP_n(l^2(\mathbb{N}))})\},\\
G_n &:= \min\left\{\frac{s + it}{2^n} : s, t \in \{-2^{2n}, ..., 2^{2n}\}\right\},\\
\Upsilon_m(z) &:= z + \{w \in \mathbb{C} : |\mathrm{Re}(w)|, |\mathrm{Im}(w)| \leq 2^{-(m+1)}\}.
\end{aligned}$$

We then define the following sets for $n > m$:

$$\begin{aligned}
S_{m,n}(z) &:= \{j = m + 1, ..., n : \exists w \in \Upsilon_m(z) \cap G_j \text{ with } \mu_{m,i}(w) \leq 1/m\},\\
T_{m,n}(z) &:= \{j = m + 1, ..., n : \exists w \in \Upsilon_m(z) \cap G_j \text{ with } \mu_{m,i}(w) \leq 1/(m + 1)\},\\
E_{m,n}(z) &:= |S_{m,n}(z)| + |T_{m,n}(z)| - n,\\
I_{m,n} &:= \left\{z \in \left\{\frac{s + it}{2^m} : s, t \in \mathbb{Z}\right\} : E_{m,n}(z) > 0\right\}.
\end{aligned}$$

Finally we define for $n_1 > n_2$

$$\tilde{\Gamma}_{n_2,n_1}(A) = \bigcup_{z \in I_{n_2,n_1}} \Upsilon_{n_2}(z),$$

and set $\tilde{\Gamma}_{n_2,n_1}(A) = \{1\}$ if $n_1 \leq n_2$. Furthermore, the tower has the following desirable properties:

1. For fixed n_2, the sequence $\tilde{\Gamma}_{n_2,n_1}(A)$ is eventually constant as we increase n_1,
2. The sets $\lim_{n_1 \to \infty} \tilde{\Gamma}_{n_2,n_1}(A) =: \tilde{\Gamma}_{n_2}(A)$ are nested, converging down to $\mathrm{Sp}_{\mathrm{ess}}(A)$.

We also need the height one tower, $\hat{\Gamma}_n$, for the spectrum of operators in Ω_{N}^d discussed in Chapter 3. Note that $\hat{\Gamma}_n(A)$ is a finite set for all n. For $z \in \hat{\Gamma}_n(z)$, this also outputs an error control $E(n, z)$ such that $\mathrm{dist}(z, \mathrm{Sp}(A)) \leq E(n, z)$ and such that $E(n, z)$ converges to the true distance to the spectrum uniformly on compact subsets of $\mathbb{C}$ (with the choice of $g(x) = x$ since the operator is normal). We now fit the pieces together and initially define

$$\zeta_{n_2,n_1}(A) = \{z \in \hat{\Gamma}_{n_1}(A) : E(n_1, z) < \mathrm{dist}(z, \tilde{\Gamma}_{n_2,n_1}(A) + B_{1/n_2}(0))\}.$$

We must show that this defines an arithmetic tower in the sense of Definitions 2.1.1 and 2.1.3. Given $z \in \hat{\Gamma}_{n_1}(A)$ and using Pythagoras' theorem, along with the fact that $\tilde{\Gamma}_{n_2,n_1}(A)$ consists of finitely many squares in the complex plane aligned with the real and imaginary axes, we can compute $\mathrm{dist}(z, \tilde{\Gamma}_{n_2,n_1}(A))^2$ in finitely many arithmetic operations and comparisons. We can compute $(E(n_1, z) + 1/n_2)^2$ and check if this is less than $\mathrm{dist}(z, \tilde{\Gamma}_{n_2,n_1}(A))^2$. Hence $\zeta_{n_2,n_1}(A)$ can be computed with finitely many arithmetic operations and comparisons. There are now two cases to consider:

Case 1: $\mathrm{Sp}_d(A) \cap (\tilde{\Gamma}_{n_2}(A) + B_{1/n_2}(0))^c = \emptyset$. For large n_1, $\tilde{\Gamma}_{n_2}(A) = \tilde{\Gamma}_{n_2,n_1}(A)$ and this set contains the essential spectrum. It follows, for large n_1, since $E(n_1, z) \geq \mathrm{dist}(z, \tilde{\Gamma}_{n_2,n_1}(A))$ for all $z \in \hat{\Gamma}_{n_1}(A)$, that $\zeta_{n_2,n_1}(A) = \emptyset$.

Case 2: $\mathrm{Sp}_d(A) \cap (\tilde{\Gamma}_{n_2}(A) + B_{1/n_2}(0))^c \neq \emptyset$. In this case, this set is a finite subset of $\mathrm{Sp}_d(A)$, $\{\hat{z}_1, ..., \hat{z}_{m(n_2)}\}$, separated from the closed set $\tilde{\Gamma}_{n_2}(A) + B_{1/n_2}(0)$ (we need the $+B_{1/n_2}(0)$ for this to be true to avoid accumulation points of the discrete spectrum). There exists some $\delta_{n_2} > 0$ such that the balls $B_{2\delta_{n_2}}(\hat{z}_j)$ for $j = 1, ..., m(n_2)$ are pairwise disjoint and such that their union does note intersect $\tilde{\Gamma}_{n_2}(A) + B_{1/n_2}(0)$.Using the convergence of $\hat{\Gamma}_{n_1}(A)$ to $\mathrm{Sp}(A)$ and $E(n, z) \geq \mathrm{dist}(z, \mathrm{Sp}(A))$, it follows that for large n_1 that

$$\zeta_{n_2,n_1}(A) \subset \bigcup_{j=1}^{m(n_2)} B_{\delta_{n_2}}(\hat{z}_j), \tag{6.2.2}$$

is non-empty and that $\zeta_{n_2,n_1}(A)$ converges to $\mathrm{Sp}_d(A) \cap (\tilde{\Gamma}_{n_2}(A) + B_{1/n_2}(0))^c \neq \emptyset$ in the Hausdorff metric.

Suppose that $\zeta_{n_2,n_1}(A)$ is non-empty. Recall that we only want one output per eigenvalue in the discrete spectrum. To do this, we partition the finite set $\zeta_{n_2,n_1}(A)$ into equivalence classes as follows. For $z, w \in \zeta_{n_2,n_1}(A)$, we say that $z \sim_{n_1} w$ if there exists a finite sequence $z = z_1, z_2, ..., z_n = w \in \zeta_{n_2,n_1}(A)$ such that $B_{E(n_1,z_j)}(z_j)$ and $B_{E(n_1,z_{j+1})}(z_{j+1})$ intersect. The idea is that equivalence classes correspond to clusters of points in $\zeta_{n_2,n_1}(A)$. Given any $z \in \zeta_{n_2,n_1}(A)$ we can compute its equivalence class using finitely many arithmetic operations and comparisons. Let S_0 be the set $\{z\}$ and given S_n, let S_{n+1} be the union of any $w \in \zeta_{n_2,n_1}(A)$ such that $B_{E(n_1,w)}(w)$ and $B_{E(n_1,v)}(v)$ intersect for some $v \in S_n$. Given S_n, we can compute S_{n+1} using finitely many arithmetic operations and comparisons. The equivalence class is any S_n where $S_n = S_{n+1}$ which must happen since $\zeta_{n_2,n_1}(A)$ is finite. We let Φ_{n_2,n_1} consist of one element of each equivalence class that minimises $E(n_1,\cdot)$ over its respective equivalence class. By the above comments it is clear that Φ_{n_2,n_1} can be computed in finitely many arithmetic operations and comparisons from the given data. Furthermore, due to (6.2.2) which holds for large n_1, the separation of the $B_{2\delta_{n_2}}(\hat{z}_j)$ and the fact that $E(n_1,\cdot)$ converges uniformly on compact subsets to the distance to $\mathrm{Sp}(A)$, it follows that for large n_1 there is exactly one point in each intersection $B_{2\delta_{n_2}}(\hat{z}_j)\cap\Phi_{n_2,n_1}(A)$. But we can shrink δ_{n_2} and apply the same argument to see that $\Phi_{n_2,n_1}(A)$ converges to $\mathrm{Sp}_d(A)\cap(\tilde{\Gamma}_{n_2}(A)+B_{1/n_2}(0))^c\neq\emptyset$ in the Hausdorff metric.

Now suppose that $\zeta_{n_2,n_1}(A)$ is non-empty and $z_1,z_2\in\Phi_{n_2,n_1}(A)$ and both lie in $B_\epsilon(z)$ for some $z\in\mathrm{Sp}_d(A)$ and $\epsilon>0$ with $\mathrm{Sp}(A)\cap B_{2\epsilon}(z)=\{z\}$. It follows that z minimises the distance to the spectrum from both z_1 and z_2. Hence, $B_{E(n_1,z_1)}(z_1)$ and $B_{E(n_1,z_2)}(z_2)$ both contain the point z so that $z_1\sim_{n_1}z_2$. But then at most one of z_1,z_2 can lie in $\Phi_{n_2,n_1}(A)$ and hence $z_1=z_2$.

To finish, we must alter $\Phi_{n_2,n_1}(A)$ to take care of the case when $\zeta_{n_2,n_1}(A)=\emptyset$ and to produce a Σ_2^A algorithm. In the case that $\zeta_{n_2,n_1}(A)=\emptyset$, set $\Phi_{n_2,n_1}(A)=\emptyset$. Let $N(A)\in\mathbb{N}$ be minimal such that $\mathrm{Sp}_d(A)\cap(\tilde{\Gamma}_N(A)+B_{1/N}(0))^c\neq\emptyset$ (recall the discrete spectrum is non-empty for our class of operators). If $n_2>n_1$ then set $\Gamma_{n_2,n_1}(A)=\{0\}$, otherwise consider $\Phi_{k,n_1}(A)$ for $n_2\leq k\leq n_1$. If all of these are empty then set $\Gamma_{n_2,n_1}(A)=\{0\}$, otherwise choose minimal k with $\Phi_{k,n_1}(A)\neq\emptyset$ and let $\Gamma_{n_2,n_1}(A)=\Phi_{k,n_1}(A)$. Note that this defines an arithmetic tower of algorithms, with $\Gamma_{n_2,n_1}(A)$ non-empty. By the above case analysis, for large n_1 it holds that

$$\Gamma_{n_2,n_1}(A)=\Phi_{n_2\vee N(A),n_1}(A)$$

and it follows that

$$\lim_{n_1\to\infty}\Gamma_{n_2,n_1}(A)=:\Gamma_{n_2}(A)=\mathrm{Sp}_d(A)\cap(\tilde{\Gamma}_{n_2\vee N(A)}(A)+B_{1/n_2\vee N(A)}(0))^c.$$

Hence $\Gamma_{n_2}(A)\subset\mathrm{Sp}_d(A)$ and $\Gamma_{n_2}(A)$ converges up to $\mathrm{cl}\,(\mathrm{Sp}_d(A))$ in the Hausdorff metric. □

Proof of Corollary 6.1.2. Since $\Omega_{\mathrm{D}}\subset\Omega_{\mathrm{N}}^f$, its suffices to show that $\{\Xi_2^d,\Omega_{\mathrm{N}}^f\}\in\Sigma_2^A$ and $\{\Xi_2^d,\Omega_{\mathrm{D}}\}\notin\Delta_2^G$.

Step 1: $\{\Xi_2^d,\Omega_{\mathrm{D}}\}\notin\Delta_2^G$. The proof is almost identical to step 1 in the proof of Theorem 6.1.1. Suppose there exists some height one tower Γ_n solving the problem. Consider the matrix operators $A_m=\mathrm{diag}\{0,0,...,0,2\}\in\mathbb{C}^{m\times m}$ and $C=\mathrm{diag}\{0,0,...\}$ and set

$$A=\bigoplus_{m=1}^{\infty}A_{k_m},$$

where we choose an increasing sequence k_m inductively as follows. Set $k_1=1$ and suppose that $k_1,...,k_m$ have been chosen. $\mathrm{Sp}_d(A_{k_1}\oplus A_{k_2}\oplus...\oplus A_{k_m}\oplus C)=\{2\}$ so there exists some $n_m\geq m$ such that if $n\geq n_m$ then

$$\Gamma_n(A_{k_1}\oplus...\oplus A_{k_m}\oplus C)=1.$$

Now let $k_{m+1}\geq\max\{N(\mathrm{diag}\{1,2\}\oplus A_{k_1}\oplus...\oplus A_{k_m}\oplus C,n_m),k_m+1\}$. Arguing as in the proof of Theorem 3.1.6, it follows that $\Gamma_{n_m}(A)=\Gamma_{n_m}(A_{k_1}\oplus...\oplus A_{k_m}\oplus C)$. But $\Gamma_{n_m}(A)$ converges to 0 as A has no discrete spectrum and this contradiction finishes this step.

Step 2: $\{\Xi_2^d,\Omega_{\mathrm{N}}^f\}\in\Sigma_2^A$. Consider the height two tower, ζ_{n_2,n_1}, defined in step 2 of the proof of Theorem 6.1.1. Let $A\in\Omega_{\mathrm{N}}^f$ and if $\zeta_{n_2,n_1}(A)=\emptyset$, define $\rho_{n_2,n_1}(A)=0$, otherwise define $\rho_{n_2,n_1}(A)=1$. The discussion in the proof of Theorem 6.1.1 shows that

$$\lim_{n_1\to\infty}\rho_{n_2,n_1}(A)=:\rho_{n_2}(A)=\begin{cases}0, & \text{if } \mathrm{Sp}_d(A)\cap(\tilde{\Gamma}_{n_2}(A)+B_{1/n_2}(0))^c=\emptyset\\ 1, & \text{otherwise.}\end{cases}$$

Since $\mathrm{Sp}_d(A)\cap(\tilde{\Gamma}_{n_2}(A)+B_{1/n_2}(0))^c$ increases to $\mathrm{cl}\,(\mathrm{Sp}_d(A))$, it follows that $\lim_{n_2\to\infty}\rho_{n_2}(A)=\Xi_2^d(A)$ and that if $\rho_{n_2}(A)=1$, then $\Xi_2^d(A)=1$. Hence, ρ_{n_2,n_1} provides a Σ_2^A tower for $\{\Xi_2^d,\Omega_{\mathrm{N}}^f\}$. □

Proof of Theorem 6.1.3. **Step 1:** $\{\Xi_1^d,\Omega_1^d\}\notin\Delta_3^G$. Suppose for a contradiction that Γ_{n_2,n_1} is a height two tower solving this problem. For this proof we shall use the decision problem $\tilde{\Xi}_2$ from §2.3 which was proven in Theorem 2.3.7 to have $\mathrm{SCI}_G=3$. For convenience, we remind the reader of this decision problem. Let $(\mathcal{M},d)$ be the discrete space $\{0,1\}$, let $\tilde{\Omega}$ denote the collection of all infinite matrices $\{a_{i,j}\}_{i,j\in\mathbb{N}}$ with entries $a_{i,j}\in\{0,1\}$ and consider the problem function

$\tilde{\Xi}_2(\{a_{i,j}\})$: Does $\{a_{i,j}\}$ have only finitely many columns containing only finitely many non-zero entries?

We will gain a contradiction by using the supposed height two tower for $\{\Xi_1^d,\Omega_1^d\}$, Γ_{n_2,n_1}, to solve $\{\tilde{\Xi}_2,\tilde{\Omega}\}$.

Without loss of generality, identify $\mathcal{B}(l^2(\mathbb{N}))$ with $\mathcal{B}(X)$ where $X=\mathbb{C}^2\oplus\bigoplus_{j=1}^{\infty}X_j$ in the l^2-sense with $X_j=l^2(\mathbb{N})$. Now let $\{a_{i,j}\}\in\tilde{\Omega}$ and for the jth column define $B_j\in\mathcal{B}(X_j)$ with the following matrix representation:

$$B_j=\bigoplus_{r=1}^{M_j}A_{l_r^j},\quad A_m:=\begin{pmatrix}1 & & & & 1\\ & 0 & & & \\ & & \ddots & & \\ & & & 0 & \\ 1 & & & & 1\end{pmatrix}\in\mathbb{C}^{m\times m},$$

where if M_j is finite then $l_{M_j}^j=\infty$ with $A_\infty=\mathrm{diag}(1,0,0,...)$. The l_r^j are defined such that

$$\sum_{r=1}^{\sum_{i=1}^m a_{i,j}}l_r^j=m+\sum_{i=1}^m a_{i,j}. \tag{6.2.3}$$

Define the self-adjoint operator

$$A=\mathrm{diag}\{3,1\}\oplus\bigoplus_{j=1}^{\infty}B_j.$$

Note that no matter what the choices of l_r^j are, $3\in\mathrm{Sp}_d(A)$ and hence $A\in\Omega_1^d$. Note also that the spectrum of A is contained in $\{0,1,2,3\}$. If $\tilde{\Xi}_2(\{a_{i,j}\})=1$ then 1 is an isolated eigenvalue of finite multiplicity and hence in $\mathrm{Sp}_d(A)$. But if $\tilde{\Xi}_2(\{a_{i,j}\})=0$ then 1 is an isolated eigenvalue of infinite multiplicity so does not lie in the discrete spectrum and hence $\mathrm{Sp}_d(A)\subset\{0,2,3\}$.

Consider the intervals $J_1=[0,1/2]$, and $J_2=[3/4,\infty)$. Set $\alpha_{n_2,n_1}=\mathrm{dist}(1,\Gamma_{n_2,n_1}(A))$. Let $k(n_2,n_1)\leq n_1$ be maximal such that $\alpha_{n_2,k}(A)\in J_1\cup J_2$. If no such k exists or $\alpha_{n_2,k}(A)\in J_1$ then set $\tilde{\Gamma}_{n_2,n_1}(\{a_{i,j}\})=1$. Otherwise set $\tilde{\Gamma}_{n_2,n_1}(\{a_{i,j}\})=0$. It is clear from (6.2.3) that this defines a generalised algorithm. In particular, given N we can evaluate $\{A_{k,l}:k,l\leq N\}$ using only finitely many evaluations of $\{a_{i,j}\}$, where we can use a suitable bijection between bases of $l^2(\mathbb{N})$ and $\mathbb{C}^2\oplus\bigoplus_{j=1}^{\infty}X_j$ to view A as acting on $l^2(\mathbb{N})$. The point of the intervals J_1,J_2 is that we can show $\lim_{n_1\to\infty}\tilde{\Gamma}_{n_2,n_1}(\{a_{i,j}\})=\tilde{\Gamma}_{n_2}(\{a_{i,j}\})$ exists. If $\tilde{\Xi}_2(\{a_{i,j}\})=1$, then, for large n_2, $\lim_{n_1\to\infty}\alpha_{n_2,k}(A)<1/2$ and hence it follows that $\lim_{n_2\to\infty}\tilde{\Gamma}_{n_2}(\{a_{i,j}\})=1$. Similarly, if $\tilde{\Xi}_2(\{a_{i,j}\})=0$, then, for large n_2, we must have that

$\lim_{n_1\to\infty}\alpha_{n_2,k}(A) > 3/4$ and hence it follows that $\lim_{n_2\to\infty}\tilde{\Gamma}_{n_2}(\{a_{i,j}\}) = 0$. Hence $\tilde{\Gamma}_{n_2,n_1}$ is a height two tower of general algorithms solving $\{\tilde{\Xi}_2, \tilde{\Omega}\}$, a contradiction.

Step 2: $\{\Xi_2^d, \Omega_2^d\} \notin \Delta_3^G$. To prove this we can use a slight alteration of the argument in step 1. Replace X by $X = l^2(\mathbb{N}) \oplus \bigoplus_{j=1}^{\infty} X_j$ and A by

$$A = \mathrm{diag}\{1, 0, 2, 0, 2, ...\} \oplus \bigoplus_{j=1}^{\infty} B_j.$$

It is then clear that $\Xi_2^d(A) = 1$ if and only if $\tilde{\Xi}_2(\{a_{i,j}\}) = 1$.

Step 3: $\{\Xi_1^d, \Omega_1^d\} \in \Sigma_3^A$. For this we argue similarly to the proof of Theorem 6.1.1 step 2. It was shown in [BACH+20] that there exists a height three arithmetic tower $\tilde{\Gamma}_{n_3,n_2,n_1}$ for the essential spectrum of operators in Ω_1^d such that

- Each $\tilde{\Gamma}_{n_3,n_2,n_1}(A)$ consists of a finite collection of points in the complex plane.
- For large n_1, $\tilde{\Gamma}_{n_3,n_2,n_1}(A)$ is eventually constant and equal to $\tilde{\Gamma}_{n_3,n_2}(A)$.
- $\tilde{\Gamma}_{n_3,n_2}(A)$ is increasing with n_2 with limit $\tilde{\Gamma}_{n_3}(A)$ containing the essential spectrum. The limit $\tilde{\Gamma}_{n_3}(A)$ is also decreasing with n_3.

Furthermore, it was proven in [BACH+20] that for operators in Ω_1^d, there exists a height two arithmetic tower $\hat{\Gamma}_{n_2,n_1}$ for computing the spectrum such that

- $\hat{\Gamma}_{n_2,n_1}(A)$ is constant for large n_1.
- For any $z \in \hat{\Gamma}_{n_2}(A)$, $\mathrm{dist}(z, \mathrm{Sp}(A)) \leq 2^{-n_2}$.

Using these, we initially define

$$\zeta_{n_3,n_2,n_1}(A) = \{z \in \hat{\Gamma}_{n_2,n_1}(A) : 2^{-n_3} - 2^{-n_2} \leq \mathrm{dist}(z, \tilde{\Gamma}_{n_3,n_2,n_1}(A))\}.$$

The arguments in the proof of Theorem 6.1.1 show that this can be computed in finitely many arithmetic operations and comparisons using the relevant evaluation functions. Note that for large n_1

$$\zeta_{n_3,n_2,n_1}(A) = \{z \in \hat{\Gamma}_{n_2}(A) : 2^{-n_3} - 2^{-n_2} \leq \mathrm{dist}(z, \tilde{\Gamma}_{n_3,n_2}(A))\} =: \zeta_{n_3,n_2}(A).$$

There are now two cases to consider (we use $D_\eta(z)$ to denote the open ball of radius η about a point z):

Case 1: $\mathrm{Sp}_d(A) \cap (\tilde{\Gamma}_{n_3}(A) + D_{2^{-n_3}}(0))^c = \emptyset$. Suppose, for a contradiction, in this case that there exists $z_{m_j} \in \zeta_{n_3,m_j}(A)$ with $m_j \to \infty$. Then, without loss of generality, $z_{m_j} \to z \in \mathrm{Sp}(A)$. We also have that

$$\mathrm{dist}(z_{m_j}, \tilde{\Gamma}_{n_3,m_j}(A)) \geq 2^{-n_3} - 2^{-m_j},$$

which implies that $\mathrm{dist}(z, \tilde{\Gamma}_{n_3}(A)) \geq 2^{-n_3}$ and hence $z \in \mathrm{Sp}_d(A) \cap (\tilde{\Gamma}_{n_3}(A) + D_{2^{-n_3}}(0))^c$, the required contradiction. It follows that $\zeta_{n_3,n_2}(A)$ is empty for large n_2.

Case 2: $\mathrm{Sp}_d(A) \cap (\tilde{\Gamma}_{n_3}(A) + D_{2^{-n_3}}(0))^c \neq \emptyset$. In this case, this set is a finite subset of $\mathrm{Sp}_d(A)$, $\{\hat{z}_1, ..., \hat{z}_{m(n_3)}\}$. Each of these points is an isolated point of the spectrum. It follows that there exists $z_{n_2} \in \hat{\Gamma}_{n_2}(A)$ with $z_{n_2} \to \hat{z}_1$ and $|z_{n_2} - \hat{z}_1| \leq 2^{-n_2}$ for large n_2. Since the $\tilde{\Gamma}_{n_3,n_2}(A)$ are increasing, this implies that

$$\begin{aligned}\mathrm{dist}(z_{n_2}, \tilde{\Gamma}_{n_3,n_2}(A)) &\geq \mathrm{dist}(z_{n_2}, \tilde{\Gamma}_{n_3}(A))\\ &\geq \mathrm{dist}(\hat{z}_1, \tilde{\Gamma}_{n_3}(A)) - 2^{-n_2} \geq 2^{-n_3} - 2^{-n_2},\end{aligned}$$

so that $z_{n_2} \in \zeta_{n_3,n_2}(A)$. The same argument holds for points converging to all of $\{\hat{z}_1, ..., \hat{z}_{m(n_3)}\}$. On the other hand, the argument used in Case 1 shows that any limit points of $\zeta_{n_3,n_2}(A)$ as $n_2 \to \infty$ are contained in $\mathrm{Sp}_d(A) \cap (\tilde{\Gamma}_{n_3}(A) + D_{2^{-n_3}}(0))^c$. It follows that in this case $\zeta_{n_3,n_2}(A)$ converges to $\mathrm{Sp}_d(A) \cap (\tilde{\Gamma}_{n_3}(A) + B_{1/n_3}(0))^c \neq \emptyset$ in the Hausdorff metric as $n_2 \to \infty$.

Let $N(A) \in \mathbb{N}$ be minimal such that $\mathrm{Sp}_d(A) \cap (\tilde{\Gamma}_N(A) + D_{2^{-N}}(0))^c \neq \emptyset$ (recall the discrete spectrum is non-empty for our class of operators). If $n_3 > n_2$ then set $\Gamma_{n_3,n_2,n_1}(A) = \{0\}$, otherwise consider $\zeta_{k,n_2,n_1}(A)$ for $n_3 \leq k \leq n_2$. If all of these are empty then set $\Gamma_{n_3,n_2,n_1}(A) = \{0\}$, otherwise choose minimal k with $\zeta_{k,n_2,n_1}(A) \neq \emptyset$ and let $\Gamma_{n_3,n_2,n_1}(A) = \zeta_{k,n_2,n_1}(A)$. Note that this defines an arithmetic tower of algorithms, with $\Gamma_{n_3,n_2,n_1}(A)$ non-empty. Since we consider finitely many of the sets $\zeta_{k,n_2,n_1}(A)$, and these are constant for large n_1, it follows that $\Gamma_{n_3,n_2,n_1}(A)$ is constant for large n_1 and constructed in the same manner with replacing $\zeta_{k,n_2,n_1}(A)$ by $\zeta_{k,n_2}(A)$. Call this limit $\Gamma_{n_3,n_2}(A)$.

For large n_2,

$$\Gamma_{n_3,n_2}(A) = \zeta_{n_3 \vee N(A),n_2}(A)$$

and it follows that

$$\lim_{n_2 \to \infty} \Gamma_{n_3,n_2}(A) =: \Gamma_{n_3}(A) = \mathrm{Sp}_d(A) \cap (\tilde{\Gamma}_{n_3 \vee N(A)}(A) + D_{2^{-n_3 \vee N(A)}}(0))^c.$$

Hence $\Gamma_{n_3}(A) \subset \mathrm{Sp}_d(A)$ and $\Gamma_{n_3}(A)$ converges up to $\mathrm{cl}\,(\mathrm{Sp}_d(A))$ in the Hausdorff metric.

Step 4: $\{\Xi_2^d, \Omega_2^d\} \in \Sigma_3^A$. Consider the height three tower, ζ_{n_3,n_2,n_1}, defined in step 3. Let $A \in \Omega_2^d$ and if $\zeta_{n_3,n_2,n_1}(A) = \emptyset$, define $\rho_{n_3,n_2,n_1}(A) = 0$, otherwise define $\rho_{n_3,n_2,n_1}(A) = 1$. The discussion in step 3 shows that

$$\lim_{n_2 \to \infty} \lim_{n_1 \to \infty} \rho_{n_3,n_2,n_1}(A) =: \rho_{n_3}(A) = \begin{cases} 0, & \text{if } \mathrm{Sp}_d(A) \cap (\tilde{\Gamma}_{n_3}(A) + D_{2^{-n_3}}(0))^c = \emptyset \\ 1, & \text{otherwise.} \end{cases}$$

Since $\mathrm{Sp}_d(A) \cap (\tilde{\Gamma}_{n_3}(A) + D_{2^{-n_3}}(0))^c$ increases to $\mathrm{cl}\,(\mathrm{Sp}_d(A))$, it follows that $\lim_{n_3 \to \infty} \rho_{n_3}(A) = \Xi_2^d(A)$ and that if $\rho_{n_3}(A) = 1$, then $\Xi_2^d(A) = 1$. Hence, ρ_{n_3,n_2,n_1} provides a Σ_3^A tower for $\{\Xi_2^d, \Omega_2^d\}$. □

6.3 Proofs of Theorems on the Spectral Gap

Proof of Theorem 6.1.4. **Step 1**: $\{\Xi_{\mathrm{gap}}, \widehat{\Omega}_{\mathrm{SA}}\} \in \Sigma_2^A$. Let $A \in \widehat{\Omega}_{\mathrm{SA}}$. We can compute all n eigenvalues of $P_n A P_n$ to arbitrary precision in finitely many arithmetic operations and comparisons. In the notation of Lemmas 6.2.1, and 6.2.2 (whose analogous results also hold for the possibly unbounded $A \in \widehat{\Omega}_{\mathrm{SA}}$), consider an approximation

$$0 \leq l_n := \mu_2^n - \mu_1^n + \epsilon_n, \quad n \geq 2,$$

where we have computed $\mu_2^n - \mu_1^n$ to accuracy $|\epsilon_n| \leq 1/n$. Recall that for $A \in \widehat{\Omega}_{\mathrm{SA}}$ we restricted the class so that either the bottom of the spectrum is in the discrete spectrum with multiplicity one, or there is a closed interval in the spectrum of positive measure with the bottom of the spectrum as its left end-point. It follows that l_n converges to zero if and only if $\Xi_{\mathrm{gap}}(A) = 0$, otherwise it converges to some positive number. If $n_1 = 1$ then set $\Gamma_{n_2,n_1}(A) = 1$, otherwise consider the following.

Let $J_{n_2}^1 = [0, 1/(2n_2)]$ and $J_{n_2}^2 = (1/n_2, \infty)$. Given $n_1 \in \mathbb{N}$, consider l_k for $k \leq n_1$. If no such k exists with $l_k \in J_{n_2}^1 \cup J_{n_2}^2$ then set $\Gamma_{n_2,n_1}(A) = 0$. Otherwise, consider k maximal with $l_k \in J_{n_2}^1 \cup J_{n_2}^2$ and set $\Gamma_{n_2,n_1}(A) = 0$ if $l_k \in J_{n_2}^1$ and $\Gamma_{n_2,n_1}(A) = 1$ if $l_k \in J_{n_2}^2$. The sequence $l_{n_1} \to c \geq 0$ for some number

c. The separation of the intervals $J^1_{n_2}$ and $J^2_{n_2}$, ensures that l_{n_1} cannot be in both intervals infinitely often as $n_1 \to \infty$ and hence the first limit $\Gamma_{n_2}(A) := \lim_{n_1\to\infty} \Gamma_{n_2,n_1}(A)$ exists. If $c = 0$, then $\Gamma_{n_2}(A) = 0$ but if $c > 0$ then there exists n_2 with $1/n_2 < c$ and hence for large n_1, $l_{n_1} \in J^2_{n_2}$. It follows in this case that $\Gamma_{n_2}(A) = 1$ and we also see that if $\Gamma_{n_2}(A) = 1$ then $\Xi_{\mathrm{gap}}(A) = 1$. Hence Γ_{n_2,n_1} provides a Σ^A_2 tower.

Step 2: $\{\Xi_{\mathrm{gap}}, \widehat{\Omega}_{\mathrm{D}}\} \notin \Delta^G_2$. We argue by contradiction and assume the existence of a height one tower, Γ_n converging to Ξ_{gap}. The method of proof follows the same lines as before. For every A and n there exists a finite number $N(A, n) \in \mathbb{N}$ such that the evaluations from $\Lambda_{\Gamma_n}(A)$ only take the matrix entries $A_{ij} = \langle Ae_j, e_i\rangle$ with $i, j \le N(A, n)$ into account. List the rationals in $(0, 1)$ without repetition as $d_1, d_2, ...$ We consider the operators $A_m = \mathrm{diag}\{d_1, d_2, ..., d_m\} \in \mathbb{C}^{m\times m}$, $B_m = \mathrm{diag}\{1, 1, ..., 1\} \in \mathbb{C}^{m\times m}$ and $C = \mathrm{diag}\{1, 1, ...\}$. Let

$$A = \bigoplus_{m=1}^{\infty} (B_{k_m} \oplus A_{k_m}),$$

where we choose an increasing sequence k_m inductively as follows. In what follows, all operators considered are easily seen to be in $\widehat{\Omega}_{\mathrm{D}}$.

Set $k_1 = 1$ and suppose that $k_1, ..., k_m$ have been chosen with the property that upon defining

$$\zeta_p := \min\{d_r : 1 \le r \le k_p\},$$

we have $\zeta_p > \zeta_{p+1}$ for $p = 1, ..., m - 1$. $\mathrm{Sp}(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C) = \{d_1, d_2, ..., d_m, 1\}$ has ζ_m the minimum of its spectrum and an isolated eigenvalue of multiplicity 1, hence

$$\Xi(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C) = 1.$$

It follows that there exists some $n_m \ge m$ such that if $n \ge n_m$ then

$$\Gamma_n(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C) = 1.$$

Now let $k_{m+1} \ge \max\{N(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C, n_m), k_m + 1\}$ with $\zeta_m > \zeta_{m+1}$. The same argument used in the proof of Theorem 3.1.6 shows that $\Gamma_{n_m}(A) = \Gamma_{n_m}(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C) = 1$. But $\mathrm{Sp}(A) = [0, 1]$ is gapless and so must have $\lim_{n\to\infty}(\Gamma_n(A)) = 0$, a contradiction. □

Proof of Theorem 6.1.6. By restricting $\widetilde{\Omega}_{\mathrm{D}}$ to $\widehat{\Omega}_{\mathrm{D}}$ and composing with the map

$$\rho : \{1, 2, 3, 4\} \to \{0, 1\},$$

$\rho(1) = 1$, $\rho(2) = \rho(3) = \rho(4) = 0$, it is clear that Theorem 6.1.4 implies $\{\Xi_{\mathrm{class}}, \widetilde{\Omega}^f_{\mathrm{SA}}\}, \{\Xi_{\mathrm{class}}, \widetilde{\Omega}_{\mathrm{D}}\} \notin \Delta^G_2$. Since $\widetilde{\Omega}_{\mathrm{D}} \subset \widetilde{\Omega}^f_{\mathrm{SA}}$, we need only construct a Π^A_2 tower for $\{\Xi_{\mathrm{class}}, \widetilde{\Omega}^f_{\mathrm{SA}}\}$.

Let $A \in \widetilde{\Omega}^f_{\mathrm{SA}}$. For a given n, set $B_n = P_nAP_n$ and in the notation of Lemmas 6.2.2 and 6.2.1, let

$$0 \le l^j_n := \mu^n_{j+1} - \mu^n_1 + \epsilon^j_n, \text{ for } j < n.$$

where we again have computed $\mu^n_{j+1} - \mu^n_1$ to accuracy $\left|\epsilon^j_n\right| \le 1/n$ using only finitely many arithmetic operations and comparisons. $\Xi_{\mathrm{class}}(A) = 1$ if and only if l^1_n converges to a positive constant as $n \to \infty$ and $\Xi_{\mathrm{class}}(A) = 2$ if and only if l^1_n converges to zero as $n \to \infty$ but there exists j with l^j_n convergent to a positive constant.

Note that we can use the algorithm, denoted $\hat{\Gamma}_n$, to compute the spectrum presented in Chapter 3, with error function denoted by $E(n, \cdot)$ converging uniformly on compact subsets of $\mathbb{C}$ to the true error from above (again with the choice of $g(x) = x$ since the operator is normal). Setting

$$a_n(A) = \min_{x \in \hat{\Gamma}_n(A)} \{x + E(n,x)\},$$

we see that $a_n(A) \geq a(A) := \inf_{x \in \mathrm{Sp}(A)}\{x\}$ and that $a_n(A) \to a(A)$. Now consider

$$b_{n_2,n_1}(A) = \min\{E(k, a_k(A) + 1/n_2) + 1/k : 1 \leq k \leq n_1\}$$

then $b_{n_2,n_1}(A)$ is positive and decreasing in n_1 so converges to some limit $b_{n_2}(A)$.

Lemma 6.3.1. *Let $A \in \widetilde{\Omega}_{\mathrm{SA}}^f$ and $c_{n_2,n_1}(A) = E(n_1, a_{n_1}(A) + 1/n_2) + 1/n_1$, then*

$$\lim_{n_1 \to \infty} c_{n_2,n_1}(A) =: c_{n_2}(A) = \mathrm{dist}(a + 1/n_2, \mathrm{Sp}(A)).$$

Furthermore, if $\Xi_{\mathrm{class}}(A) \neq 4$ then for large n_2 it follows that $c_{n_2}(A) = b_{n_2}(A) = 1/n_2$.

Proof of Lemma 6.3.1. We know that $a_{n_1}(A) + 1/n_2$ converges to $a(A) + 1/n_2$ as $n_1 \to \infty$. Furthermore, $\mathrm{dist}(z, \mathrm{Sp}(A))$ is continuous in z and $E(n_1, z)$ converges uniformly to $\mathrm{dist}(z, \mathrm{Sp}(A))$ on compact subsets of $\mathbb{C}$. Hence, the limit $c_{n_2}(A)$ exists and is equal to $\mathrm{dist}(a(A) + 1/n_2, \mathrm{Sp}(A))$. It is clear that $b_{n_2}(A) \leq c_{n_2}(A)$. Suppose now that $\Xi_{\mathrm{class}}(A) \neq 4$, then for large n_1, say bigger than some N, and for large enough n_2,

$$\begin{aligned} E(n_1, a_{n_1}(A) + 1/n_2) &\geq \mathrm{dist}(a_{n_1}(A) + 1/n_2, \mathrm{Sp}(A)) \\ &= |a_{n_1}(A) + 1/n_2 - a(A)| \\ &\geq 1/n_2 = \mathrm{dist}(a(A) + 1/n_2, \mathrm{Sp}(A)). \end{aligned}$$

Now choose n_2 large such that the above inequality holds and $1/n_2 \leq 1/N$. Then $b_{n_2,n_1}(A) \geq 1/n_2$. Taking limits finishes the proof. □

If $n_2 \geq n_1$ then set $\Gamma_{n_2,n_1}(A) = 1$. Otherwise, for $1 \leq j \leq n_2$, let $k^j_{n_2,n_1}$ be maximal with $1 \leq k^j_{n_2,n_1} < n_1$ such that $l^j_{k^j_{n_2,n_1}} \in J^1_{n_2} \cup J^2_{n_2}$ if such $k^j_{n_2,n_1}$ exist, where $J^1_{n_2}$ and $J^2_{n_2}$ are as in the proof of Theorem 6.1.4. If $k^1_{n_2,n_1}$ exists with $l^1_{k^1_{n_2,n_1}} \in J^2_{n_2}$ then set $\Gamma_{n_2,n_1}(A) = 1$. Otherwise, if any of $k^m_{n_2,n_1}$ exists with $l^m_{k^m_{n_2,n_1}} \in J^2_{n_2}$ for $2 \leq m \leq n_2$ then set $\Gamma_{n_2,n_1}(A) = 2$. Suppose that neither of these two cases hold. In this case compute $b_{n_2,n_1}(A)$. If $b_{n_2,n_1}(A) \geq 1/n_2$ then set $\Gamma_{n_2,n_1}(A) = 3$, otherwise set $\Gamma_{n_2,n_1}(A) = 4$. We now must show this provides a Π^A_2 tower solving our problem.

First we show convergence of the first limit. Fix n_2 and consider n_1 large. The separation of the intervals $J^1_{n_2}$ and $J^2_{n_2}$ ensures that each sequence $\{l^j_n\}_{n \in \mathbb{N}}$ cannot visit each interval infinitely often. Since $b_{n_1,n_2}(A)$ is non-increasing in n_1, we also see that the question whether $b_{n_2,n_1}(A) \geq 1/n_2$ eventually has a constant answer. These observations ensure convergence of the first limit $\Gamma_{n_2}(A) = \lim_{n_1 \to \infty} \Gamma_{n_2,n_1}(A)$.

If $\Xi_{\mathrm{class}}(A) = 1$ then for large n_2, $l^1_{n_1}$ must eventually be in $J^2_{n_2}$ and hence $\Gamma_{n_2}(A) = 1$. It is also clear that if $\Gamma_{n_2}(A) = 1$ then $l^1_{n_1}$ converges to a positive constant, which implies $\Xi_{\mathrm{class}}(A) = 1$. If $\Xi_{\mathrm{class}}(A) = 2$ then for large n_2, $l^m_{n_1}$ eventually lies in $J^2_{n_2}$ for some $2 \leq m \leq n_2$, but $l^1_{n_1}$ eventually in $J^1_{n_2}$. It follows that $\Gamma_{n_2}(A) = 2$. If $\Gamma_{n_2}(A) = 2$, then we know that there exists some $l^m_{n_1}$ convergent to $l \geq 1/n_2$ and hence we know $\Xi_{\mathrm{class}}(A)$ is either 1 or 2.

Now suppose that $\Xi_{\mathrm{class}}(A) = 3$, then for fixed n_2 and any $1 \leq m \leq n_2$, $l^m_{n_1}$ eventually lies in $J^1_{n_2}$ and hence our lowest level of the tower must eventually depend on whether $b_{n_2,n_1}(A) \geq 1/n_2$. From Lemma 6.3.1, $b_{n_2}(A) = c_{n_2}(A) = 1/n_2$ for large n_2. It follows that for large n_2, $b_{n_2}(A) \geq 1/n_2$ for all n_1 and

$\Gamma_{n_2}(A) = 3$. Furthermore, if $\Gamma_{n_2}(A) = 3$ then we know that $c_{n_2}(A) \geq b_{n_2}(A) \geq 1/n_2$, which implies $\Xi_{\text{class}}(A) \neq 4$. Finally, note that if $\Xi_{\text{class}}(A) = 4$ but there exists n_2 with $\Gamma_{n_2}(A) \neq 4$ then the above implies the contradiction $\Xi_{\text{class}}(A) \neq 4$. The above imply Γ_{n_2,n_1} realises the Π_2^A classification. □

6.4 Numerical Example for Discrete Spectra

Although it is hard to analyse the convergence of a height two tower, we can take advantage of the extra structure in this problem. The algorithm constructed in Theorem 6.1.1, referred to as `DiscreteSpec` in this section, computes $\Gamma_{n_2,n_1}(A)$ such that $\lim_{n_1\rightarrow\infty} \Gamma_{n_2,n_1}(A)$ is a finite subset of $\mathrm{Sp}_d(A)$. Furthermore, for each $z \in \mathrm{Sp}_d(A)$, there is at most one point in $z_{n_1} \in \Gamma_{n_2,n_1}(A)$ approximating z. We can use the methods of Chapter 3 (`DistSpec`) to gain an error bound of $\mathrm{dist}(z_{n_1}, \mathrm{Sp}(A))$, which, for large n_1, will be equal to $|z - z_{n_1}|$ since z is an isolated point of $\mathrm{Sp}(A)$. As we increase n_2, more and more of the discrete spectrum (in general portions nearer the essential spectrum) are approximated.

Our example is the almost Mathieu operator on $l^2(\mathbb{Z})$, given by

$$(H_\alpha x)_n = x_{n-1} + x_{n+1} + 2\lambda\cos(2\pi n\alpha + \nu)x_n, \quad \lambda = 1 \text{ (critical coupling)}.$$

The case of $\lambda = 1$ was studied in Hofstadter's classic paper [Hof76] (Hofstadter butterfly). The Hamiltonian represents a crystal electron in a uniform magnetic field, and the spectrum can be interpreted as the allowed energies of the system. For rational choices of α, the operator is periodic with purely absolutely continuous spectrum depending on ν. For irrational α, the spectrum is a Cantor set and does not depend on ν. Hence it follows that there is no discrete spectrum. In general, we cannot work with infinite precision, so approximate irrational α by rational approximations. We choose to work with $\nu = 0$ but found similar results for other values. To generate a discrete spectrum, we add a perturbation of the potential of the form

$$V(n) = V_n/(|n| + 1), \tag{6.4.1}$$

where V_n are independent and uniformly distributed in $[-2, 2]$. The perturbation is compact so preserves the essential spectrum, allowing us to test the algorithm. This type of problem is well-studied in the more general setting of Jacobi operators [Tes00, HS02], and physically models defects in the crystal.

Figure 6.1 shows a typical result for a realisation of the random potential. The figure shows the output of finite section and the algorithm of Chapter 3 (with a uniform error bound of 10^{-2}) for computing the total spectrum. We have also shown the output of `DiscreteSpec`, which separates the discrete spectrum from the essential spectrum. For each α we took n_2 large enough (obtained by comparing with the output of the height two tower for computing the essential spectrum) for expected limit inclusions

$$\Gamma_{n_2}(A) \subset \mathrm{Sp}_d(A) \subset \Gamma_{n_2}(A) + B_{0.01}(0). \tag{6.4.2}$$

Recall that $\Gamma_{n_2}(A) \subset \mathrm{Sp}_d(A)$ always holds and taking n_2 larger caused sharper inclusion bounds on the right-hand side of (6.4.2). Additionally, we confirmed that (6.4.2) does indeed hold by using the height one tower to compute the spectrum (Chapter 3) with and without the random potential. Note that it is difficult to detect spectral pollution when using finite section with the additional perturbation (6.4.1). In contrast, `DiscreteSpec` computes the discrete spectrum without spectral pollution and allows us to separate the discrete spectrum from the essential spectrum.

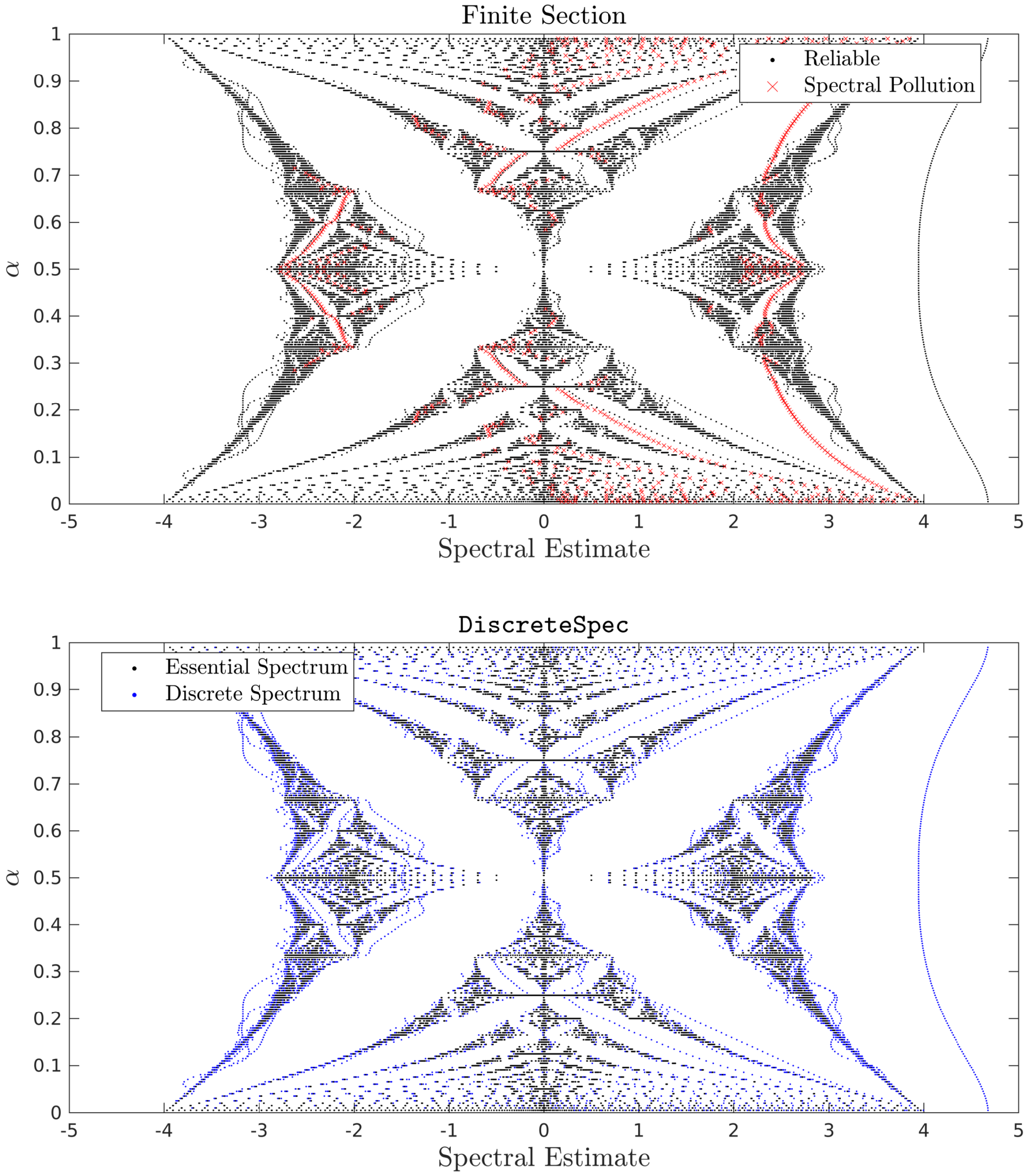


Figure 6.1: Top: Output of finite section. Spectral pollution detected by the algorithm of Chapter 3 is shown as red crosses. Bottom: Output of `DiscreteSpec` and the splitting into the essential spectrum and the discrete spectrum. The output captures the discrete spectrum down to a distance ≈ 0.01 away from the essential spectrum, which can be made smaller for larger n_2.

Chapter 7

Geometric Features and Detecting Finite Section Failure

In this chapter, based on [Colns], we address certain geometric features of the spectrum. We begin with some remarks on the finite section method, the most common approach to computing spectra. A highlight of this chapter is the proof that computing an error flag for finite section is harder than computing the spectrum itself (the problem solved in Chapter 3). This also settles the problem of computing or detecting gaps in the essential spectrum of self-adjoint operators, which has received considerable attention in the community. Furthermore, we classify various types of spectral radii, polynomial operator norms and capacity (which is useful for the analysis of Krylov numerical methods) in the SCI hierarchy. Even in the simplest case of computing the usual spectral radius, the only previous computational results are for normal operators (where the spectral radius is equal to the operator norm). In the non-normal case, this becomes a highly non-trivial problem, requiring three limits in the general case for the class of bounded operators on $l^2(\mathbb{N})$.

7.1 The Finite Section Method and when it fails

To motivate parts of this chapter, we begin with some brief remarks on the finite section method, the most common approach to approximate spectra (which, while successful for many problems, can also fail catastrophically). There has been considerable attention towards methods that detect gaps in the essential spectrum (spectral gaps) and eigenvalues within these gaps for self-adjoint operators [RS78, Kla80, Dav98, ZJ00, BBG00, CL90, LS14]. When computing spectra via the finite section method, it is well-known that spurious eigenvalues (spectral pollution) can occur anywhere within these gaps (see [LS09, Mar10] and the theorems below). There is a large literature that studies the precise nature of spectral pollution and possible ways to avoid it. This is an issue in applied areas such as computational chemistry, elasticity, electromagnetism and hydrodynamics [DG81, SH84, LS09, STY$^+$04, JWP96]. The computation is often done with finite element, finite difference or spectral methods by discretising the operator on a suitable finite-dimensional space, and then using algorithms for finite-dimensional matrix eigenvalue problems on the discretised operator [Rap77, RSHSPV97, BBG00, BDG99, BP06, BCJ09, ABP06, Zha07, BHP07, BPW09, BBG13, CW13]. Related to this is a more subtle issue, namely, that most numerical methods for eigenvalue problems come with convergence rates (often with hidden constants) and it is common knowledge that only a small portion of numerical eigenvalues are reliable. However, this knowledge is typically only qualitative rather than quantitative, and it is not clear in general what portion of the computation can be trusted (even when a method converges) [WT88, Zha15]. In other words, how do we know that an eigenvalue or portion of the spectrum is resolved?

To state our theorems in this chapter, we recall the definition of the essential numerical range:

$$W_e(A) = \bigcap_{K \text{ compact}} \mathrm{cl}(W(A+K)),$$

where $W(A) = \{\langle Ax, x\rangle : \|x\| = 1\}$ is the usual numerical range. If A is hyponormal ($A^*A - AA^* \geq 0$) then $W_e(A)$ is the convex hull of the essential spectrum [Sal72]. We also recall two theorems:

Theorem 7.1.1 ([Pok79])**.** *Let $A \in \mathcal{B}(\mathcal{H})$ and $\{P_n\}$ be a sequence of finite-dimensional projections converging strongly to the identity. Suppose that $S \subset W_e(A)$. Then there exists a sequence $\{Q_n\}$ of finite-dimensional projections such that $P_n < Q_n$ (so $Q_n \to I$ strongly) and*

$$d_{\mathrm{H}}(\mathrm{Sp}(A_n) \cup S, \mathrm{Sp}(\tilde{A}_n)) \to 0, \quad n \to \infty,$$

where

$$A_n = P_n A|_{P_n\mathcal{H}}, \qquad \tilde{A}_n = Q_n A|_{Q_n\mathcal{H}}$$

and d_{H} denotes the Hausdorff distance.

Theorem 7.1.2 ([Pok79])**.** *Let $A \in \mathcal{B}(\mathcal{H})$ and $\{P_n\}$ be a sequence of finite-dimensional projections converging strongly to the identity. If $\lambda \notin W_e(A)$ then $\lambda \in \mathrm{Sp}(A)$ if and only if*

$$\mathrm{dist}(\lambda, \mathrm{Sp}(P_n A|_{P_n\mathcal{H}})) \longrightarrow 0, \qquad n \to \infty.$$

These theorems say that the failure of the finite section method is confined to the essential numerical range and can be arbitrarily bad on $W_e(A)\backslash\mathrm{Sp}(A)$.[1] This is one of the key results motivating the quest for an algorithm that detects gaps in the essential spectrum of self-adjoint operators (in this case, these gaps correspond exactly to $W_e(A)\backslash\mathrm{Sp}(A)$).

7.2 The Set-up

Throughout this chapter and the next, A will be a bounded operator on $l^2(\mathbb{N})$ realised as a matrix with respect to the canonical basis. By a choice of basis we can, as in previous chapters, deal with arbitrary separable Hilbert spaces.

There are two basic natural sets of information that we allow our algorithms to read when computing spectral properties of A. The first is the set of evaluation functions Λ_1 consisting of the family of all functions $f^1_{i,j} : A \mapsto \langle Ae_j, e_i\rangle$, $i, j \in \mathbb{N}$, which provide the entries of the matrix representation of A with respect to the canonical basis $\{e_i\}_{i\in\mathbb{N}}$. The second, which we denote by Λ_2, is the family Λ_1 together with all functions $f^2_{i,j} : A \mapsto \langle Ae_j, Ae_i\rangle$ and $f^3_{i,j} : A \mapsto \langle A^*e_j, A^*e_i\rangle$, $i, j \in \mathbb{N}$, which provide the entries of the matrix representation of A^*A and AA^* with respect to the canonical basis $\{e_i\}_{i\in\mathbb{N}}$. In general, the classification of a computational problem in the SCI hierarchy depends on the evaluation set Λ. We

[1] In the non-normal case it is possible for finite section to not capture all of the spectrum - parts of the spectrum may be unattainable. This is distinct from spectral pollution. Theorem 7.1.1 says that, up to a different choice of projections, this can be avoided on $W_e(A)$.

have included Λ_2 in these two chapters since it is natural for problems posed in variational form. When considering classes with functions f (and $\{c_n\}$) and g as in (3.1.1) and (3.1.2), we will add these to the relevant evaluation set and, with the usual abuse of notation, still use the notation Λ_i. A small selection of the problems also require additional information, such as when testing if a set intersects a spectral set. However, any changes to Λ_i will be pointed out where appropriate.

7.3 Main Results

7.3.1 Spectral radii, operator norms and capacity of spectrum

The spectral radius $r(A)$ of a bounded operator A is the supremum of the absolute values of member of the spectrum (which is attained). Let Ω_{N} denote the class of normal operators in Ω_{B} and Ω_{D} denote the self-adjoint diagonal operators in Ω_{N}. We also denote by Ω_f the class of operators in Ω_{B} with dispersion bounded by f (see §3.1.1). Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ be an increasing function such that g maps $[0, \infty)$ onto itself continuously and strictly monotonously. Let Ω_g be the class of bounded operators with

$$\|R(z, A)\|^{-1} \geq g(\mathrm{dist}(z, \mathrm{Sp}(A))), \tag{7.3.1}$$

for $z \in \mathbb{C}$. Note that such a g is always guaranteed to exist, however, the classification in the SCI hierarchy depends on whether one knows an estimate for g or not. For example, in the self-adjoint and normal cases $g(x) = x$ is the trivial choice of g. Operators with $g(x) = x$ are known as G_1 in the operator theory literature and include the well-studied class of hyponormal operators [Put79]. It is known that if A is G_1 then: if $\mathrm{Sp}(A)$ is real then A is self-adjoint [Nie62], if $\mathrm{Sp}(A)$ is contained in the unit circle then A is unitary [Don63], and if $\mathrm{Sp}(A)$ is finite then A is normal [Sta65].

We let $\Xi_r(A) := r(A)$. Our proofs show that the computational problem of the operator norm or numerical radius of any $A \in \Omega_{\mathrm{B}}$ lies in Σ_1^A. Hence we can easily get an upper bound (that may not be sharp) for $\Xi_r(A)$ in one limit. If an operator lies in Ω_g with $g(x) = x$, then it is well-known that the convex hull of the spectrum is equal to the closure of the numerical range (the operator is convexoid) [Orl64] and hence the computational problem lies in Σ_1^A. One might expect that the computation of $\Xi_r(A)$ is strictly easier than that of the spectrum, particularly in light of Gelfand's famous formula $\Xi_r(A) = \lim_{n\to\infty} \|A^n\|^{\frac{1}{n}}$. However, the following shows that this intuition is false in general, and only occurs if an operator is convexoid. Controlling the resolvent via a function g as in (7.3.1) makes the problem easier than the general Ω_{B}, but is not sufficient to reduce the SCI of the problem to 1.

Theorem 7.3.1. *Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ be a strictly increasing, continuous function that vanishes only at 0 with $\lim_{x\to\infty} g(x) = \infty$. Suppose also that for some $\delta \in (0, 1)$ it holds that $g(x) \leq (1 - \delta)x$. Then:*

$$\begin{aligned}
&\Delta_1^G \not\ni \{\Xi_r, \Omega_{\mathrm{D}}, \Lambda_1\} \in \Sigma_1^A, && \Delta_1^G \not\ni \{\Xi_r, \Omega_{\mathrm{N}}, \Lambda_1\} \in \Sigma_1^A, && \Delta_1^G \not\ni \{\Xi_r, \Omega_f \cap \Omega_g, \Lambda_1\} \in \Sigma_1^A,\\
&\Delta_2^G \not\ni \{\Xi_r, \Omega_g, \Lambda_1\} \in \Sigma_2^A, && \Delta_2^G \not\ni \{\Xi_r, \Omega_f, \Lambda_1\} \in \Pi_2^A, && \Delta_3^G \not\ni \{\Xi_r, \Omega_{\mathrm{B}}, \Lambda_1\} \in \Pi_3^A.
\end{aligned}$$

When considering the evaluation set Λ_2, the only changes are the following classifications:

$$\Delta_1^G \not\ni \{\Xi_r, \Omega_g, \Lambda_2\} \in \Sigma_1^A, \qquad \Delta_2^G \not\ni \{\Xi_r, \Omega_{\mathrm{B}}, \Lambda_2\} \in \Pi_2^A.$$

Next, we consider the essential spectral radius. Define the essential spectrum of $A \in \Omega_{\mathrm{B}}$ as

$$\mathrm{Sp}_{\mathrm{ess}}(A) = \bigcap_{B \in \Omega_C} \mathrm{Sp}(A + B),$$

where Ω_C denotes the class of compact operators. The essential spectral radius, $\Xi_{er}(A)$, is simply the supremum of the absolute values over $\mathrm{Sp}_{\mathrm{ess}}(A)$.

Theorem 7.3.2. *We have the following classifications for* $i = 1, 2$*:*

$$\Delta_2^G \not\ni \{\Xi_{er}, \Omega_{\mathrm{D}}, \Lambda_i\} \in \Pi_2^A, \qquad \Delta_2^G \not\ni \{\Xi_{er}, \Omega_{\mathrm{N}}, \Lambda_i\} \in \Pi_2^A, \qquad \Delta_2^G \not\ni \{\Xi_{er}, \Omega_f, \Lambda_i\} \in \Pi_2^A.$$

For general operators,

$$\Delta_3^G \not\ni \{\Xi_{er}, \Omega_{\mathrm{B}}, \Lambda_1\} \in \Pi_3^A, \quad \Delta_2^G \not\ni \{\Xi_{er}, \Omega_{\mathrm{B}}, \Lambda_2\} \in \Pi_2^A.$$

As two final problems in this section, given a polynomial p (of degree at least two), we consider the problem of computing $\Xi_{r,p} = \|p(A)\|$ and the capacity of the spectrum defined by

$$\Xi_{cap}(A) = \inf_{\text{monic polynomial } p} \|p(A)\|^{1/\deg(p)}.$$

Operators with $\Xi_{cap}(A) = 0$ are known as quasialgebraic, and a theorem of Halmos shows that this definition of capacity agrees with the usual potential-theoretic definition of capacity of the set $\mathrm{Sp}(A)$ [Hal71]. This quantity is of particular interest in Krylov methods where, for instance, it is related to the speed of convergence[2] [Nev93, Nev95]. Vaguely speaking, the capacity is a measure of the size of $\mathrm{Sp}(A)$ (a measure of its ability to hold electrical charge as opposed to volume). We will also see some other measures of size in Chapter 8 when considering the Lebesgue measure and fractal dimensions of $\mathrm{Sp}(A)$.

Theorem 7.3.3. *We have the following classifications for* $i = 1, 2$ *and* $\hat{\Omega} = \Omega_{\mathrm{D}}, \Omega_f$*:*

$$\Delta_1^G \not\ni \{\Xi_{r,p}, \hat{\Omega}, \Lambda_i\} \in \Sigma_1^A, \qquad \Delta_2^G \not\ni \{\Xi_{cap}, \hat{\Omega}, \Lambda_i\} \in \Pi_2^A.$$

Whereas for $\tilde{\Omega} = \Omega_{\mathrm{N}}, \Omega_g$ *or* Ω_{B}*:*

$$\Delta_2^G \not\ni \{\Xi_{r,p}, \tilde{\Omega}, \Lambda_1\} \in \Sigma_2^A, \qquad \Delta_3^G \not\ni \{\Xi_{cap}, \tilde{\Omega}, \Lambda_1\} \in \Pi_3^A$$
$$\Delta_1^G \not\ni \{\Xi_{r,p}, \tilde{\Omega}, \Lambda_2\} \in \Sigma_1^A, \qquad \Delta_2^G \not\ni \{\Xi_{cap}, \tilde{\Omega}, \Lambda_2\} \in \Pi_2^A.$$

7.3.2 Gaps in essential spectra and detecting algorithm failure for finite section

We will show that detecting whether spectral pollution can occur is strictly harder than computing the spectrum for self-adjoint operators. In other words, detecting the failure of the finite section method is strictly harder than the problem it was designed to solve!

Let $\Xi_{we}(A) = W_e(A)$. For a given open set U in $\mathbb{F}$ ($\mathbb{F}$ being $\mathbb{C}$ or $\mathbb{R}$), let $\Xi_{poll}^{\mathbb{F}}$ be the decision problem

$$\Xi_{poll}^{\mathbb{F}}(A, U) = \begin{cases} 1, & \text{if } \mathrm{cl}\,(U) \cap (W_e(A) \backslash \mathrm{Sp}(A)) \neq \emptyset \\ 0, & \text{otherwise.} \end{cases}$$

$\Xi_{poll}^{\mathbb{F}}$ decides whether spectral pollution can occur on the closed set $\mathrm{cl}\,(U)$, which is assumed to have non-empty interior. For the self-adjoint case (where $\mathbb{F} = \mathbb{R}$), this is equivalent to asking whether there exists a point in the open set U which also lies in a gap of the essential spectrum. To incorporate U into Λ_i, we allow access to a countable number of open balls $\{U_m\}_{m \in \mathbb{N}}$ whose union is U. If $\mathbb{F}$ is $\mathbb{R}$ then each U_m is of the form (a_m, b_m) with $a_m, b_m \in \mathbb{Q} \cup \{\pm\infty\}$, whereas if $\mathbb{F}$ is $\mathbb{C}$ then each U_m is equal to $D_{r_m}(z_m)$ (the open ball of radius r_m centred at z_m) with $r_m \in \mathbb{Q}_+$ and $z_m \in \mathbb{Q} + i\mathbb{Q}$. We add pointwise evaluations of $\{(a_m, b_m)\}$ or $\{(r_m, z_m)\}$ to Λ_i. Let Ω_{SA} denote the class of bounded self-adjoint operators.

[2] This is an idealisation since the capacity studies operator norms while true Krylov processes look at $p(A)x$ with one or several vectors x. However, from local spectral theory (e.g. [M92]) it follows that generically the asymptotic speeds are the same.

Theorem 7.3.4. *Let $\Omega = \Omega_{\mathrm{N}}, \Omega_{\mathrm{SA}}$ or Ω_{B} and let $i = 1, 2$. Then*

$$\Delta_2^G \not\ni \{\Xi_{we}, \Omega, \Lambda_i\} \in \Pi_2^A.$$

Furthermore, for $i = 1, 2$ the following classifications hold, valid also if we restrict to the case $U = U_1$ or to $U = U_1 = \mathbb{F}$:

$$\Delta_3^G \not\ni \{\Xi_{poll}^{\mathbb{R}}, \Omega_{\mathrm{SA}}, \Lambda_i\} \in \Sigma_3^A, \qquad \Delta_3^G \not\ni \{\Xi_{poll}^{\mathbb{C}}, \Omega_{\mathrm{B}}, \Lambda_i\} \in \Sigma_3^A.$$

Remark 7.3.5. *One can show that $\{\mathrm{Sp}(\cdot), \Omega_{\mathrm{SA}}, \Lambda_1\} \in \Sigma_2^A$ and $\{\mathrm{Sp}(\cdot), \Omega_{\mathrm{SA}}, \Lambda_2\} \in \Sigma_1^A$. Hence determining $\Xi_{poll}^{\mathbb{R}}$ is strictly harder than the spectral computational problem and requires two extra limits if $\Lambda = \Lambda_2$. Even in the general case, $\{\mathrm{Sp}(\cdot), \Omega_{\mathrm{B}}, \Lambda_2\} \in \Pi_2^A$ and hence the spectral problem is strictly easier. The proofs also make clear that we get the same classification of $\Xi_{poll}^{\mathbb{F}}$ for other classes such as Ω_{N}, Ω_g etc.*

7.4 Proofs of Theorems in §7.3.1

We begin with the proof of Theorem 7.3.1, dealing with the evaluation set Λ_1 first. Suppose that $\tilde{\Gamma}_{n_k,...,n_1}$ is a Π_k^A tower of algorithms to compute the spectrum of a class of operators, where the output is a finite set for each $n_1, ..., n_k$. It is then clear that

$$\Gamma_{n_k,...,n_1}(A) = \sup_{z \in \tilde{\Gamma}_{n_k,...,n_1}(A)} |z| + \frac{1}{2^{n_k}}$$

provides a Π_k^A tower of algorithms for the spectral radius. Strictly speaking, the above may not be an arithmetic tower owing to the absolute value. But it can be approximated to arbitrary precision (from above say), the error of which can be absorbed in the first limit. In what follows, we always assume this is done without further comment. Similarly if $\tilde{\Gamma}_{n_k,...,n_1}$ provides a Σ_k^A tower of algorithms for the spectrum (output a finite set for each $n_1, ..., n_k$),

$$\Gamma_{n_k,...,n_1}(A) = \sup_{z \in \tilde{\Gamma}_{n_k,...,n_1}(A)} |z| - \frac{1}{2^{n_k}}$$

provides a Σ_k^A tower of algorithms for the spectral radius. If we only have a height k tower with no Σ_k or Π_k type error control for the spectrum, then taking the supremum of absolute values shows we get a height k tower for the spectral radius.

The fact that $\{\Xi_r, \Omega_{\mathrm{D}}\} \in \Sigma_1^A$, $\{\Xi_r, \Omega_f \cap \Omega_g\} \in \Sigma_1^A$, $\{\Xi_r, \Omega_g\} \in \Sigma_2^A$, $\{\Xi_r, \Omega_f\} \in \Pi_2^A$ and $\{\Xi_r, \Omega_{\mathrm{B}}\} \in \Pi_3^A$ hence follow from Chapter 3 and the results of [BACH$^+$20]. It is clear that $\{\Xi_r, \Omega_{\mathrm{D}}\} \notin \Delta_1^G$ and this also shows that $\{\Xi_r, \Omega_{\mathrm{N}}\} \notin \Delta_1^G$ and $\{\Xi_r, \Omega_f \cap \Omega_g\} \notin \Delta_1^G$. Hence, we must show the positive result that $\{\Xi_r, \Omega_{\mathrm{N}}\} \in \Sigma_1^A$ and prove the lower bounds $\{\Xi_r, \Omega_g\} \notin \Delta_2^G$, $\{\Xi_r, \Omega_f\} \notin \Delta_2^G$ and $\{\Xi_r, \Omega_{\mathrm{B}}\} \notin \Delta_3^G$.

Proof of Theorem 7.3.1 for Λ_1. Throughout this proof we use the evaluation set Λ_1 (dropped from notation for convenience).

Step 1: $\{\Xi_r, \Omega_{\mathrm{N}}\} \in \Sigma_1^A$. Recall that the spectral radius of a normal operator $A \in \Omega_{\mathrm{B}}$ is equal to its operator norm. Consider the finite section matrices $P_nAP_n \in \mathbb{C}^{n\times n}$. It is straightforward to show that

$$\|P_nAP_n\| \uparrow \|A\| \quad \text{as } n \to \infty.$$

The norm $\|P_nAP_n\|$ is the square root of the largest eigenvalue of the semi-positive definite self-adjoint matrix $(P_nAP_n)^*(P_nAP_n)$. This can be estimated from below to an accuracy of $1/n$, which then yields a Σ_1^A algorithm for $\{\Xi_r, \Omega_{\mathrm{N}}\}$.

Step 2: $\{\Xi_r, \Omega_g\} \notin \Delta_2^G$. Recall that we assumed the existence of a $\delta \in (0,1)$ such that $g(x) \leq (1-\delta)x$. Let $\epsilon > 0$, then it is easy to see that the matrices

$$S_{\pm}(\epsilon) = \begin{pmatrix} 1 & 0 \\ \pm\epsilon & 1 \end{pmatrix}$$

have norm bounded by $1+\epsilon+\epsilon^2$ and are clearly inverse of each other. Choose ϵ small such that $(1+\epsilon+\epsilon^2)^2 \leq 1/(1-\delta)$. If $B \in \mathbb{C}^{2\times 2}$ is normal, it follows that $\hat{B} := S_+(\epsilon)BS_-(\epsilon)$ lies in Ω_g and has the same spectrum as B. We choose

$$\hat{B} = S_+(\epsilon)\begin{pmatrix} 1 & -\epsilon \\ -\epsilon & 0 \end{pmatrix} S_-(\epsilon) = \begin{pmatrix} 1+\epsilon^2 & -\epsilon \\ \epsilon^3 & -\epsilon^2 \end{pmatrix}.$$

The crucial property of $\hat{B}$ is that the first entry $1+\epsilon^2$ is strictly greater in magnitude than the two eigenvalues $(1 \pm \sqrt{1+4\epsilon^2})/2$.

Now suppose for a contradiction that a height one tower, Γ_n, solves the problem. We will gain a contradiction by showing that $\Gamma_n(A)$ does not converge for an operator of the form,

$$A = \bigoplus_{r=1}^{\infty} A_{l_r}, \quad A_m := \begin{pmatrix} 1+\epsilon^2 & & & & -\epsilon \\ & 0 & & & \\ & & \ddots & & \\ & & & 0 & \\ \epsilon^3 & & & & -\epsilon^2 \end{pmatrix} \in \mathbb{C}^{m\times m},$$

where we only consider $l_k \geq 3$. Each A_m is unitarily equivalent to the matrix $\hat{B} \oplus 0 \in \mathbb{C}^{m\times m}$ and has spectrum equal to $\{0, (1\pm\sqrt{1+4\epsilon^2})/2\}$. Any A of the above form is unitarily equivalent to a direct sum of an infinite number of $\hat{B}$'s and the zero operator and hence lies in Ω_g. Now suppose that $l_1, ..., l_k$ have been chosen and consider the operator

$$B_k = A_{l_1} \oplus ... \oplus A_{l_k} \oplus C, \quad C = \mathrm{diag}\{1+\epsilon^2, 0, ...\}.$$

The spectrum of B_k is $\{0, (1\pm\sqrt{1+4\epsilon^2})/2, 1+\epsilon^2\}$ and hence there exist $\eta > 0$ and $n(k) \geq k$ such that $\Gamma_{n(k)}(B_k) > (1+\sqrt{1+4\epsilon^2})/2+\eta$. But $\Gamma_{n(k)}(B_k)$ can only depend on the evaluations of the matrix entries $\{B_k\}_{ij} = \langle B_k e_j, e_i\rangle$ with $i,j \leq N(B_k, n(k))$ (as well as evaluations of the function g) into account. If we choose $l_{k+1} > N(B_k, n(k))$ then by the assumptions in Definition 2.1.1, $\Gamma_{n(k)}(A) = \Gamma_{n(k)}(B_k) > (1+\sqrt{1+4\epsilon^2})/2+\eta$. But $\Gamma_n(A)$ must converge to $(1+\sqrt{1+4\epsilon^2})/2$, a contradiction.

Step 3: $\{\Xi_r, \Omega_f\} \notin \Delta_2^G$. Suppose for a contradiction that a height one tower, Γ_n, solves the problem. We will gain a contradiction by showing that $\Gamma_n(A)$ does not converge for an operator of the form,

$$A = \bigoplus_{r=1}^{\infty} C_{l_r} \oplus A_{l_r}, \quad A_m := \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & & 1 \\ & & & & 0 \end{pmatrix} \in \mathbb{C}^{m\times m}, \quad C_m = \mathrm{diag}\{0,0,...,0\} \in \mathbb{C}^{m\times m},$$

where we assume that $l_r \geq r$ to ensure that the spectrum of A is equal to the unit disc $B_1(0)$. Note that the function $f(n) = n+1$ will do for the bounded dispersion with $c_n = 0$. Now suppose that $l_1, ..., l_k$ have been chosen and consider the operator

$$B_k = \big(C_{l_1} \oplus A_{l_1}\big) \oplus ... \oplus \big(C_{l_k} \oplus A_{l_k}\big) \oplus C, \quad C = \mathrm{diag}\{0, 0, ...\}.$$

The spectrum of B_k is $\{0\}$ and hence there exists $n(k) \geq k$ such that $\Gamma_{n(k)}(B_k) < 1/4$. But $\Gamma_{n(k)}(B_k)$ can only depend on the evaluations of the matrix entries $\{B_k\}_{ij} = \langle B_k e_j, e_i\rangle$ with $i, j \leq N(B_k, n(k))$ (as well as evaluations of the function f) into account. If we choose $l_{k+1} > N(B_k, n(k))$ then by the assumptions in Definition 2.1.1, $\Gamma_{n(k)}(A) = \Gamma_{n(k)}(B_k) < 1/4$. But $\Gamma_n(A)$ must converge to 1, a contradiction.

Step 4: $\{\Xi_r, \Omega_\mathrm{B}\} \notin \Delta_3^G$. Suppose as a contradiction that Γ_{n_2,n_1} is a height two (general) tower and without loss of generality assume it to be non-negative. Let $(\mathcal{M}, d)$ be the space $[0, 1]$ with the usual metric, let $\tilde{\Omega}$ denote the collection of all infinite matrices $\{a_{i,j}\}_{i,j\in\mathbb{N}}$ with entries $a_{i,j} \in \{0, 1\}$ and recall the problem function

$$\tilde{\Xi}_1(\{a_{i,j}\}) : \text{ Does } \{a_{i,j}\} \text{ have a column containing infinitely many non-zero entries?}$$

It was shown in Theorem 2.3.7 that $\mathrm{SCI}(\tilde{\Xi}_1, \tilde{\Omega})_G = 3$. We will gain a contradiction by using the supposed height two tower to solve $\{\tilde{\Xi}_1, \tilde{\Omega}\}$.

Without loss of generality, identify Ω_B with $\mathcal{B}(X)$ where $X = \bigoplus_{j=1}^{\infty} X_j$ in the l^2-sense with $X_j = l^2(\mathbb{N})$. Now let $\{a_{i,j}\} \in \tilde{\Omega}$ and define $B_j \in \mathcal{B}(X_j)$ with the matrix representation

$$(B_j)_{k,i} = \begin{cases} 1, & \text{if } k = i \text{ and } a_{k,j} = 0 \\ 1, & \text{if } k < i \text{ and } a_{l,j} = 0 \text{ for } k < l < i \\ 0, & \text{otherwise } 0 \leq n \leq 1. \end{cases}$$

Let $\mathcal{I}_j$ be the index set of all i where $a_{i,j} = 1$. B_j acts as a unilateral shift on $\mathrm{cl}\,(\mathrm{span})\,\{e_k : k \in \mathcal{I}_j\}$ and the identity on its orthogonal complement. It follows that

$$\mathrm{Sp}(B_j) = \begin{cases} 1, & \text{if } \mathcal{I}_j = \emptyset \\ \{0, 1\}, & \text{if } \mathcal{I}_j \text{ is finite and non-empty} \\ \mathbb{D} \quad \text{(the unit disc)}, & \text{if } \mathcal{I}_j \text{ is infinite.} \end{cases}$$

For the matrix $\{a_{i,j}\}$ define $A \in \Omega_\mathrm{B}$ by

$$A = \bigoplus_{j=1}^{\infty} \left(B_j - \frac{1}{2} I_j\right),$$

where I_j denotes the identity operator on $\mathbb{C}^{j\times j}$, then $\mathrm{Sp(A)} = \mathrm{cl}\left(\cup_{j=1}^{\infty}\mathrm{Sp}(B_j)\right) - \frac{1}{2}$.

Hence we see that

$$\Xi_r(A) = \begin{cases} \frac{1}{2}, & \text{if } \tilde{\Xi}_1(\{a_{i,j}\}) = 0 \\ \frac{3}{2}, & \text{if } \tilde{\Xi}_1(\{a_{i,j}\}) = 1. \end{cases}$$

We then set $\tilde{\Gamma}_{n_2,n_1}(\{a_{i,j}\}) = \min\{\max\{\Gamma_{n_2,n_1}(A) - 1/2, 0\}, 1\}$. It is clear that this defines a generalised algorithm mapping into $[0, 1]$. In particular, given N we can evaluate $\{A_{k,l} : k, l \leq N\}$ using only finitely many evaluations of $\{a_{i,j}\}$, where we can use a bijection between canonical bases of $l^2(\mathbb{N})$ and $\bigoplus_{j=1}^{\infty} X_j$ to view A as acting on $l^2(\mathbb{N})$. But then $\tilde{\Gamma}_{n_2,n_1}$ provides a height two tower for $\{\tilde{\Xi}_1, \tilde{\Omega}\}$, a contradiction. □

Proof of Theorem 7.3.1 for Λ_2. Here we prove the changes for Ξ_r when we consider the evaluation set Λ_2. It is clear that the classifications in Σ_1^A do not change. It is also easy to use the algorithms in Chapter 3 (now using Λ_2 to collapse the first limit and approximate γ_n) to prove $\{\Xi_r, \Omega_g, \Lambda_2\} \in \Sigma_1^A$. Similarly we can use the algorithm for the spectrum of operators in Ω_f for Ω_{B} using Λ_2 to collapse the first limit and hence $\{\Xi_r, \Omega_{\mathrm{B}}, \Lambda_2\} \in \Pi_2^A$. Since $\Omega_f \subset \Omega_{\mathrm{B}}$, it follows that we only need to prove $\{\Xi_r, \Omega_f, \Lambda_2\} \notin \Delta_2^G$. This is proven using the same example and a similar argument to step 3 of the proof of Theorem 7.3.1. □

Proof of Theorem 7.3.2. We begin by proving the results for Λ_1. For the lower bounds, it is enough to show that $\{\Xi_{er}, \Omega_{\mathrm{D}}, \Lambda_1\} \notin \Delta_2^G$ and $\{\Xi_{er}, \Omega_{\mathrm{B}}, \Lambda_1\} \notin \Delta_3^G$. For the upper bounds, we must show that $\{\Xi_{er}, \Omega_f, \Lambda_1\} \in \Pi_2^A$, $\{\Xi_{er}, \Omega_{\mathrm{B}}, \Lambda_1\} \in \Pi_3^A$ and $\{\Xi_{er}, \Omega_{\mathrm{N}}, \Lambda_1\} \in \Pi_2^A$. The lower bounds for Λ_2 follow from $\{\Xi_{er}, \Omega_{\mathrm{D}}, \Lambda_1\} \notin \Delta_2^G$ and for the upper bounds it is enough to prove $\{\Xi_{er}, \Omega_{\mathrm{B}}, \Lambda_2\} \in \Pi_2^A$.

Step 1: $\{\Xi_{er}, \Omega_{\mathrm{D}}, \Lambda_1\} \notin \Delta_2^G$. This is the same argument as in step 3 of the proof of Theorem 7.3.1, however now we replace A_m by $A_m = \mathrm{diag}\{1, 1, ..., 1\} \in \mathbb{C}^{m\times m}$ and use the fact that $\Xi_{er}(B_k) = 0$. It follows that given the proposed height one tower Γ_n and the constructed A, $\Xi_{er}(A) = 1$ but $\Gamma_{n(k)}(A) < 1/4$, the required contradiction.

Step 2: $\{\Xi_{er}, \Omega_{\mathrm{B}}, \Lambda_1\} \notin \Delta_3^G$. This is the same argument as step 4 of the proof of Theorem 7.3.1.

Step 3: $\{\Xi_{er}, \Omega_f, \Lambda_1\} \in \Pi_2^A$, $\{\Xi_{er}, \Omega_{\mathrm{B}}, \Lambda_1\} \in \Pi_3^A$ and $\{\Xi_{er}, \Omega_{\mathrm{B}}, \Lambda_2\} \in \Pi_2^A$. $\{\Xi_{er}, \Omega_f, \Lambda_1\} \in \Pi_2^A$ follows immediately from the existence of a Π_2^A tower of algorithms for the essential spectrum of operators in Ω_f proven in [BACH+20]. The output of this tower is a finite collection of rectangles with complex rational vertices, hence we can gain an approximation of the maximum absolute value over this output to any given precision. This can be used to construct a Π_2^A tower for $\{\Xi_{er}, \Omega_f, \Lambda_1\}$. Similarly, $\{\Xi_{er}, \Omega_{\mathrm{B}}, \Lambda_1\} \in \Pi_3^A$ follows from the Π_3^A tower of algorithms for $\{\mathrm{Sp}_{\mathrm{ess}}, \Omega_{\mathrm{B}}, \Lambda_1\}$ constructed in [BACH+20]. Finally, we can use Λ_2 to collapse the first limit of the algorithm for the essential spectrum in [BACH+20], giving a Π_2^A algorithm and this can be used to show $\{\Xi_{er}, \Omega_{\mathrm{B}}, \Lambda_2\} \in \Pi_2^A$.

Step 4: $\{\Xi_{er}, \Omega_{\mathrm{N}}, \Lambda_1\} \in \Pi_2^A$. A Π_2^A tower is constructed in the proof of Theorem 7.3.4 for the essential numerical range, $W_e(A)$, of normal operators (using Λ_1) and this outputs a finite collection of points. For normal operators A, $W_e(A)$ is the convex hull of the essential spectrum and hence $\sup_{z\in W_e(A)} |z|$ is equal to $\Xi_{er}(A)$. Hence a Π_2^A tower for $\{\Xi_{er}, \Omega_{\mathrm{N}}, \Lambda_1\}$ follows by taking the maximum absolute value over the tower for $W_e(A)$. □

Proof of Theorem 7.3.3. Some general remarks are in order to simplify the proof. First, note that given a height k arithmetical tower $\widehat{\Gamma}_{n_k,...,n_1}(\cdot, p)$ for $\Xi_{r,p}$ and a class Ω', we can build a Π_{k+1}^A tower for $\{\Xi_{cap}, \Omega'\}$ as follows. Let $p_1, p_2, ...$ be an enumeration of the monic polynomials with rational coefficients and $\tilde{\Gamma}_{n_k,...,n_1}(\cdot, p)$ be an approximation to $\left|\widehat{\Gamma}_{n_k,...,n_1}(\cdot, p)\right|^{1/\deg(p)}$ to accuracy $1/n_1$ using finitely many arithmetic operations and comparisons. Define

$$\Gamma_{n_{k+1},...,n_1}(A) = \min_{1\leq m\leq n_{k+1}} \tilde{\Gamma}_{n_k,...,n_1}(A, p_m).$$

The fact that this is a convergent Π_{k+1}^A tower is clear. This, together with inclusions of the considered classes of operators, means that to prove the positive results we only need to prove $\{\Xi_{r,p}, \Omega_f, \Lambda_1\} \in \Sigma_1^A$, $\{\Xi_{r,p}, \Omega_{\mathrm{B}}, \Lambda_1\} \in \Sigma_2^A$ and $\{\Xi_{r,p}, \Omega_{\mathrm{B}}, \Lambda_2\} \in \Sigma_1^A$. Likewise, for the negative results we only need to prove $\{\Xi_{cap}, \Omega_{\mathrm{D}}, \Lambda_2\} \notin \Delta_2^G$ (the fact that $\{\Xi_{r,p}, \Omega_{\mathrm{D}}, \Lambda_2\} \notin \Delta_1^G$ is obvious), $\{\Xi_{cap}, \Omega_{\mathrm{N}}, \Lambda_1\} \notin \Delta_3^G$ and $\{\Xi_{r,p}, \Omega_{\mathrm{N}}, \Lambda_2\} \notin \Delta_2^G$. We shall prove these results with Ω_{N} replaced by the class of self-adjoint bounded operators denoted by Ω_{SA}.

Remark 7.4.1 (Efficiently computing the capacity)**.** *Listing the monic polynomials with rational coefficients in the above proof is very inefficient. In practice, it is much better to split the domain of interest into* *intervals (or squares if in the complex plane, but we stick to the self-adjoint case in the following discussion). Suppose that each interval has dyadic endpoints and a diameter of 2^{-n_2} and that our operator is self-adjoint with known bounded dispersion. One can then apply Lemma 8.1.7 (denoting the index of that tower by n_1) to obtain an interval covering of the spectrum which will converge as $n_1 \to \infty$, modulo the possibility of isolated points of the spectrum located at the endpoints of the intervals. Since the capacity of a compact set is unaltered by adding finitely many points, we do not have to worry about the endpoints - the limit of the capacity of this covering as $n_1 \to \infty$ will be the capacity of a covering of the spectrum. As $n_2 \to \infty$, we can use the fact that capacity is right-continuous as a set function (for compact sets E_n, E with $E_n \downarrow E$, one has $\mathrm{cap}(E_n) \downarrow \mathrm{cap}(E)$) to obtain a Π_2^A algorithm. The point of this is that it reduces the computation of the resulting tower Γ_{n_2,n_1} to computing the capacity of finite unions of disjoint closed intervals in $\mathbb{R}$. In our numerical example, we made use of the method in [LSN17], which uses conformal mappings and can deal with thousands of intervals.*

Step 1: $\{\Xi_{r,p}, \Omega_f, \Lambda_1\} \in \Sigma_1^A$. The function f and sequence $\{c_n\}$ allows us to compute the matrix elements of $p(A)$ for any $A \in \Omega_f$ and polynomial p to arbitrary accuracy. We can then use the same argument as step 1 of the proof of Theorem 7.3.1, approximating $\|P_n p(A) P_n\|$ instead of $\|P_n A P_n\|$.

Step 2: $\{\Xi_{r,p}, \Omega_{\mathrm{B}}, \Lambda_1\} \in \Sigma_2^A$ and $\{\Xi_{r,p}, \Omega_{\mathrm{B}}, \Lambda_2\} \in \Sigma_1^A$. For the first result, we note that

$$\lim_{m\to\infty} \|P_n p(P_m A P_m) P_n\| = \|P_n p(A) P_n\|$$

and let $\Gamma_{n,m}(A,p)$ be an approximation of $\|P_n p(P_m A P_m) P_n\|$ to accuracy $1/m$, which can be computed in finitely many arithmetic operations and comparisons. To prove $\{\Xi_{r,p}, \Omega_{\mathrm{B}}, \Lambda_2\} \in \Sigma_1^A$, for any given $A \in \Omega_{\mathrm{B}}$ we can use Λ_2 to compute a function f_A and sequence $\{c_n(A)\}$ bounding the dispersion such that $A \in \Omega^{f_A}$ and use step 1.

Step 3: $\{\Xi_{cap}, \Omega_{\mathrm{SA}}, \Lambda_1\} \notin \Delta_3^G$. Suppose as a contradiction that Γ_{n_2,n_1} is a height two (general) tower for the problem and without loss of generality, assume it to be non-negative. Our strategy will be as in the proof of Theorem 7.3.1. Let $(\mathcal{M}, d)$ be the space $[0,1]$ with the usual metric, let $\tilde{\Omega}$ denote the collection of all infinite matrices $\{a_{i,j}\}_{i,j\in\mathbb{N}}$ with entries $a_{i,j} \in \{0,1\}$ and consider the problem function

$$\tilde{\Xi}_2(\{a_{i,j}\}) : \text{ Does } \{a_{i,j}\} \text{ have (only) finitely many columns with (only) finitely many 1's?}$$

Recall that it was shown in Theorem 2.3.7 that $\mathrm{SCI}(\tilde{\Xi}_2, \tilde{\Omega})_G = 3$. We will gain a contradiction by using the supposed height two tower to solve $\{\tilde{\Xi}_2, \tilde{\Omega}\}$. Without loss of generality, identify Ω_{SA} with self adjoint operators in $\mathcal{B}(X)$ where $X = \bigoplus_{j=1}^{\infty} X_j$ in the l^2-sense with $X_j = l^2(\mathbb{N})$. To proceed we need the following elementary lemma, which will be useful in constructing examples of spectral pollution.

Lemma 7.4.2. *Let $z_1, z_2, ..., z_k \in [-1,1]$ and let $a_j = \sqrt{1 - z_j^2}$ (say positive square root). Then the* *symmetric matrix*

$$B(z_1,...,z_k)=\left(\begin{array}{ccccc|ccccc}
z_1 & 0 & \cdots & & & a_1 & 0 & \cdots & & \\
0 & z_2 & 0 & \cdots & & 0 & a_2 & 0 & \cdots & \\
\vdots & 0 & \ddots & & & \vdots & 0 & \ddots & & \\
 & \vdots & & & & & \vdots & & & \\
 & & & & z_k & & & & & a_k \\ \hline
a_1 & 0 & \cdots & & & -z_1 & 0 & \cdots & & \\
0 & a_2 & 0 & \cdots & & 0 & -z_2 & 0 & \cdots & \\
\vdots & 0 & \ddots & & & \vdots & 0 & \ddots & & \\
 & \vdots & & & & & \vdots & & & \\
 & & & & a_k & & & & & -z_k
\end{array}\right)\in\mathbb{C}^{2k\times 2k}$$

has eigenvalues ± 1 *(repeated* k *times).*

Proof. By a change of basis, the above matrix is equivalent to a block diagonal matrix with blocks

$$\begin{pmatrix} z_j & a_j \\ a_j & -z_j \end{pmatrix}.$$

These blocks have eigenvalues $\{-1,1\}$. □

Now choose a sequence of rational numbers $\{z_j\}_{j\in\mathbb{N}}\in[-1,1]$ that is also dense in $[-1,1]$ and let $B_j=B(z_1,...,z_j)$. For each column of a given $\{a_{i,j}\}\in\tilde{\Omega}$, let the infinite matrix $C^{(j)}$ be defined as follows. If $k,l<j+1$ then $C^{(j)}_{kl}=z_k\delta_{k,l}$. Let $r(i)$ denote the row of the ith one of the column $\{a_{i,j}\}_{i\in\mathbb{N}}$ (with $r(i)=\infty$ if $\sum_m a_{m,j}<i$ and $r(0)=0$). If $r(i)<\infty$ then for $k\leq l$ define

$$C^{(j)}_{kl}=\begin{cases} a_p\delta_{k,l-(r(i)-r(i-1)-1)}, & p=1,...,j, l=r(i)+j\cdot(2i-1)+p-1\\ -z_p\delta_{k,l}, & p=1,...,j, l=r(i)+j\cdot(2i-1)+p-1\\ z_p\delta_{k,l}, & p=1,...,j, l=r(i)+2j\cdot i+p-1\\ 0, & \text{otherwise,}\end{cases}$$

and extend $C^{(j)}_{kl}$ below the diagonal to a symmetric matrix. The key property of this matrix is that if the column $\{a_{i,j}\}_{i\in\mathbb{N}}$ has infinitely many 1s, then its is unitarily equivalent to an infinite direct sum of infinitely many B_j together with the zero operator acting on some subspace (whose dimension is equal to the number of zeros in the column). In this case $\mathrm{Sp}(C^{(j)})=\{-1,1,0\}$ or $\{-1,1\}$. On the other hand, if $\{a_{i,j}\}_{i\in\mathbb{N}}$ has finitely many 1s, then $C^{(j)}$ is unitarily equivalent the direct sum of a finite number of B_j, the diagonal operator $\mathrm{diag}\{z_1,...,z_j\}$ and the zero operator acting on some subspace. In this case $\{z_1,...,z_j\}\subset\mathrm{Sp}(C^{(j)})$. Let $A=\bigoplus_{j=1}^{\infty}C^{(j)}$, then it is clear that if $\tilde{\Xi}_2(\{a_{i,j}\})=1$, then $\mathrm{Sp}(A)$ is a finite set, otherwise it is the entire interval $[-1,1]$.

Now we use the following facts for bounded self-adjoint operators A. If $\mathrm{Sp}(A)$ is a finite set then $\Xi_{cap}(A)=0$ whereas if $\mathrm{Sp}(A)=[-1,1]$ then $\Xi_{cap}(A)=1/2$ (this can be proven easily using the minimal l^{∞} norm property of monic Chebyshev polynomials). We then define $\tilde{\Gamma}_{n_2,n_1}(\{a_{i,j}\})=\min\{\max\{1-2\Gamma_{n_2,n_1}(A),0\},1\}$. It is clear that this defines a generalised algorithm. In particular, given N we can evaluate $\{A_{k,l}:k,l\leq N\}$ using only finitely many evaluations of $\{a_{i,j}\}$, where we can use a bijection between canonical bases of $l^2(\mathbb{N})$ and $\bigoplus_{j=1}^{\infty} X_j$ to view A as acting on $l^2(\mathbb{N})$. We also have the convergence $\lim_{n_2\to\infty}\lim_{n_1\to\infty}\tilde{\Gamma}_{n_2,n_1}(\{a_{i,j}\}) = \tilde{\Xi}_2(\{a_{i,j}\})$, a contradiction.

Step 4: $\{\Xi_{cap}, \Omega_{\mathrm{D}}, \Lambda_2\} \notin \Delta_2^G$. This is the same argument as in step 3 of the proof of Theorem 7.3.1, however now we replace A_m by $A_m = \mathrm{diag}\{d_1, d_2, ..., d_m\} \in \mathbb{C}^{m\times m}$, where $\{d_m\}$ is a dense subsequence of $[-1,1]$, and use the fact that $\Xi_{cap}(B_k) = 0$. It follows that given the proposed height one tower Γ_n and the constructed A, $\Xi_{cap}(A) = 1/2$ but $\Gamma_{n(k)}(A) < 1/4$, the required contradiction.

Step 5: $\{\Xi_{r,p}, \Omega_{\mathrm{SA}}, \Lambda_2\} \notin \Delta_2^G$. Recall that we are given some polynomial p of degree at least two. We assume without loss of generality that the zeros of p are ± 1 and $|p(0)| > 1$ (the more general case is similar). The argument is similar to step 3 of the proof of Theorem 7.3.1, but we spell it out since it uses Lemma 7.4.2. Suppose for a contradiction that a height one tower, Γ_n, solves the problem. We will gain a contradiction by showing that $\Gamma_n(A)$ does not converge for an operator of the form,

$$A = \bigoplus_{r=1}^{\infty} B(z_1, ..., z_{l_r}),$$

and define

$$C = \mathrm{diag}\{z_1, z_2, ...\} \in \Omega_{\mathrm{B}}.$$

Where we assume that $l_r \geq r$ to ensure that the spectrum of A is equal to $\{-1,1\}$ and hence $\Xi_{r,p}(A) = 0$. Now suppose that $l_1, ..., l_k$ have been chosen and consider the operator

$$B_k = B(z_1) \oplus ... \oplus B(z_1, ..., z_{l_k}) \oplus C.$$

The spectrum of B_k is $[-1,1]$ so that $\Xi_{r,p}(B_k) > 1$ and hence there exists $n(k) \geq k$ such that $\Gamma_{n(k)}(B_k) > 1/4$. But $\Gamma_{n(k)}(B_k)$ can only depend on the evaluations of the matrix entries $\{B_k\}_{ij} = \langle B_k e_j, e_i\rangle$ with $i, j \leq N(B_k, n(k))$ (as well as evaluations of the function f) into account. If we choose $l_{k+1} > N(B_k, n(k))$ then by the assumptions in Definition 2.1.1, $\Gamma_{n(k)}(A) = \Gamma_{n(k)}(B_k) > 1/4$. But $\Gamma_n(A)$ must converge to 0, a contradiction. □

7.5 Proof of Theorem 7.3.4

Proof of Theorem 7.3.4 for Ξ_{we}. For the lower bounds, it is enough to note that $\{\Xi_{we}, \Omega_{\mathrm{D}}, \Lambda_2\} \notin \Delta_2^G$ by the same argument as step 1 of the proof of Theorem 7.3.2. The construction is exactly the same but yields $d_{\mathrm{H}}(\Gamma_{n(k)}(A), \{0\}) \leq 1/2$, whereas $\Xi_{we}(A) = [0,1]$. Hence the proposed height one tower cannot converge. To construct a Π_2^A tower for general operators, we need the following Lemma:

Lemma 7.5.1. *Let $B \in \mathbb{C}^{n\times n}$ and $\epsilon > 0$. Then using finitely many arithmetic operations and comparisons, we can compute points $z_1, ..., z_k \in \mathbb{Q} + i\mathbb{Q}$ such that*

$$d_{\mathrm{H}}(\{z_1, ..., z_k\}, W(B)) \leq \epsilon.$$

Proof. Recall from step 1 of the proof of Theorem 7.3.1 that we can compute an upper bound $M \in \mathbb{Q}_+$ for $\|B\|$ in finitely many arithmetic operations and comparisons. Now choose points $x_1, ..., x_k \in \mathbb{Q}^n$, each of norm at most 1, such that $d_{\mathrm{H}}(\{x_1, ..., x_k\}, \{x \in \mathbb{C}^n : \|x\| = 1\}) < \epsilon/(3M)$. These can be computed in finitely many arithmetic operations and comparisons using generalised polar coordinates and approximations of trigonometric identities. It follows that

$$d_{\mathrm{H}}(\{\langle Bx_1, x_1\rangle, ..., \langle Bx_k, x_k\rangle\}, W(B)) \leq 2\epsilon/3.$$

We then let each $z_j \in \mathbb{Q} + i\mathbb{Q}$ be a $\epsilon/4$ approximation of $\langle Bx_j, x_j\rangle$, which can be computed in finitely many arithmetic operations and comparisons. □

Remark 7.5.2 (Efficient computation). *In practice, there are much more efficient methods of computation. For example, the method of Johnson [Joh78], reduces the computation of $W(B)$ for $B \in \mathbb{C}^{n\times n}$ to a series of $n \times n$ Hermitian eigenvalue problems.*

It is well-known that for $A \in \Omega_{\mathrm{B}}$,

$$\mathrm{cl}\left(W(P_nA|_{P_n\mathcal{H}})\right) \uparrow \mathrm{cl}\left(W(A)\right),$$
$$\mathrm{cl}\left(W((I-P_n)A|_{(I-P_n)\mathcal{H}})\right) \downarrow W_e(A).$$

Given A, let $\Gamma_{n_2,n_1}(A)$ be a finite collection of points produced by the algorithm in Lemma 7.5.1 applied to $B = (I - P_{n_2})P_{n_1+n_2+1}A|_{P_{n_1+n_2+1}(I-P_{n_2})\mathcal{H}}$ and $\epsilon = 1/n_1$. The above limits show that Γ_{n_2,n_1} provides a Π_2^A tower for $\{\Xi_{er}, \Omega_{\mathrm{B}}, \Lambda_1\}$. □

Proof of Theorem 7.3.4 for $\Xi_{poll}^{\mathbb{F}}$. We will prove that $\{\Xi_{poll}^{\mathbb{R}}, \Omega_{\mathrm{D}}, \Lambda_i\} \notin \Delta_3^G$ and $\{\Xi_{poll}^{\mathbb{C}}, \Omega_{\mathrm{B}}, \Lambda_1\} \in \Sigma_3^A$. The construction of towers for $\Xi_{poll}^{\mathbb{R}}$ are similar, as are the arguments for lower bounds.

Step 1: $\{\Xi_{poll}^{\mathbb{C}}, \Omega_{\mathrm{B}}, \Lambda_1\} \in \Sigma_3^A$. Let $\tilde{\Gamma}_{n_2,n_1}$ be the Π_2^A tower for $\{\Xi_{er}, \Omega_{\mathrm{B}}, \Lambda_1\}$ constructed above. Let

$$\gamma_{n_2,n_1}(z;A) = \min\{\sigma_{\inf}(P_{n_1}(A - zI)|_{P_{n_2}\mathcal{H}}), \sigma_{\inf}(P_{n_1}(A^* - \bar{z}I)|_{P_{n_2}\mathcal{H}})\}$$

and note that this can be approximated to any given accuracy in finitely many arithmetic operations and comparisons. We assume that we approximate from below to an accuracy of $1/n_1$ and call this approximation $\tilde{\gamma}_{n_2,n_1}$. The function $\gamma_{n_2,n_1}(z;A)$ is Lipschitz continuous with Lipschitz constant bounded by 1. Define the set

$$V_{n_1} = \bigcup_{m=1}^{n_1} U_m,$$

where U_m are the approximations to the open set U. By taking squares of distances to ball centres, we can decide whether a point $z \in \mathbb{Q} + i\mathbb{Q}$ has $\mathrm{dist}(z, V_{n_1}) < \eta$ for any given $\eta \in \mathbb{Q}_+$. Let $\Upsilon_{n_2,n_1}(A,U)$ be the finite collection of all $z \in \tilde{\Gamma}_{n_2,n_1}(A)$ with $\mathrm{dist}(z, V_{n_1}) < 1/n_2 - 1/n_1$. If $\Upsilon_{n_2,n_1}(A,U)$ is empty then set $Q_{n_2,n_1}(A,U) = 0$, otherwise set

$$Q_{n_2,n_1}(A,U) := \sup_{z\in\Upsilon_{n_2,n_1}(A,U)} \tilde{\gamma}_{n_2,n_1}(z;A) - \frac{1}{n_1}.$$

The above remarks show that this can be computed using finitely many arithmetic operations and comparisons.

For notational convenience, we let $W_{n_2} = \mathrm{cl}\left(W((I-P_{n_2})A|_{(I-P_{n_2})\mathcal{H}})\right)$ and also let $W_{n_2,n_1} = W((I-P_{n_2})P_{n_1+n_2+1}A|_{P_{n_1+n_2+1}(I-P_{n_2})\mathcal{H}})$. We claim that the set $\Upsilon_{n_2,n_1}(A,U)$ converges to

$$\Upsilon_{n_2}(A,U) := \mathrm{cl}\left(\left\{z \in W_{n_2} : \mathrm{dist}\left(z, \mathrm{cl}\left(U\right)\right) < \frac{1}{n_2}\right\}\right),$$

as $n_1 \to \infty$, meaning also if $\Upsilon_{n_2}(A,U)$ is empty then $\Upsilon_{n_2,n_1}(A,U)$ is empty for large n_1. If $z \in \Upsilon_{n_2,n_1}(A,U)$, then there exists $\hat{z} \in W_{n_2,n_1} \subset W_{n_2}$ with $|z - \hat{z}| \leq 1/n_1$. Since

$$\mathrm{dist}\left(z, \mathrm{cl}\left(U\right)\right) \leq \mathrm{dist}(z, V_{n_1}) < 1/n_2 - 1/n_1,$$

it follows that $\mathrm{dist}\left(\hat{z}, \mathrm{cl}\left(U\right)\right) < 1/n_2$ and hence $\Upsilon_{n_2}(A,U)$ is non-empty. So to prove convergence we only need to deal with the case $\Upsilon_{n_2}(A,U) \neq \emptyset$. The above argument also shows that any limit point of a subsequence $z_{m(j)} \in \Upsilon_{n_2,m(j)}(A,U)$ must lie in $\Upsilon_{n_2}(A,U)$. Hence to prove the claim, we need to only prove that for any $z \in \Upsilon_{n_2}(A,U)$, there exists z_{n_1} that are contained in $\Upsilon_{n_2,n_1}(A,U)$ for large n_1 and converge to z.

Let $z \in W_{n_2}$ with $\mathrm{dist}\,(z, \mathrm{cl}\,(U)) < 1/n_2$, then there exists $\epsilon > 0$ and $j > 0$ such that $\mathrm{dist}(z, U_j) < 1/n_2 - \epsilon$. There also exists $z_{n_1} \in \tilde{\Gamma}_{n_2,n_1}(A)$ with $z_{n_1} \rightarrow z$. It must hold for $n_1 > j$ that

$$\begin{aligned}\mathrm{dist}(z_{n_1}, V_{n_1}) \leq \mathrm{dist}(z_{n_1}, V_j) &\leq |z_{n_1} - z| + \mathrm{dist}(z, U_j)\\ &< |z_{n_1} - z| + \frac{1}{n_2} - \epsilon.\end{aligned}$$

This last quantity is smaller than $1/n_2 - 1/n_1$ for large n_1 and hence $z_{n_1} \in \Upsilon_{n_2,n_1}(A,U)$ for large n_1. It follows for any $z \in \Upsilon_{n_2}(A,U)$, there exists z_{n_1} that are contained in $\Upsilon_{n_2,n_1}(A,U)$ for large n_1 and converge to z.

Define

$$Q_{n_2}(A,U) := \sup_{z \in \Upsilon_{n_2}(A,U)} \gamma_{n_2}(z;A),$$

where we recall that $\gamma_{n_2}(z;A) = \min\{\sigma_{\inf}((A - zI)|_{P_{n_2}\mathcal{H}}), \sigma_{\inf}((A^* - \bar{z}I)|_{P_{n_2}\mathcal{H}})\}$. If $z \in \Upsilon_{n_2,n_1}(A,U)$, then the above shows that there exists $\hat{z} \in \Upsilon_{n_2}(A,U)$ with $|z - \hat{z}| \leq 1/n_1$. It follows that

$$\begin{aligned}\tilde{\gamma}_{n_2,n_1}(z;A) - \frac{1}{n_1} &\leq \gamma_{n_2,n_1}(z;A) - \frac{1}{n_1}\\ &\leq \gamma_{n_2,n_1}(\hat{z};A) \leq \gamma_{n_2}(z;A),\end{aligned}$$

where we have used the bound on the Lipschitz constant and the fact that γ_{n_2,n_1} converge up to γ_{n_2} (and uniformly on compact subsets of $\mathbb{C}$). It follows that $Q_{n_2,n_1}(A,U) \leq Q_{n_2}(A,U)$ and this also covers the case that $\Upsilon_{n_2}(A,U) = \emptyset$ if we define the supremum over the empty set to be 0. The set convergence proven above and uniform convergence of $\tilde{\gamma}_{n_2,n_1}$ implies that $Q_{n_2,n_1}(A,U)$ converges to $Q_{n_2}(A,U)$. It is also clear that the $\Upsilon_{n_2}(A,U)$ are nested and converge down to $W_e(A) \cap \mathrm{cl}\,(U)$ since W_{n_2} converges down to $W_e(A)$. The function γ_{n_2} also converges down to

$$\gamma(z;A) = \|R(z,A)\|^{-1}$$

uniformly on compact subsets of $\mathbb{C}$ and hence $Q_{n_2}(A,U)$ converges down to

$$Q(A,U) = \sup_{z \in W_e(A) \cap \mathrm{cl}(U)} \|R(z,A)\|^{-1}.$$

Define

$$\Gamma_{n_3,n_2,n_1}(A,U) = 1 - \chi_{[0,1/n_3]}(Q_{n_2,n_1}(A,U)) \in \{0,1\}.$$

The above show that

$$\lim_{n_1 \rightarrow \infty} \Gamma_{n_3,n_2,n_1}(A,U) = 1 - \chi_{[0,1/n_3]}(Q_{n_2}(A,U)) =: \Gamma_{n_3,n_2}(A,U).$$

Since $\chi_{[0,1/n_3]}$ has right limits and $Q_{n_2}(A,U)$ are non-increasing,

$$\lim_{n_2 \rightarrow \infty} \Gamma_{n_3,n_2}(A,U) = 1 - \chi_{[0,1/n_3]}(Q(A,U)\pm) := \Gamma_{n_3}(A,U),$$

where ± denotes one of the right or left limits (it is possible to have either). Now if $\Xi_{poll}^{\mathbb{C}}(A,U) = 0$, then $\Gamma_{n_3}(A,U) = 0$ for all n_3. But if $\Xi_{poll}^{\mathbb{C}}(A,U) = 1$, then for large n_3, $\Gamma_{n_3}(A,U) = 1$. Moreover, in this latter case, $\Gamma_{n_3}(A, U) = 1$ signifies the existence of $z \in W_e(A) \cap \mathrm{cl}(U)$ with $\gamma(z; A) > 0$ and hence $z \notin \mathrm{Sp}(A)$. Hence Γ_{n_3,n_2,n_1} provides a Σ_3^A tower.

Step 2: $\{\Xi_{poll}^{\mathbb{R}}, \Omega_{\mathrm{D}}, \Lambda_2\} \notin \Delta_3^G$. We will argue for the case that $U = U_1 = \mathbb{R}$ and the restricted case is similar. Assume for a contradiction that this is false and $\widehat{\Gamma}_{n_2,n_1}$ is a general height two tower for $\{\Xi_{poll}^{\mathbb{R}}, \Omega_{\mathrm{D}}, \Lambda_2\}$. We follow the same strategy as the proof of Theorem 7.3.1 step 4. Let $(\mathcal{M}, d)$ be discrete space $\{0, 1\}$ and $\tilde{\Omega}$ denote the collection of all infinite matrices $\{a_{i,j}\}_{i,j\in\mathbb{N}}$ with entries $a_{i,j} \in \{0, 1\}$ and consider the problem function

$$\tilde{\Xi}_1(\{a_{i,j}\}) : \text{ Does } \{a_{i,j}\} \text{ have a column containing infinitely many non-zero entries?}$$

For $j \in \mathbb{N}$, let $\{b_{i,j}\}_{i\in\mathbb{N}}$ be a dense subset of $I_j := [1-1/2^{2j-1}, 1-1/2^{2j}]$. Given a matrix $\{a_{i,j}\}_{i,j\in\mathbb{N}} \in \tilde{\Omega}$, construct a matrix $\{c_{i,j}\}_{i,j\in\mathbb{N}}$ by letting $c_{i,j} = a_{i,j} b_{r(i,j),j}$ where

$$r(i, j) = \max\left\{1, \sum_{k=1}^{i} a_{k,j}\right\}.$$

Now consider any bijection $\phi : \mathbb{N} \to \mathbb{N}^2$ and define the diagonal operator

$$A = \mathrm{diag}(c_{\phi(1)}, c_{\phi(2)}, c_{\phi(3)}, ...).$$

The algorithm $\widehat{\Gamma}_{n_2,n_1}$ thus translates to an algorithm Γ'_{n_2,n_1} for $\{\tilde{\Xi}_1, \tilde{\Omega}\}$. Namely, we define the algorithm $\Gamma'_{n_2,n_1}(\{a_{i,j}\}_{i\in\mathbb{N}}) = \widehat{\Gamma}_{n_2,n_1}(A)$. The fact that ϕ is a bijection shows that the lowest level Γ'_{n_2,n_1} are generalised algorithms (and are consistent). In particular, given N, we can find $\{A_{i,j} : i, j \leq N\}$ using finitely many evaluations of the matrix values $\{c_{k,l}\}$ (the same is true for A^*A and AA^* since the operator is diagonal). But for any given $c_{k,l}$ we can evaluate this entry using only finitely many evaluations of the matrix values $\{a_{m,n}\}$ by the construction of r. Finally note that

$$\mathrm{Sp}(A) = \{1\} \cup \left(\bigcup_{j:\{a_{i,j}\}_{i\in\mathbb{N}} \text{ has infinitely many 1s}} I_j\right) \cup Q,$$

where Q lies in the discrete spectrum. The intervals I_j are also separated. It follows that there is a gap in the essential spectrum if and only if there exists a column $\{a_{i,j}\}_{i\in\mathbb{N}}$ with infinitely many 1s. Otherwise the essential spectrum is $\{1\}$. It follows that $\tilde{\Xi}(\{a_{i,j}\}) = \Xi_{poll}^{\mathbb{R}}(A, \mathbb{R})$ and hence we get a contradiction. □

7.6 Numerical Examples

The SCI-sharp towers of algorithms constructed in this chapter can be efficiently implemented for large scale computations. Moreover, they have desirable convergence properties, converging monotonically or being eventually constant, as captured by the Σ/Π classification. Generically, this monotonicity holds in all of the limits, and not just the final limit: many of the towers undergo *oscillation phenomena* where each subsequent limit is monotone but in the opposite sense/direction than the limit beforehand. We can take advantage of this when analysing the algorithms numerically. The algorithms also highlight suitable information that lowers the SCI classification to Σ_1/Π_1. Other advantages for the algorithms based on approximating the resolvent norm include locality, numerical stability and speed/parallelisation.

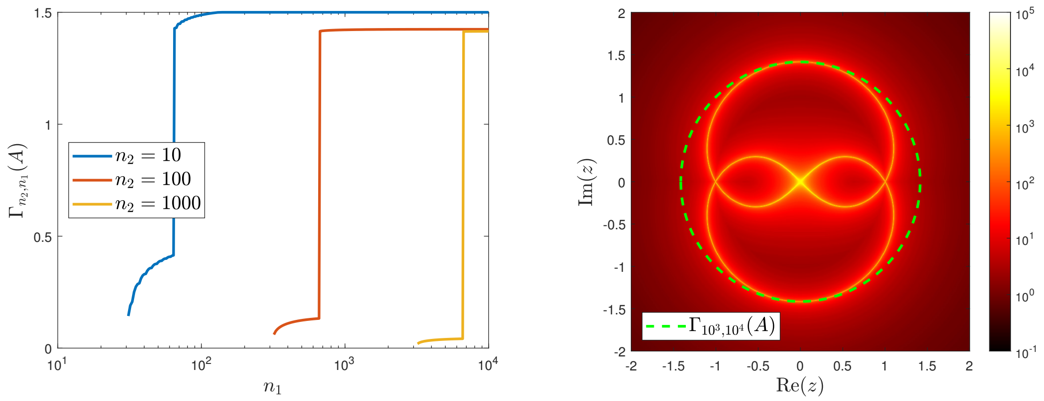


Figure 7.1: Left: Output of the algorithm for computing the spectral radius. Right: Pseudospectrum computed using the method of Chapter 3 (the colour scale corresponds to the resolvent norm $\|(A - zI)^{-1}\|$) which provides error control. We have show the output of $\Gamma_{10^3,10^4}(A)$ via the green dashed circle.

7.6.1 Numerical example for spectral radius

We begin with the spectral radius and consider the upper-triangular non-normal operator on $l^2(\mathbb{Z})$ defined by its action on the canonical basis via

$$Ae_j = e_{j-2} + i^j e_{j-1}.$$

In this case, the operator norm of A is 2 and the approximation of the spectrum by finite section is $\{0\}$. Hence, to compute the spectral radius, one must resort to the techniques used in our tower of algorithms based on rectangular truncations. Recall that the SCI classification for computing the spectral radius of such operators (where the dispersion is known[3]) is Π_2^A (see Theorem 7.3.1 for further classifications). The first parameter, n_1, controls the size of the rectangular truncation (as well as the grid resolution), whereas the second, n_2, controls the resolvent norm cut-off ($\epsilon = 1/n_2$).

Figure 7.1 (left) shows the output of the tower of algorithms $\Gamma_{n_2,n_1}(A)$ for computing the spectral radius. We see the expected monotonicity: $\Gamma_{n_2,n_1}(A)$ is increasing in n_1 but decreasing in n_2. It appears that $\lim_{n_1\to\infty}\Gamma_{10^2,n_1}(A) \approx \lim_{n_1\to\infty}\Gamma_{10^3,n_1}(A) \approx 1.4149$. The fact that these two values for different n_2 are similar suggests that we have reached convergence. Though, of course, the proof that the problem does not lie in Δ_2^G shows that we can never apply a choice of subsequences to gain convergence in one limit over the whole class Ω_f. Nevertheless, the approximate value of 1.4149 is confirmed in Figure 7.1 (right) where we have shown pseudospectra, computed using the algorithm of Chapter 3.

7.6.2 Numerical examples for essential numerical range

To demonstrate the algorithm for computing the essential numerical range, we first consider the Laurent operator A_0 acting on $l^2(\mathbb{Z})$ with symbol

$$a(t) = \frac{t^4 + t^{-1}}{2}.$$

In this case, $\mathrm{Sp}(A_0) = \mathrm{Sp}_{\mathrm{ess}}(A_0) = \{a(z) : |z| = 1\}$. We consider the operator $A = A_0 + E$ where the compact perturbation E is given by

$$Ee_j = -\frac{3i}{1+|j|}e_{j-1}.$$

[3] For this example and others on $l^2(\mathbb{Z})$, we reorder the basis so that the operator A acts on $l^2(\mathbb{N})$.

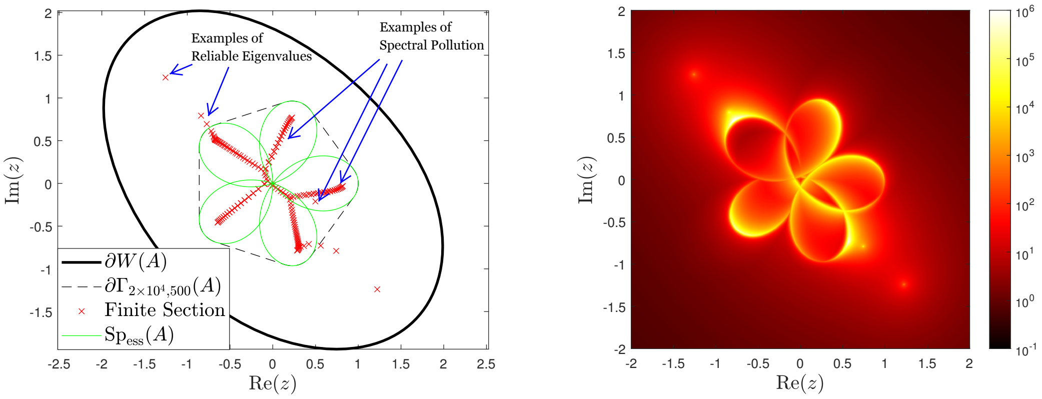


Figure 7.2: Left: The boundaries of $\partial W(A)$ and $\partial\Gamma_{2\times 10^4,500}(A)$. We have also shown the essential spectrum of A (whose convex hull, in this example, corresponds to $W_e(A)$) and the output of finite section for a 200×200 truncation. Right: Pseudospectrum computed using the method of Chapter 3 (the colour scale corresponds to the resolvent norm $\|(A - zI)^{-1}\|$) which provides error control. This confirms that eigenvalues, computed using finite section, outside $\partial\Gamma_{2\times 10^4,500}(A)$ are accurate and, in this example, indicates that the other eigenvalues correspond to spectral pollution.

Recall that the SCI classification for computing the essential numerical range is Π_2^A (see Theorem 7.3.4). The first parameter, n_1, controls the size of the truncation, whereas the second, n_2, controls how far along the matrix the truncations $(I - P_{n_2})P_{n_1+n_2}A|_{P_{n_1+n_2}(I-P_{n_2})\mathcal{H}}$ are taken with respect to the canonical basis once we have represented the operator as an operator on $l^2(\mathbb{N})$. (An alternative to reordering the basis so that the operator acts on $l^2(\mathbb{N})$ is to use truncations in 'both directions' on $l^2(\mathbb{Z})$ by letting P_n be the projection onto the span of $\{e_j : |j| \leq n\}$.)

Figure 7.2 (left) shows the output of the algorithm $\Gamma_{n_2,n_1}(A)$ to compute the essential numerical range for $n_2 = 20000$ and $n_1 = 500$. We show the boundary $\partial\Gamma_{n_2,n_1}(A)$ since the essential numerical range is convex. In this example, $W_{\mathrm{e}}(A)$ is the convex hull of $\mathrm{Sp}_{\mathrm{ess}}(A_0)$, which allows us to verify the output of the algorithm. We also show 200 eigenvalues of finite section (computed using extended precision to avoid numerical instabilities associated with non-normal truncations), the majority of which are due to truncation and provide an example of spectral pollution. This is confirmed when we compare to the pseudospectrum, also shown in Figure 7.2 (right), computed using the algorithm of Chapter 3. However, eigenvalues outside $W_{\mathrm{e}}(A)$ correspond to true eigenvalues of A (see Theorem 7.1.2).

The algorithm can also be extended to unbounded operators, as outlined in [Colns].[4] For example, we consider the complex Schrödinger operator

$$T = -\frac{d^2}{dx^2} + (2i + 1)\cos(x). \tag{7.6.1}$$

By using a Gabor basis, we can represent T as a closed operator on $l^2(\mathbb{N})$ such that the linear span of the canonical basis (corresponding to the Gabor basis) forms a core. We compute the matrix elements (corresponding to inner products with the basis functions) with error control using quadrature. Figure 7.3 shows the output for $n_2 = 10^4$ and various n_1. We see the expected monotonicity as n_1 increases and the output for $n_1 = 2000$ has converged to visible accuracy in the plot.

[4] The essential numerical range for unbounded operators was defined and studied in [BMT20].

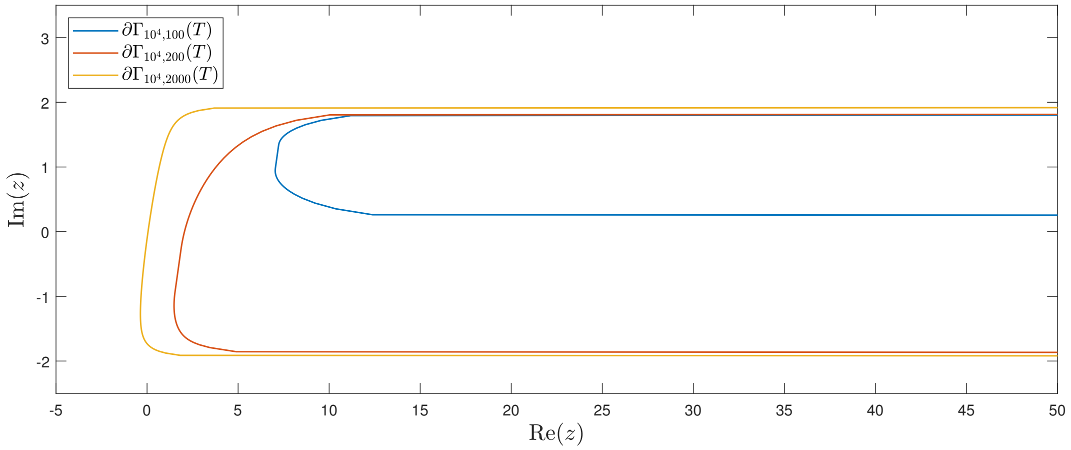


Figure 7.3: The output of the algorithm for computing the essential numerical range of closed operators, applied to the complex Schrödinger operator T in (7.6.1).

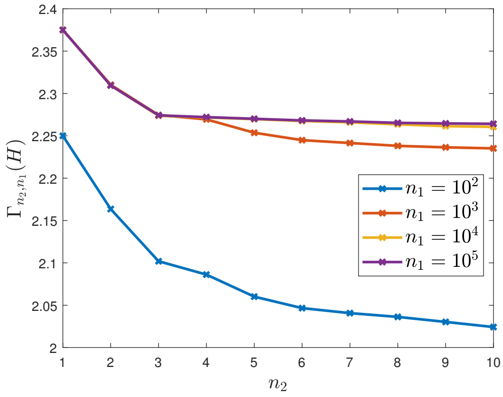


Figure 7.4: Output of the algorithm for computing the capacity of $\mathrm{Sp}(H_0)$.

7.6.3 Numerical example for capacity

We now consider the transport Hamiltonian H_0 on a Penrose tile discussed in §3.4.1 of Chapter 3. Recall that by choosing a suitable ordering of the vertices, we can represent H_0 as an operator acting on $l^2(\mathbb{N})$ of bounded dispersion with $f(n) - n \sim \mathcal{O}(\sqrt{n})$. Recall also that the SCI classification for computing the capacity of the spectrum of such operators is Π_2^A (see Theorem 7.3.3 for further classifications). The first parameter, n_1, controls the size of the truncation used to test if intervals intersect the spectrum via Lemma 8.1.7, whereas the second, n_2, controls the spacings of the interval coverings (which have width 2^{-n_2}). In this example, we used the conformal mapping method of [LSN17] to accurately and rapidly compute the capacity of finite unions of intervals in $\mathbb{R}$. See Remark 7.4.1 for a discussion of computational efficiency.

Figure 7.4 shows the output of $\Gamma_{n_2,n_1}(H_0)$ and we see the expected monotonicity: the output is increasing in n_1 but decreasing in n_2. By comparing the outputs for $n_1 = 10^4$ and $n_1 = 10^5$, it appears we have convergence up to around $n_2 = 8$. This suggests an upper bound (since the output is non-increasing in n_2) of approximately 2.26 for the capacity of $\mathrm{Sp}(H_0)$ ($\mathrm{Sp}(H_0)$ is shown in Figure 3.1).

Chapter 8

Lebesgue Measure and Fractal Dimensions of Spectra

In this chapter, based on [Colns], we consider the SCI of computing the Lebesgue measure of the spectrum (and pseudospectrum) and different fractal dimensions of the spectrum (box-counting and Hausdorff). This chapter is motivated by recent progress in the field of Schrödinger operators with random or almost periodic potentials [Avi09, Avi08, AJ09, AK06, AV07, Pui04, Süt89]. Cantor-like spectra occur in many families of one-dimensional operators. Fractal dimensions of spectra are important in many applications. For example, in quantum mechanics, they lead to upper bounds on the spreading of wavepackets, and are related to time-dependent quantities associated with wave functions [HTHK94, KPG92, KKKG97]. Fractal spectra appear in a wide variety of contexts, such as exciting new results in multilayer materials (e.g. bilayer graphene) [DWM$^+$13, GG13a, HSYY$^+$13, PGY$^+$13], strained materials [NBLOLT17, RTN14] or quasicrystals [BRS16, TGB$^+$14, KST87, LRF$^+$11].

Whilst results are known for specific one-dimensional examples such as the almost Mathieu operator [AK06] or the Fibonacci Hamiltonian [Süt89], the problems of computing the Lebesgue measure and fractal dimensions of spectra remain open in the general case [DGS15]. This is reflected by the difficulty of performing rigorous numerical studies, despite many examples studied in the physics literature (see the references in [AJM17, BS91, Sir89]). In general, there are no known algorithms for determining the Lebesgue measure and fractal dimension of spectra for general operators or even banded self-adjoint operators.

We solve these problems and design towers of algorithms that are numerically implementable. These are demonstrated numerically on a two-dimensional model of a quasicrystal. In particular, we provide numerical evidence that a portion of the spectrum of the graphical Laplacian on a Penrose tile is fractal with fractal dimension approximately 0.8. However, we find that determining the Lebesgue measure and fractal dimensions are hard in the sense of the SCI. This helps to explain the difficulty encountered in studying these properties numerically or theoretically.

8.1 Main Results

We continue to use the set-up of Chapter 7 described in §7.2 and recall the following classes of bounded operators from §7.3, for which we prove classifications:

- Ω_f: operators with dispersion bounded by f
- Ω_g: operators with resolvent bounded by g
- Ω_{D}: self-adjoint and diagonal operators
- Ω_{SA}: self-adjoint operators
- Ω_{N}: normal operators
- Ω_{B}: general bounded operators.

We first discuss the Lebesgue measure, and then move onto the computation of the box-counting dimension and Hausdorff dimension.

8.1.1 Lebesgue measure of spectra

A basic property of $\mathrm{Sp}(A)$, also connected to physical applications in quantum mechanics, is its Lebesgue measure. Well-studied operators such as the almost Mathieu operator at critical coupling [AK06] or the Fibonacci Hamiltonian [Süt89] have spectra with Lebesgue measure zero. The Lebesgue measure on $\mathbb{C}$ will be denoted by Leb and, when considering classes of self-adjoint operators, the Lebesgue measure on $\mathbb{R}$ will be denoted by $\mathrm{Leb}_{\mathbb{R}}$. We will also consider

$$\widehat{\mathrm{Sp}}_\epsilon(A) = \{z \in \mathbb{C} : \|R(z,A)\|^{-1} < \epsilon\},$$

whose closure is $\mathrm{Sp}_\epsilon(A)$. For a class $\Omega \subset \Omega_{\mathrm{B}}$, there are three questions we are interested in and answer in this section:

1. Given $A \in \Omega$, can we compute $\mathrm{Leb}(\mathrm{Sp}(A))$?
2. Given $A \in \Omega$ and $\epsilon > 0$, can we compute $\mathrm{Leb}(\widehat{\mathrm{Sp}}_\epsilon(A))$?
3. Given $A \in \Omega$, can we determine whether $\mathrm{Leb}(\mathrm{Sp}(A)) = 0$?

We do not consider the final question for the pseudospectrum since $\mathrm{Leb}(\widehat{\mathrm{Sp}}_\epsilon(A)) > 0$. It might appear that answering the third question is at least as easy as the first. However, this could be false (and in general is), since we consider a problem function with range in a different metric space. For the first two questions, we consider the metric space $([0,\infty), d)$ with the Euclidean metric. For question three we consider the discrete metric on $\{0,1\}$, where 1 is interpreted as 'yes', and 0 as 'no'. Finally, we consider the computation of $\mathrm{Leb}(\widehat{\mathrm{Sp}}_\epsilon(A))$ since it is not immediately clear that the level sets

$$S_\epsilon(A) := \{z \in \mathbb{C} : \|R(z,A)\|^{-1} = \epsilon\} \tag{8.1.1}$$

always have Lebesgue measure zero. Again, this is analogous to the case of approximating the pseudospectra for bounded operators, where one uses the crucial property that the pseudospectrum cannot jump - it cannot be constant on open subsets of $\mathbb{C}$ for bounded operators acting on a separable Hilbert space [Sha08]. Assuming that the sets in (8.1.1) are null is the measure theoretic equivalent. Note, however, that it is straightforward to show that $S_\epsilon(A)$ is null for $A \in \Omega_{\mathrm{N}}$ through the formula $\|R(z,A)\|^{-1} = \mathrm{dist}(z, \mathrm{Sp}(A))$.

The above problem functions are denoted by Ξ_1^L, Ξ_2^L and Ξ_3^L respectively. In analogy to computing the spectra/pseudospectra themselves, Ξ_2^L is, in fact, the easiest to compute and can be done in one limit for a large class of operators. We also have from the dominated convergence theorem that

$$\lim_{\epsilon\downarrow 0}\mathrm{Leb}(\widehat{\mathrm{Sp}}_\epsilon(A))=\mathrm{Leb}(\mathrm{Sp}(A)). \tag{8.1.2}$$

Unless otherwise told, we will assume that given $A\in\Omega_f$, we know a sequence $\{c_n\}_{n\in\mathbb{N}}$ that converges to zero such that $D_{f,n}(A)\leq c_n$. When considering Ω_{D} or Ω_{SA}, we use $\mathrm{Leb}_{\mathbb{R}}$.

Lebesgue measure of spectrum and pseudospectrum

Theorem 8.1.1. *Given the above set-up, we have the following classifications*

$$\Delta_2^G\not\ni\{\Xi_1^L,\Omega_f,\Lambda_i\}\in\Pi_2^A,\quad \Delta_2^G\not\ni\{\Xi_1^L,\Omega_{\mathrm{D}},\Lambda_i\}\in\Pi_2^A\quad i=1,2,$$

and for $\Omega=\Omega_{\mathrm{B}},\Omega_{\mathrm{SA}}$, Ω_{N} *or* Ω_g,

$$\Delta_3^G\not\ni\{\Xi_1^L,\Omega,\Lambda_1\}\in\Pi_3^A,\quad \Delta_2^G\not\ni\{\Xi_1^L,\Omega,\Lambda_2\}\in\Pi_2^A.$$

The constructed algorithm is local, and we can easily adapt it to find the Lebesgue measure of $\mathrm{Sp}(A)$ intersected with any compact interval or cube in one or two dimensions, respectively. It also does not need the sequence $\{c_n\}$. In other words, the evaluations of $\{c_n\}$ can be dropped from Λ_i, and the theorem remains true. The algorithm can also be restricted to $\mathbb{R}$ where it converges to $\mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}(A)\cap\mathbb{R})$.

We now turn to the SCI classification of $\mathrm{Leb}(\widehat{\mathrm{Sp}}_\epsilon(A))$ which is useful since it provides a route to computing $\mathrm{Leb}(\mathrm{Sp}(A))$ for any $A\in\Omega_{\mathrm{B}}$ via (8.1.2). This is a similar state of affairs to the computation of the spectrum itself - one can approximate the spectrum via pseudospectra.

Theorem 8.1.2. *Given the above set-up, we have the following classifications*

$$\Delta_1^G\not\ni\{\Xi_2^L,\Omega_f,\Lambda_i\}\in\Sigma_1^A,\quad \Delta_1^G\not\ni\{\Xi_2^L,\Omega_{\mathrm{D}},\Lambda_i\}\in\Sigma_1^A\quad i=1,2,$$

and for $\Omega=\Omega_{\mathrm{B}},\Omega_{\mathrm{SA}}$, Ω_{N} *or* Ω_g,

$$\Delta_2^G\not\ni\{\Xi_2^L,\Omega,\Lambda_1\}\in\Sigma_2^A,\quad \Delta_1^G\not\ni\{\Xi_2^L,\Omega,\Lambda_2\}\in\Sigma_1^A.$$

Heuristically, the pseudospectrum is less refined than the spectrum, making the measure easier to estimate. Another viewpoint is the analysis of the continuity points of the maps Ξ_1^L and Ξ_2^L:

Proposition 8.1.3. *In the above set-up, the following hold:*

1. Ξ_1^L *is continuous at* $A\in\Omega_{\mathrm{D}}$ *if and only if* $\mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}(A))=0$.
2. Ξ_2^L *is continuous at all* $A\in\Omega_{\mathrm{D}}$ *if* $\epsilon>0$.

Exercise: Prove Proposition 8.1.3.

When is $\mathrm{Leb}(\mathrm{Sp}(A))=0$?

In this section, let $(\mathcal{M},d)$ be the set $\{0,1\}$ with the discrete topology and consider the problem function

$$\Xi_3^L(A)=\begin{cases}0, & \text{if }\mathrm{Leb}(\mathrm{Sp}(A))>0\\ 1, & \text{otherwise.}\end{cases}$$

It is straightforward to build a height three tower for this problem based on the monotonicity of `LebSpec` (the algorithm constructed in Theorem 8.1.1). The next theorem shows that this is optimal - even for the set of diagonal self-adjoint bounded operators. This demonstrates just how hard it is to answer decision problem questions about the spectrum with finite amounts of information, particularly when the questions involve a tool such as Lebesgue measure, which ignores countable sets.

Theorem 8.1.4. *Given the above set-up, we have the following classifications*

$$\Delta_3^G \not\ni \{\Xi_3^L, \Omega_f, \Lambda_i\} \in \Pi_3^A, \quad \Delta_3^G \not\ni \{\Xi_3^L, \Omega_{\mathrm{D}}, \Lambda_i\} \in \Pi_3^A, \quad i = 1, 2,$$

and for $\Omega = \Omega_{\mathrm{B}}, \Omega_{\mathrm{SA}}$, Ω_{N} *or* Ω_g,

$$\Delta_4^G \not\ni \{\Xi_3^L, \Omega, \Lambda_1\} \in \Pi_4^A, \quad \Delta_3^G \not\ni \{\Xi_3^L, \Omega, \Lambda_2\} \in \Pi_3^A.$$

8.1.2 Fractal dimensions of spectra

If the spectrum of an operator has zero Lebesgue measure, it is natural to ask about its fractal dimension. This question is not just borne out of mathematical curiosity. For instance, the fractal dimension leads to an upper bound on the spreading of an initially localised wavepacket, and there has been much work by physicists on relating the fractal dimension to time-dependent quantities associated with wave functions. However, estimating the fractal dimension is extremely difficult. One possible reason is that it is not possible to construct a height one tower of algorithms, even for the most basic definition of fractal dimension, the box-counting dimension. The Hausdorff dimension is even worse and has SCI ≥ 3. In this section, we exclusively treat self-adjoint operators and seek fractal dimensions of subsets of $\mathbb{R}$.[1]

Box-counting dimension

Let F be a bounded set in some Euclidean space and $N_\delta(F)$ be the number of closed boxes of side length $\delta > 0$ required to cover F. Define the upper and lower box-counting dimensions as

$$\overline{\dim}_B(F) = \limsup_{\delta \downarrow 0} \frac{\log(N_\delta(F))}{\log(1/\delta)},$$
$$\underline{\dim}_B(F) = \liminf_{\delta \downarrow 0} \frac{\log(N_\delta(F))}{\log(1/\delta)}.$$

When both are equal, we can replace the $\liminf$ and $\limsup$ by $\lim$ and we define the common value as the box-counting dimension $\dim_B(F)$, an example of a fractal dimension. The major drawback of this definition is lack of countable stability. For instance, the box-counting dimension of $\{0, 1, 1/2, 1/3, ...\}$ is $1/2$. Examples also exist of closed Cantor sets for which the upper and lower dimensions do not agree [Fal03]. A natural example occurring as the spectrum of a discrete Schrödinger operator is presented in [Colns], where this effect can be seen numerically. In the one-dimensional case, it is easy to prove that if F is measurable with $\overline{\dim}_B(F) < 1$ then $\mathrm{Leb}_{\mathbb{R}}(F) = 0$. The converse is false by considering countable unions of Cantor sets whose Hausdorff dimension tends to 1 and similar results hold in higher dimensions. We shall show that we can compute the box-counting dimension in two limits.

[1] The proofs for general self-adjoint operators can be adapted with an additional limit and the use of two-dimensional covering boxes to treat the class of general bounded operators. Some care is required involving boundaries of covering boxes for the Hausdorff dimension, but for brevity, we omit the details.

Let Ω_f^{BD} be the class of self-adjoint operators in Ω_f whose upper and lower box-counting dimensions of the spectrum agree. Let $\Omega_{\mathrm{SA}}^{BD}$ be the class of self-adjoint operators whose upper and lower box-counting dimensions of the spectrum agree, and denote by Ω_{D}^{BD} the class of diagonal operators in $\Omega_{\mathrm{SA}}^{BD}$.

Theorem 8.1.5. *Let Ξ_B be the evaluation of box-counting dimension of spectra, then for $i = 1, 2$ and $\Omega = \Omega_f^{BD}$ or Ω_{D}^{BD}*

$$\Delta_2^G \not\ni \{\Xi_B, \Omega, \Lambda_i\} \in \Pi_2^A,$$

whereas

$$\Delta_3^G \not\ni \{\Xi_B, \Omega_{\mathrm{SA}}^{BD}, \Lambda_1\} \in \Pi_3^A, \quad \Delta_2^G \not\ni \{\Xi_B, \Omega_{\mathrm{SA}}^{BD}, \Lambda_2\} \in \Pi_2^A.$$

Remark 8.1.6. *The algorithms we construct for Ξ_B also converge without the assumption that the upper and lower box-counting dimensions agree to a quantity $\Gamma(A)$ with*

$$\underline{\dim}_B(\mathrm{Sp}(A)) \leq \Gamma(A) \leq \overline{\dim}_B(\mathrm{Sp}(A)).$$

Hausdorff dimension

A more complicated, yet robust notion of fractal dimension is related to the Hausdorff measure. For the connection and various other measures that give rise to the same dimension we refer the reader to [Fal03, Mat95]. Let $F \subset \mathbb{R}^n$ be a Borel set in n-dimensional Euclidean space and let $\mathcal{C}_\delta(F)$ denote the class of (countable) δ-covers[2] of F. One first defines the quantity (for $d \geq 0$)

$$\mathcal{H}_\delta^d(F) = \inf\left\{\sum_i \mathrm{diam}(U_i)^d : \{U_i\} \in \mathcal{C}_\delta(F)\right\},$$

and the d-dimensional Hausdorff measure of F by

$$\mathcal{H}^d(F) = \lim_{\delta \downarrow 0} \mathcal{H}_\delta^d(F).$$

There is a unique $d' = \dim_H(F) \geq 0$, the Hausdorff dimension of F, such that $\mathcal{H}^d(F) = 0$ for $d > d'$ and $\mathcal{H}^d(F) = \infty$ for $d < d'$.

One can prove that

$$\dim_H(F) \leq \underline{\dim}_B(F) \leq \overline{\dim}_B(F).$$

A useful property of the Hausdorff dimension that makes it hard to compute is its countable stability (if F is countable then $\dim_H(F) = 0$). The following lemma is used in the construction of an algorithm for computing the Hausdorff dimension but is interesting in its own right so is listed here.

Lemma 8.1.7. *Let $(a, b) \subset \mathbb{R}$ be a finite open interval and let $A \in \Omega_f \cap \Omega_{\mathrm{SA}}$. Then determining whether $\mathrm{Sp}(A) \cap (a, b) \neq \emptyset$ using Λ_i is a problem with $\mathrm{SCI}_A = 1$. Furthermore, we can design an algorithm that halts if and only the answer is 'yes', that is, the problem lies in Σ_1^A. Similarly the problem lies in Σ_2^A when considering Ω_{SA} with Λ_1 (or Σ_1^A when we allow access to Λ_2).*

Theorem 8.1.8. *Let Ξ_H be the evaluation of the Hausdorff dimension of spectra, then for $i = 1, 2$ and $\Omega = \Omega_{\mathrm{D}}$ or $\Omega_f \cap \Omega_{\mathrm{SA}}$*

$$\Delta_3^G \not\ni \{\Xi_H, \Omega, \Lambda_i\} \in \Sigma_3^A,$$

whereas

$$\Delta_4^G \not\ni \{\Xi_H, \Omega_{\mathrm{SA}}, \Lambda_1\} \in \Sigma_4^A, \quad \Delta_3^G \not\ni \{\Xi_H, \Omega_{\mathrm{SA}}, \Lambda_2\} \in \Sigma_3^A.$$

[2] That is, the set of covers $\{U_i\}_{i \in I}$ with I at most countable and with $\mathrm{diam}(U_i) \leq \delta$.

8.2 Proofs of Theorems on Lebesgue Measure

We will use the function `DistSpec` discussed in Chapter 3, written in (highly non-efficient form):

```
Function DistSpec($A$,$n$,$f(n)$,$z$)
    Input  : $n \in \mathbb{N}$, $f(n) \in \mathbb{N}$, matrix $A$, $z \in \mathbb{C}$
    Output: $y \in \mathbb{R}_+$, an approximation to the function $z \mapsto \|R(z,A)\|^{-1}$
    $B = (A - zI)(1 : f(n), 1 : n), \quad C = (A - zI)^*(1 : f(n), 1 : n)$
    $S = B^*B, \quad T = C^*C$
    $\nu = 1, l = 0$
    while $\nu = 1$ do
        $l = l + 1$
        $p = \texttt{IsPosDef}(S - \frac{l^2}{n^2}), \quad q = \texttt{IsPosDef}(T - \frac{l^2}{n^2})$
        $\nu = \min(p, q)$
    end
    $y = \frac{l}{n}$
end
```

For ease of notation, we suppress the dispersion function f in calling `DistSpec` but assume that we know $D_{f,n}(A) \leq c_n$ with $c_n \to 0$ as $n \to \infty$. However, the proof of convergence also works when using $c_n = 0$ (which does not necessarily bound $D_{f,n}(A)$). The key observation is the following:

Observation: If $A \in \Omega_f$, then the function $F_n(z) := \texttt{DistSpec}(A, n, f(n), z) + c_n$ converges uniformly to $\|R(z,A)\|^{-1}$ from above on compact subsets of $\mathbb{C}$. By taking successive minima, we can assume without loss of generality that F_n is non-increasing in n.

The other ingredient needed is the following proposition

Proposition 8.2.1. *Given a finite union of disks in the complex plane, the Lebesgue measure of their intersection with the interior of a rectangle can be computed within arbitrary precision using finitely many arithmetical operations and comparisons on the centres and radii of the discs as well the position of the rectangle.*

Exercise: Prove Proposition 8.2.1.

Proof of Theorem 8.1.1. **Step 1:** $\{\Xi_1^L, \Omega_f, \Lambda_i\}, \{\Xi_1^L, \Omega_{\mathrm{D}}, \Lambda_i\} \in \Pi_2^A$. It is enough to consider Λ_1. We will estimate $\mathrm{Leb}(\mathrm{Sp}(A))$ by estimating the Lebesgue measure of the resolvent set on the closed square $[-C, C]^2$, where $\|A\| \leq C$. We do not assume C is known. For $n_1, n_2 \in \mathbb{N}$, let

$$\texttt{Grid}(n_1, n_2) = \left(\frac{1}{2^{n_2}}\mathbb{Z} + \frac{1}{2^{n_2}}i\mathbb{Z}\right) \cap [-n_1, n_1]^2.$$

Letting $B(x,r), D(x,r)$ denote the closed and open balls of radius r around x respectively[3] in $\mathbb{C}$ (or $\mathbb{R}$ where appropriate), we define

$$U(n_1, n_2, A) = [-n_1, n_1] \times [-n_1, n_1] \cap (\cup_{z \in \texttt{Grid}(n_1,n_2)} B(z, F_{n_1}(z))).$$

[3] We set $D(x, 0) = \emptyset$.

Note that $\mathrm{Leb}(U(n_1,n_2,A))$ can be computed up to arbitrary predetermined precision using only arithmetic operations and comparisons by Proposition 8.2.1. Using this we can define

$$\Gamma_{n_2,n_1}(A) = 4n_1^2 - \mathrm{Leb}(U(n_1,n_2,A))$$

where, without loss of generality, we assume that we have computed the exact value of the Lebesgue measure (since we can absorb this error in the first limit). It is obvious that Γ_{n_2,n_1} are general arithmetical algorithms, the only non-trivial part is convergence.

We will now show that the algorithm `LebSpec` converges and realises the Π_2^A classification. There exists a compact set K such that $\|R(z,A)\|^{-1} > 1$ on K^c and without loss of generality we can make C larger, $C \in \mathbb{N}$ and take $K = [-C,C]^2$. For $n_1 \geq C$

$$U(n_1,n_2,A) = ([-C,C]^2 \cap (\cup_{z\in\mathtt{Grid}(n_1,n_2)} B(z,F_{n_1}(z)))) \cup ([-n_1,n_1]^2 \backslash [-C,C]^2)$$

since $F_n(z) \geq \|R(z,A)\|^{-1}$. It follows that for large n_1

$$\Gamma_{n_2,n_1}(A) = 4C^2 - \mathrm{Leb}([-C,C]^2 \cap (\cup_{z\in\mathtt{Grid}(n_1,n_2)} B(z,F_{n_1}(z)))).$$

As $n_1 \to \infty$, $[-C,C]^2 \cap (\cup_{z\in\mathtt{Grid}(n_1,n_2)} B(z,F_{n_1}(z)))$ converges to the closed set

$$X(n_2,A) = [-C,C]^2 \cap (\cup_{z\in\mathtt{Grid}(C,n_2)} B(z,\|R(z,A)\|^{-1}))$$

from above and hence

$$\lim_{n_1\to\infty} \Gamma_{n_2,n_1}(A) = 4C^2 - \mathrm{Leb}(X(n_2,A)),$$

from below. Consider the relatively open set

$$V(n_2,A) = [-C,C]^2 \cap (\cup_{z\in\mathtt{Grid}(C,n_2)} D(z,\|R(z,A)\|^{-1})).$$

Clearly $\mathrm{Leb}(X(n_2,A)) = \mathrm{Leb}(V(n_2,A))$ since the sets differ by a finite collection of circular arcs or points (recall we defined the open ball of radius zero to be the empty set). Hence we must show that

$$\lim_{n_2\to\infty} \mathrm{Leb}(V(n_2,A)) = \mathrm{Leb}(\rho_C(A)),$$

where $\rho_C(A) = [-C,C]^2 \backslash \mathrm{Sp}(A)$. For $z \in \rho_C(A)$,

$$\mathrm{dist}(z,\mathrm{Sp}(A)) \geq \|R(z,A)\|^{-1}$$

and hence we get $V(n_2,A) \subset \rho_C(A)$. Since $\rho_C(A)$ is relatively open, a simple density argument using the continuity of $\|R(z,A)\|^{-1}$ yields $V(n_2,A) \uparrow \rho_C(A)$ as $n_2 \to \infty$ since the grid refines itself. So we get

$$\mathrm{Leb}(V(n_2,A)) \uparrow \mathrm{Leb}(\rho_C(A)).$$

This proves the convergence and also shows that $\Gamma_{n_2}(A) \downarrow \Xi_1^L(A)$, thus yielding the Π_2^A classification. The same argument works in the one-dimensional case when considering self-adjoint operators Ω_{D} and $\mathrm{Leb}_{\mathbb{R}}$. Simply restrict everything to the real line and consider the interval $[-C,C]$ rather than a square.

Step 2: $\{\Xi_1^L, \Omega_f, \Lambda_i\}, \{\Xi_1^L, \Omega_{\mathrm{D}}, \Lambda_i\} \notin \Delta_2^G$. It is enough to consider Λ_2. We will only show that $\mathrm{SCI}(\Xi_1^L, \Omega_{\mathrm{D}}, \Lambda_2)_G \geq 2$ for which we use $\mathrm{Leb}_{\mathbb{R}}$ and the two-dimensional case is similar. Suppose for a contradiction that there exists a height one tower Γ_n, then $\Lambda_{\Gamma_n}(A)$ is finite for each $A \in \Omega_{\mathrm{D}}$. Hence, for every A and n there exists a finite number $N(A, n) \in \mathbb{N}$ such that the evaluations from $\Lambda_{\Gamma_n}(A)$ only take the matrix entries $A_{ij} = \langle Ae_j, e_i\rangle$ with $i, j \leq N(A, n)$ into account.

Pick any sequence $a_1, a_2, ...$ dense in the unit interval $[0, 1]$. Consider the matrix operators $A_m = \mathrm{diag}\{a_1, a_2, ..., a_m\} \in \mathbb{C}^{m\times m}$, $B_m = \mathrm{diag}\{0, 0, ..., 0\} \in \mathbb{C}^{m\times m}$ and $C = \mathrm{diag}\{0, 0, ...\}$. Set $A = \bigoplus_{m=1}^{\infty}(B_{k_m} \oplus A_{k_m})$ where we choose an increasing sequence k_m inductively as follows. Set $k_1 = 1$ and suppose that $k_1, ..., k_m$ have been chosen. $\mathrm{Sp}(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C) = \{0, a_1, a_2, ..., a_{k_m}\}$ and hence $\mathrm{Leb}(\mathrm{Sp}(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C)) = 0$ so there exists some $n_m \geq m$ such that if $n \geq n_m$ then

$$\Gamma_n(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C) \leq \frac{1}{2}.$$

Now let $k_{m+1} \geq \max\{N(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C, n_m), k_m + 1\}$. Any evaluation function $f_{i,j} \in \Lambda$ is simply the $(i, j)^{\text{th}}$ matrix entry and hence by construction

$$f_{i,j}(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C) = f_{i,j}(A),$$

for all $f_{i,j} \in \Lambda_{\Gamma_{n_m}}(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C)$. By assumption (iii) in Definition 2.1.1 it follows that $\Lambda_{\Gamma_{n_m}}(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C) = \Lambda_{\Gamma_{n_m}}(A)$ and hence by assumption (ii) in the same definition that $\Gamma_{n_m}(A) = \Gamma_{n_m}(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C) \leq 1/2$. But $\lim_{n\to\infty}(\Gamma_n(A)) = \mathrm{Leb}(\mathrm{cl}(\{0, a_1, a_2, ...\})) = 1$ a contradiction.

Step 3: $\{\Xi_1^L, \Omega, \Lambda_1\} \in \Pi_3^A$ for $\Omega = \Omega_{\mathrm{B}}, \Omega_{\mathrm{SA}}, \Omega_{\mathrm{N}}$ or Ω_g. We will deal with the case of Ω_{B}. The cases of Ω_{N} and Ω_g then follow via $\Omega_{\mathrm{N}} \subset \Omega_g \subset \Omega_{\mathrm{B}}$ and the one-dimensional Lebesgue measure case for Ω_{SA} is similar.

A careful analysis of the proof in step 1 yields that

- $\Gamma_{n_2,n_1}(A)$ converges to $\Gamma_{n_2}(A)$ from below as $n_1 \to \infty$.
- $\Gamma_{n_2}(A)$ converges to $\mathrm{Leb}(\mathrm{Sp}(A))$ monotonically from above as $n_2 \to \infty$.

We can ensure that the first limit converges from below by always slightly overestimating the Lebesgue measure of $U(n_1, n_2)$ (with error converging to zero) and using Proposition 8.2.1. These observations will be used later to answer question 3. We do not need to know c_n for the above proof to work, but we will need it for the first of the above facts. A slight alteration of the proof/algorithm by inserting an extra limit deals with the general case.

Define the function

$$\gamma_{n,m}(z; A) = \min\{\sigma_{\inf}(P_m(A - zI)|_{P_n\mathcal{H}}), \sigma_{\inf}(P_m(A^* - \bar{z}I)|_{P_n\mathcal{H}})\},$$

where $\sigma_{\inf}$ denotes the injection modulus/smallest singular value. One can show that $\gamma_{n,m}$ converges uniformly on compact subsets to

$$\gamma_n(z; A) = \min\{\sigma_{\inf}((A - zI)|_{P_n\mathcal{H}}), \sigma_{\inf}((A^* - \bar{z}I)|_{P_n\mathcal{H}})\},$$

as $m \to \infty$ and that this converges uniformly down to $\|R(z, A)\|^{-1}$ on compact subsets as $n \to \infty$ [Han11]. With a slight abuse of notation, we can approximate $\gamma_{n,m}(z; A)$ to within $1/m$ by $\mathtt{DistSpec}(A, n, m, z)$ (where the spacing of the search routine is $1/m$) so that this converges uniformly on compact subsets to $\gamma_n(z; A)$. In exactly the same manner as before, define

$$U(n_1, n_2, n_3, A) = [-n_2, n_2]^2 \cap (\cup_{z\in\mathtt{Grid}(n_2,n_3)} B(z, \gamma_{n_2,n_1}(z; A))),$$
$$\Gamma_{n_3,n_2,n_1}(A) = (2n_2)^2 - \mathrm{Leb}(U(n_1, n_2, n_3, A))$$

The stated uniform convergence means that the argument in step 1 carries through and we have a height three tower, realising the Π_3^A classification.

Step 4: $\{\Xi_1^L, \Omega_{\mathrm{SA}}, \Lambda_1\} \notin \Delta_3^G$. The proof is exactly the same argument as the proof of step 3 of Theorem 7.3.3. However, in this case to gain the contradiction, we then define $\tilde{\Gamma}_{n_2,n_1}(\{a_{i,j}\}) = \min\{\max\{1 - \Gamma_{n_2,n_1}(A)/2, 0\}, 1\}$ where $\Gamma_{n_2,n_1}(A)$ is the supposed height two tower for $\{\Xi_1^L, \Omega_{\mathrm{SA}}, \Lambda_1\}$.

Step 5: $\{\Xi_1^L, \Omega, \Lambda_1\} \notin \Delta_3^G$ for $\Omega = \Omega_{\mathrm{B}}, \Omega_{\mathrm{N}}$, or Ω_g. Since $\Omega_{\mathrm{N}} \subset \Omega_g \subset \Omega_{\mathrm{B}}$, we only need to deal with Ω_{N}. We can use a similar argument as in step 4, but now replacing each $C^{(j)}$ by

$$D^{(j)} = \bigoplus_{k=1}^{j} ih_k C^{(j)},$$

where $h_1, h_2, ...$ is a dense sequence in $[0,1]$ and this operators acts on $X_j = \bigoplus_{k=1}^{j} l^2(\mathbb{N})$. This ensures that the spectrum of the operator yields a positive two-dimensional Lebesgue measure if and only if $\tilde{\Xi}_2(\{a_{i,j}\}) = 0$. The rest of the argument is entirely analogous.

Step 6: $\Delta_2^G \not\ni \{\Xi_1^L, \Omega, \Lambda_2\} \in \Pi_2^A$ for $\Omega = \Omega_{\mathrm{B}}, \Omega_{\mathrm{SA}}, \Omega_{\mathrm{N}}$ or Ω_g. The impossibility result follows by considering diagonal operators. For the existence of Π_2^A algorithms, we can use the construction in step 3, but the knowledge of matrix values of A^*A allows us to skip the first limit and approximate γ_n directly. □

Proof of Theorem 8.1.2. Using the convergence

$$\lim_{\epsilon \downarrow 0} \mathrm{Leb}(\widehat{\mathrm{Sp}}_\epsilon(A)) = \mathrm{Leb}(\mathrm{Sp}(A)),$$

the lower bounds in Theorem 8.1.1 immediately imply the lower bounds in Theorem 8.1.2. Hence we only need to construct the appropriate algorithms.

Step 1: $\{\Xi_2^L, \Omega_f, \Lambda_1\}, \{\Xi_2^L, \Omega_{\mathrm{D}}, \Lambda_1\} \in \Sigma_1^A$. Let $A \in \Omega_f$ and

$$E_n = \frac{1}{n}(\mathbb{Z} + i\mathbb{Z}) \cap \{z \in \mathbb{C} : F_n(z) \leq \epsilon\} \cap [-n, n]^2.$$

Clearly, we can compute E_n with finitely many arithmetic operations and comparisons, and we set

$$\Gamma_n(A) = \mathrm{Leb}\big(\cup_{z \in E_n} D(z, \max\{0, \epsilon - F_n(z)\})\big).$$

Proposition 8.2.1 shows that, without loss of generality, we can assume $\Gamma_n(A)$ can be computed exactly with finitely many arithmetic operations and comparisons.

Suppose that $F_n(z) < \epsilon$ and that $|w| < \epsilon - F_n(z)$. If $z \in \mathrm{Sp}(A)$ then clearly

$$\|R(z+w, A)\|^{-1} \leq |w| < \epsilon - F_n(z) \leq \epsilon,$$

and this holds trivially if $z + w \in \mathrm{Sp}(A)$ so assume that neither of $z, z+w$ are in the spectrum. The resolvent identity yields

$$\|R(z+w, A)\| \geq \|R(z, A)\| - |w| \, \|R(z+w, A)\| \, \|R(z, A)\|,$$

which rearranges to

$$\|R(z+w, A)\|^{-1} \leq \|R(z, A)\|^{-1} + |w| < \epsilon.$$

It follows that $\cup_{z \in E_n} D(z, \max\{0, \epsilon - F_n(z)\})$ is in $\widehat{\mathrm{Sp}}_\epsilon(A)$ and hence that $\Gamma_n(A) \leq \Xi_2^L(A)$. Without loss of generality by taking successive maxima we can assume that $\Gamma_n(A)$ is increasing. Together these will yield Σ_1^A once convergence is shown. Using the uniform convergence of F_n and density of $1/n(\mathbb{Z}+i\mathbb{Z})\cap[-n,n]^2$ we see that pointwise convergence holds:

$$\chi_{\cup_{z\in E_n} D(z,\max\{0,\epsilon-F_n(z)\}} \rightarrow \chi_{\widehat{\mathrm{Sp}}_\epsilon(A)},$$

where χ_E denotes the indicator function of a set E. It follows by the dominated convergence theorem that $\Gamma_n(A)\rightarrow \mathrm{Leb}(\widehat{\mathrm{Sp}}_\epsilon(A))$. The proof for Ω_{D} is similar by restricting everything to the real line.

Step 2: $\{\Xi_2^L,\Omega,\Lambda_1\}\in\Sigma_2^A$ for $\Omega=\Omega_{\mathrm{B}},\Omega_{\mathrm{SA}},\,\Omega_{\mathrm{N}}$ or Ω_g. To prove this, we simply replace F_{n_1} by the functions γ_{n_2,n_1} and set

$$\Gamma_{n_2,n_1}(A)=\mathrm{Leb}\big(\cup_{z\in E_{n_2}} D(z,\max\{0,\epsilon-\gamma_{n_2,n_1}(z;A)\})\big).$$

Step 3: $\{\Xi_2^L,\Omega,\Lambda_2\}\in\Sigma_1^A$ for $\Omega=\Omega_{\mathrm{B}},\Omega_{\mathrm{SA}},\,\Omega_{\mathrm{N}}$ or Ω_g. The knowledge of matrix values of A^*A allows us to skip the first limit in the construction of step 2 and approximate γ_n directly. □

Finally, we deal with the question of determining if the Lebesgue measure is zero. Recall that for this problem, $(\mathcal{M},d)$ denotes the set $\{0,1\}$ endowed with the discrete topology and we consider the problem function

$$\Xi_3^L(A)=\begin{cases}0, & \text{if } \mathrm{Leb}(\mathrm{Sp}(A))>0\\ 1, & \text{otherwise.}\end{cases}$$

Proof of Theorem 8.1.4. We will show that $\{\Xi_3^L,\Omega_f,\Lambda_1\}\in\Pi_3^A$ and $\{\Xi_3^L,\Omega_{\mathrm{D}},\Lambda_2\}\notin\Delta_3^G$. The analogous statements $\{\Xi_3^L,\Omega_{\mathrm{D}},\Lambda_1\}\in\Pi_3^A$ and $\{\Xi_3^L,\Omega_f,\Lambda_2\}\notin\Delta_3^G$ follow from similar arguments.

The lower bound argument can also be used when considering Λ_2 and $\Omega=\Omega_{\mathrm{B}},\Omega_{\mathrm{SA}},\,\Omega_{\mathrm{N}}$ or Ω_g. We will also prove the lower bound $\{\Xi_3^L,\Omega_{\mathrm{SA}},\Lambda_1\}\notin\Delta_4^G$. The remaining lower bounds for Λ_1 follow from a similar argument and construction as in step 5 of the proof of Theorem 8.1.1 to ensure we are dealing with two-dimensional Lebesgue measure. Finally, we prove that $\{\Xi_3^L,\Omega_{\mathrm{B}},\Lambda_1\}\in\Pi_4^A$. The upper bounds for $\Omega=\Omega_{\mathrm{SA}},\,\Omega_{\mathrm{N}}$ or Ω_g and Λ_1 follow from an almost identical argument. When considering Λ_2, we can collapse the first limit in exactly the same manner as we did for solving Ξ_1^L.

Step 1: $\{\Xi_3^L,\Omega_f,\Lambda_1\}\in\Pi_3^A$. First we use the algorithm used to compute Ξ_1^L in Theorem 8.1.1, which we shall denote by $\widetilde{\Gamma}$, to build a height 3 tower for $\{\Xi_3^L,\Omega_f\}$. As above, Ω_f denotes the set of bounded operators with the usual assumption of bounded dispersion (now with known bounds c_n). Recall that we observed

- $\widetilde{\Gamma}_{n_2,n_1}(A)$ converges to $\widetilde{\Gamma}_{n_2}(A)$ from below as $n_1\rightarrow\infty$.
- $\widetilde{\Gamma}_{n_2}(A)$ converges to $\mathrm{Leb}(\mathrm{Sp}(A))$ monotonically from above as $n_2\rightarrow\infty$.

We can alter our algorithms, by taking maxima, so that we can assume without loss of generality that $\widetilde{\Gamma}_{n_2,n_1}(A)$ converges to $\widetilde{\Gamma}_{n_2}(A)$ monotonically from below as $n_1\rightarrow\infty$. Now let

$$\Gamma_{n_3,n_2,n_1}(A)=\chi_{[0,1/n_3]}(\widetilde{\Gamma}_{n_2,n_1}(A)).$$

Note that $\chi_{[0,1/n_3]}$ is left continuous on $[0,\infty)$ with right limits. Hence by the assumed monotonicity

$$\lim_{n_1\rightarrow\infty}\Gamma_{n_3,n_2,n_1}(A)=\chi_{[0,1/n_3]}(\widetilde{\Gamma}_{n_2}(A)).$$

It follows that

$$\lim_{n_2\rightarrow\infty}\lim_{n_1\rightarrow\infty}\Gamma_{n_3,n_2,n_1}(A)=\chi_{[0,1/n_3]}(\mathrm{Leb}(\mathrm{Sp}(A))\pm),$$

where ± denotes one of the right or left limits (it is possible to have either). It is then easy to see that

$$\lim_{n_3\to\infty}\lim_{n_2\to\infty}\lim_{n_1\to\infty}\Gamma_{n_3,n_2,n_1}(A)=\Xi_3^L(A).$$

It is also clear that the answer to the question is 0 if $\Gamma_{n_3}(A)=0$, which yields the Π_3^A classification.

Step 2: $\{\Xi_3^L,\Omega_{\mathrm{D}},\Lambda_1\}\notin\Delta_3^G$. Assume for a contradiction that this is false and $\widehat{\Gamma}_{n_2,n_1}$ is a general height two tower for $\{\Xi_3^L,\Omega_{\mathrm{D}}\}$. Let $(\mathcal{M},d)$ be discrete space $\{0,1\}$ and $\tilde{\Omega}$ denote the collection of all infinite matrices $\{a_{i,j}\}_{i,j\in\mathbb{N}}$ with entries $a_{i,j}\in\{0,1\}$ and consider the problem function

$$\tilde{\Xi}_1(\{a_{i,j}\}):\text{ Does }\{a_{i,j}\}\text{ have a column containing infinitely many non-zero entries?}$$

Recall that it was shown in Theorem 2.3.7 in Chapter 2 §2.3 that $\mathrm{SCI}(\tilde{\Xi}_1,\tilde{\Omega})_G=3$. We will gain a contradiction by using the supposed height two tower to solve $\{\tilde{\Xi}_1,\tilde{\Omega}\}$.

For $j\in\mathbb{N}$, let $\{b_{i,j}\}_{i\in\mathbb{N}}$ be a dense subset of $I_j:=[1-1/2^{j-1},1-1/2^j]$. Given a matrix $\{a_{i,j}\}_{i,j\in\mathbb{N}}\in\tilde{\Omega}$, construct a matrix $\{c_{i,j}\}_{i,j\in\mathbb{N}}$ by letting $c_{i,j}=a_{i,j}b_{r(i,j),j}$ where

$$r(i,j)=\max\left\{1,\sum_{k=1}^{i}a_{k,j}\right\}.$$

Now consider any bijection $\phi:\mathbb{N}\to\mathbb{N}^2$ and define the diagonal operator

$$A=\mathrm{diag}(c_{\phi(1)},c_{\phi(2)},c_{\phi(3)},...).$$

The algorithm $\widehat{\Gamma}_{n_2,n_1}$ thus translates to an algorithm defined by Γ'_{n_2,n_1} for $\{\tilde{\Xi}_1,\tilde{\Omega}\}$. Namely, we set $\Gamma'_{n_2,n_1}(\{a_{i,j}\}_{i\in\mathbb{N}})=\widehat{\Gamma}_{n_2,n_1}(A)$. The fact that ϕ is a bijection shows that the lowest level Γ'_{n_2,n_1} are generalised algorithms (and are consistent). In particular, given N, we can find $\{A_{i,j}:i,j\leq N\}$ using finitely many evaluations of the matrix values $\{c_{k,l}\}$. But for any given $c_{k,l}$ we can evaluate this entry using only finitely many evaluations of the matrix values $\{a_{m,n}\}$ by the construction of r. Finally note that

$$\mathrm{Sp}(A)=\left(\bigcup_{j:\sum_i a_{i,j}=\infty}I_j\right)\cup Q,$$

where Q is at most countable. Hence

$$\mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}(A))=\sum_{j:\sum_i a_{i,j}=\infty}\frac{1}{2^j}.$$

It follows that $\tilde{\Xi}_1(\{a_{i,j}\})=\Xi_3^L(A)$ and hence we get a contradiction.

Step 3: $\{\Xi_3^L,\Omega_{\mathrm{SA}},\Lambda_1\}\notin\Delta_4^G$. Suppose for a contradiction that Γ_{n_3,n_2,n_1} is a height three tower of general algorithms for the problem $\{\Xi_3^L,\Omega_{\mathrm{SA}},\Lambda_1\}$. Let $(\mathcal{M},d)$ be the space $\{0,1\}$ with the discrete metric, let $\tilde{\Omega}$ denote the collection of all infinite arrays $\{a_{m,i,j}\}_{m,i,j\in\mathbb{N}}$ with entries $a_{m,i,j}\in\{0,1\}$ and consider the problem function

$$\tilde{\Xi}_4(\{a_{m,i,j}\}):\text{ For every }m\text{, does }\{a_{m,i,j}\}_{i,j}\text{ have (only) finitely many columns}$$
$$\text{with (only) finitely many 1's?}$$

Recall that it was shown in Theorem 2.3.7 in Chapter 2 §2.3 that $\mathrm{SCI}(\tilde{\Xi}_4,\tilde{\Omega})_G=4$. We will gain a contradiction by using the supposed height three tower to solve $\{\tilde{\Xi}_4,\tilde{\Omega}\}$.

The construction follows step 3 of the proof of Theorem 7.3.3 closely. For fixed m, recall the construction of the operator $A_m:=A(\{a_{m,i,j}\}_{i,j})$ from that proof, the key property being that if $\{a_{m,i,j}\}_{i,j}$ has (only) finitely many columns with (only) finitely many 1's then $\mathrm{Sp}(A_m)$ is a finite subset of $[-1,1]$, otherwise it is the whole interval $[-1,1]$. Now consider the intervals $I_m = [1-2^{m-1}, 1-2^m]$ and affine maps, α_m, that act as a bijection from $[-1,1]$ to I_m. Without loss of generality, identify Ω_{SA} with self adjoint operators in $\mathcal{B}(X)$ where $X = \bigoplus_{i=1}^{\infty}\bigoplus_{j=1}^{\infty} X_{i,j}$ in the l^2-sense with $X_{i,j} = l^2(\mathbb{N})$. We then consider the operator

$$T(\{a_{m,i,j}\}_{m,i,j}) = \bigoplus_{m=1}^{\infty} \alpha_m(A_m).$$

The same arguments in the proof of Theorem 7.3.3 show that the map

$$\tilde{\Gamma}_{n_3,n_2,n_1}(\{a_{m,i,j}\}_{m,i,j}) = \Gamma_{n_3,n_2,n_1}(T(\{a_{m,i,j}\}_{m,i,j}))$$

is a general tower using the relevant pointwise evaluation functions of the array $\{a_{m,i,j}\}_{m,i,j}$. If it holds that $\tilde{\Xi}_4(\{a_{m,i,j}\}) = 1$, then $\mathrm{Sp}(T(\{a_{m,i,j}\}_{m,i,j}))$ is countable and hence $\Xi_3^L(T(\{a_{m,i,j}\}_{m,i,j})) = 1$. On the other hand, if $\tilde{\Xi}_4(\{a_{m,i,j}\}) = 0$, then there exists m with $\mathrm{Sp}(A_m) = [-1,1]$ and hence $I_m \subset \mathrm{Sp}(T(\{a_{m,i,j}\}_{m,i,j}))$ so that $\Xi_3^L(T(\{a_{m,i,j}\}_{m,i,j})) = 0$. It follows that $\tilde{\Gamma}_{n_3,n_2,n_1}$ provides a height three tower for $\{\tilde{\Xi}_4, \tilde{\Omega}\}$, a contradiction.

Step 4: $\{\Xi_3^L, \Omega_{\mathrm{B}}, \Lambda_1\} \in \Pi_4^A$. Recall the tower of algorithms to solve $\{\Xi_1^L, \Omega_{\mathrm{B}}, \Lambda_1\}$, and denote it by $\widetilde{\Gamma}$. Our strategy will be the same as in step 1 but with an extra limit. It is easy to show that

- $\widetilde{\Gamma}_{n_3,n_2,n_1}(A)$ converges to $\widetilde{\Gamma}_{n_3,n_2}(A)$ from above as $n_1 \to \infty$.
- $\widetilde{\Gamma}_{n_3,n_2}(A)$ converges to $\widetilde{\Gamma}_{n_3}(A)$ from below as $n_2 \to \infty$.
- $\widetilde{\Gamma}_{n_3}(A)$ converges to $\mathrm{Leb}(\mathrm{Sp}(A))$ from above as $n_3 \to \infty$.

Again, by taking successive maxima or minima where appropriate, we can assume that all of these are monotonic. Now let

$$\Gamma_{n_4,n_3,n_2,n_1}(A) = \chi_{[0,1/n_4]}(\widetilde{\Gamma}_{n_3,n_2,n_1}(A)).$$

Note that $\chi_{[0,1/n_4]}$ is left continuous on $[0,\infty)$ with right limits. Hence by the assumed monotonicity and arguments as in step 1, it is then easy to see that

$$\lim_{n_4\to\infty}\lim_{n_3\to\infty}\lim_{n_2\to\infty}\lim_{n_1\to\infty}\Gamma_{n_4,n_3,n_2,n_1}(A) = \Xi_3^L(A).$$

It is also clear that the answer to the question is 0 if $\Gamma_{n_4}(A) = 0$, which yields the Π_4^A classification. □

8.3 Proofs of Theorems on Fractal Dimensions

We begin with the box-counting dimension. For the construction of towers of algorithms, it is useful to use a slightly different (but equivalent - see [Fal03]) definition of the upper and lower box-counting dimensions. Let $F \subset \mathbb{R}$ be bounded and $N'_\delta(F)$ denote the number of δ-mesh intervals that intersect F. A δ-mesh interval is an interval of the form $[m\delta, (m+1)\delta]$ for $m \in \mathbb{Z}$. Then

$$\overline{\dim}_B(F) = \limsup_{\delta\downarrow 0}\frac{\log(N'_\delta(F))}{\log(1/\delta)},$$
$$\underline{\dim}_B(F) = \liminf_{\delta\downarrow 0}\frac{\log(N'_\delta(F))}{\log(1/\delta)}.$$

Proof of Theorem 8.1.5. Since $\Omega_{BD}^{\mathrm{D}} \subset \Omega_f^{BD} \subset \Omega_{\mathrm{SA}}^{BD}$, it is enough to prove that $\{\Xi_B, \Omega_f^{BD}, \Lambda_1\} \in \Pi_2^A$, $\{\Xi_B, \Omega_{\mathrm{SA}}^{BD}, \Lambda_2\} \in \Pi_2^A$, $\{\Xi_B, \Omega_{\mathrm{SA}}^{BD}, \Lambda_1\} \in \Pi_3^A$, $\{\Xi_B, \Omega_{\mathrm{SA}}^{BD}, \Lambda_1\} \notin \Delta_3^A$ and $\{\Xi_B, \Omega_{\mathrm{D}}^{BD}, \Lambda_2\} \notin \Delta_2^A$.

Step 1: $\{\Xi_B, \Omega_f^{BD}, \Lambda_1\} \in \Pi_2^A$. Recall the existence of a height one tower, $\tilde{\Gamma}_n$, using Λ_1 for $\mathrm{Sp}(A)$, $A \in \Omega_f^{BD}$ from Chapter 3. Furthermore, $\tilde{\Gamma}_n(A)$ outputs a finite collection $\{z_{1,n}, ..., z_{k_n,n}\} \subset \mathbb{Q}$ such that $\mathrm{dist}(z_{j,n}, \mathrm{Sp}(A)) \leq 2^{-n}$. Define the intervals

$$I_{j,n} = [z_{j,n} - 2^{-n}, z_{j,n} + 2^{-n}]$$

and let $\mathcal{I}_m$ denote the collection of all 2^{-m}-mesh intervals. Let $\Upsilon_{m,n}(A)$ be any union of finitely many such mesh intervals with minimal length $|\Upsilon_{m,n}(A)|$ ('length' being the number of intervals $\in \mathcal{I}_m$ that make up $\Upsilon_{m,n}(A)$) such that

$$\Upsilon_{m,n}(A) \cap I_{j,l} \neq \emptyset, \quad \text{for } 1 \leq l \leq n, 1 \leq j \leq k_l.$$

There may be more than one such collection so we can gain a deterministic algorithm by enumerating each $\mathcal{I}_m$ and choosing the first such collection in this enumeration. It is then clear that $|\Upsilon_{m,n}(A)|$ is increasing in n. Furthermore, to determine $\Upsilon_{m,n}(A)$, there are only finitely many intervals in $\mathcal{I}_m$ to consider, namely those that have non-empty intersection with at least one $I_{j,l}$ with $1 \leq l \leq n, 1 \leq j \leq k_l$. It follows that $\Upsilon_{m,n}(A)$ and hence $|\Upsilon_{m,n}(A)|$ can be computed in finitely may arithmetic operations and comparisons using Λ_1.

Suppose that $I = [a, b] \in \mathcal{I}_m$ has $(a, b) \cap \mathrm{Sp}(A) \neq \emptyset$. Then for large n there exists $z_{j,n} \in I$ such that $I_{j,n} \subset I$ and hence $I \subset \Upsilon_{m,n}(A)$ for large n. If $z \in \mathrm{Sp}(A) \cap 2^{-m}\mathbb{Z}$ then a similar argument shows that $z \subset \Upsilon_{m,n}(A)$ for large n. Since $\mathrm{Sp}(A)$ is bounded and $\mathrm{Sp}(A) \cap 2^{-m}\mathbb{Z}$ finite, it follows that $\mathrm{Sp}(A) \subset \Upsilon_{m,n}(A)$ for large n and hence

$$N_{2^{-m}}(\mathrm{Sp}(A)) \leq \liminf_{n\to\infty} |\Upsilon_{m,n}(A)|.$$

Let $W_m(A)$ be the union of all intervals in $\mathcal{I}_m$ that intersect $\mathrm{Sp}(A)$. It is clear that $W_m(A) \cap I_{j,l} \neq \emptyset$ for $1 \leq l \leq n, 1 \leq j \leq k_l$ and hence $|\Upsilon_{m,n}(A)| \leq N'_{2^{-m}}(\mathrm{Sp}(A))$. It follows that $\lim_{n\to\infty} |\Upsilon_{m,n}(A)| = \delta_m(A)$ exists with

$$N_{2^{-m}}(\mathrm{Sp}(A)) \leq \delta_m(A) \leq N'_{2^{-m}}(\mathrm{Sp}(A)). \tag{8.3.1}$$

For $n_2 > n_1$ set $\Gamma_{n_2,n_1}(A) = 0$, otherwise set

$$\Gamma_{n_2,n_1}(A) = \max_{n_2 \leq k \leq n_1} \max_{1 \leq j \leq n_1} \frac{\log(|\Upsilon_{k,j}(A)|)}{k \log(2)}.$$

The above monotone convergence and (8.3.1) shows that

$$\lim_{n_1\to\infty} \Gamma_{n_2,n_1}(A) = \Gamma_{n_2}(A) = \sup_{k \geq n_2} \frac{\log(\delta_k(A))}{k\log(2)} \geq \limsup_{k\to\infty} \frac{\log(\delta_k(A))}{k\log(2)},$$
$$\lim_{n_2\to\infty} \Gamma_{n_2}(A) = \limsup_{k\to\infty} \frac{\log(\delta_k(A))}{k\log(2)}.$$

Hence, by the assumption that the box-counting dimension exists, we have constructed a Π_2^A tower.

Step 2: $\{\Xi_B, \Omega_{\mathrm{SA}}^{BD}, \Lambda_2\} \in \Pi_2^A$ and $\{\Xi_B, \Omega_{\mathrm{SA}}^{BD}, \Lambda_1\} \in \Pi_3^A$. The first of these is exactly as in step 1, using Λ_2 to construct the relevant Σ_1^A tower for the spectrum. The proof that $\{\Xi_B, \Omega_{\mathrm{SA}}^{BD}, \Lambda_1\} \in \Pi_3^A$ uses a height two tower, $\tilde{\Gamma}_{n_2,n_1}$, using Λ_1 for $\mathrm{Sp}(A)$, $A \in \Omega_{\mathrm{SA}}^{BD}$ (or any self-adjoint A) constructed in [BACH+20]. This tower has the property that each $\tilde{\Gamma}_{n_2,n_1}(A)$ is a finite subset of $\mathbb{Q}$ and, for fixed n_2, is constant for large n_1. Moreover if $z \in \lim_{n_1\to\infty} \tilde{\Gamma}_{n_2,n_1}(A)$ then $\mathrm{dist}(z, \mathrm{Sp}(A)) \leq 2^{-n_2}$. It follows that we can use the same construction as step 1 with an additional limit at the start to reach the finite set $\lim_{n_1\to\infty}\tilde{\Gamma}_{n_2,n_1}(A)$.

Step 3: $\{\Xi_B,\Omega_{\mathrm{D}}^{BD},\Lambda_2\}\notin\Delta_2^A$. This is exactly the same argument as step 2 of the proof of Theorem 8.1.1 with Lebesgue measure replaced by box-counting dimension.

Step 4: $\{\Xi_B,\Omega_{\mathrm{SA}}^{BD},\Lambda_1\}\notin\Delta_3^A$. This is exactly the same argument as step 4 of the proof of Theorem 8.1.1 with Lebesgue measure replaced by box-counting dimension. □

We now turn to the Hausdorff dimension. Recall Lemma 8.1.7 on the problem of determining whether $\mathrm{Sp}(A)\cap(a,b)\neq\emptyset$.

Proof of Lemma 8.1.7. We start with the class $\Omega_f\cap\Omega_{\mathrm{SA}}$. We can interpret this problem as a decision problem and the following algorithm as one that halts on output yes. Let $c=(a+b)/2$ and $\delta=(b-a)/2$ then the idea is to simply test whether $\texttt{DistSpec}(A,n,f(n),c)+c_n<\delta$. If the answer is yes then we output yes, otherwise we output no and increase n by one. Note that $\mathrm{Sp}(A)\cap(a,b)\neq\emptyset$ if and only if $\|R(c,A)\|^{-1}<\delta$ and hence as $\texttt{DistSpec}(A,n,f(n),c)+c_n$ converges down to $\|R(c,A)\|^{-1}$ we see that this provides a convergent algorithm. For Ω_{SA} we require an additional limit by replacing $\texttt{DistSpec}(A,n,f(n),c)+c_n$ with the function $\gamma_{n_2,n_1}(z;A)$. If we have access to Λ_2 then this can be avoided in the usual way. □

To build our algorithm for the Hausdorff dimension, we use an alternative, equivalent definition for compact sets that can be found in [FMSG15, FMSG14]. We consider the case of subsets of $\mathbb{R}$. Let ρ_k denote the set of all closed binary cubes of the form $[2^{-k}m,2^{-k}(m+1)],m\in\mathbb{Z}$. Set

$$\mathcal{A}_k(F)=\left\{\{U_i\}_{i\in I}: I\text{ is finite },F\subset\cup_{i\in I}U_i,U_i\in\cup_{l\geq k}\rho_l\right\}$$

and define

$$\tilde{\mathcal{H}}_k^d(F)=\inf\left\{\sum_i\mathrm{diam}(U_i)^d:\{U_i\}_{i\in I}\in\mathcal{A}_k(F)\right\},\quad \tilde{\mathcal{H}}^d(F)=\lim_{k\to\infty}\tilde{\mathcal{H}}_k^d(F).$$

The following can be found in [FMSG14] (Theorem 3.13):

Theorem 8.3.1 ([FMSG14])**.** *Let F be a bounded subset of $\mathbb{R}$. Then there exists a unique $d'=\dim_{H'}(F)$ such that $\tilde{\mathcal{H}}^d(F)=0$ for $d>d'$ and $\tilde{\mathcal{H}}^d(F)=\infty$ for $d<d'$. Furthermore, $d'=\dim_H(\mathrm{cl}(F))$.*

Denoting the dyadic rationals by $\mathbb{D}$, we shall compute $\dim_H(\mathrm{Sp}(A))$ via approximating the above applied to $F=\mathrm{Sp}(A)\cap\mathbb{D}^c$ and using the lemma 8.1.7.

Proof of Theorem 8.1.8. It is enough to prove the lower bounds $\{\Xi_H,\Omega_{\mathrm{D}},\Lambda_2\}\notin\Delta_3^G$, $\{\Xi_H,\Omega_{\mathrm{SA}},\Lambda_1\}\notin\Delta_4^G$ and construct the towers of algorithms for the inclusions $\{\Xi_H,\Omega_f\cap\Omega_{\mathrm{SA}},\Lambda_1\}\in\Sigma_3^A$, $\{\Xi_H,\Omega_{\mathrm{SA}},\Lambda_1\}\in\Sigma_4^A$ and $\{\Xi_H,\Omega_{\mathrm{SA}},\Lambda_2\}\in\Sigma_3^A$.

Step 1: $\{\Xi_H,\Omega_{\mathrm{D}},\Lambda_2\}\notin\Delta_3^G$. Suppose for a contradiction that a height two tower, Γ_{n_2,n_1}, exists for $\{\Xi_H,\Omega_{\mathrm{D}}\}$ (taking values in $[0,1]$ without loss of generality). We repeat the argument in the proof of Theorem 8.1.4. Consider the same problem

$$\tilde{\Xi}_1(\{a_{i,j}\}):\ \text{Does }\{a_{i,j}\}\text{ have a column containing infinitely many non-zero entries?}$$

but now mapping to $[0,1]$ with the usual metric, and the same operator $A = \mathrm{diag}(c_{\phi(1)}, c_{\phi(2)}, c_{\phi(3)}, ...)$ with

$$\mathrm{Sp}(A) = \left(\bigcup_{j:\sum_i a_{i,j}=\infty} I_j \right) \cup Q,$$

where Q is at most countable. We use the fact that the Hausdorff dimension satisfies

$$\dim_H(\cup_{j=1}^{\infty} X_j) = \sup_{j\in\mathbb{N}} \dim_H(X_j)$$

and that $\dim_H(Q) = 0$ for any countable Q, to note that the equality $\Xi_H(A) = \tilde{\Xi}_1(\{a_{i,j}\})$ holds. We then set $\tilde{\Gamma}_{n_2,n_1}(\{a_{i,j}\}_{i,j}) = \Gamma_{n_2,n_1}(A)$ to provide a height two tower for $\tilde{\Xi}_1$. But this contradicts Theorem 2.3.7.

Step 2: $\{\Xi_H, \Omega_{\mathrm{SA}}, \Lambda_1\} \notin \Delta_4^G$. Suppose for a contradiction that Γ_{n_3,n_2,n_1} is a height three tower of general algorithms for the problem $\{\Xi_H, \Omega_{\mathrm{SA}}, \Lambda_1\}$ (taking values in $[0,1]$ without loss of generality). Let $(\mathcal{M}, d)$ be the space $[0,1]$ with the usual metric, let $\tilde{\Omega}$ denote the collection of all infinite arrays $\{a_{m,i,j}\}_{m,i,j\in\mathbb{N}}$ with entries $a_{m,i,j} \in \{0,1\}$ and consider the problem function

$\tilde{\Xi}_4(\{a_{m,i,j}\})$: For every m, does $\{a_{m,i,j}\}_{i,j}$ have (only) finitely many columns
with (only) finitely many 1's?

Recall that it was shown in Theorem 2.3.7 in Chapter 2 §2.3 that $\mathrm{SCI}(\tilde{\Xi}_4, \tilde{\Omega})_G = 4$. We will gain a contradiction by using the supposed height three tower to solve $\{\tilde{\Xi}_4, \tilde{\Omega}\}$.

We use the same construction as in step 3 of the proof of Theorem 8.1.4. If $\tilde{\Xi}_4(\{a_{m,i,j}\}) = 1$, then $\mathrm{Sp}(T(\{a_{m,i,j}\}_{m,i,j}))$ is countable and hence $\Xi_H(T(\{a_{m,i,j}\}_{m,i,j})) = 0$. On the other hand, if $\tilde{\Xi}_4(\{a_{m,i,j}\}) = 0$, then there exists m with $\mathrm{Sp}(A_m) = [-1,1]$ and hence $I_m \subset \mathrm{Sp}(T(\{a_{m,i,j}\}_{m,i,j}))$ so that $\Xi_H(T(\{a_{m,i,j}\}_{m,i,j})) = 1$. It follows that $\tilde{\Gamma}_{n_3,n_2,n_1}(\{a_{m,i,j}\}_{m,i,j}) = 1 - \Gamma_{n_3,n_2,n_1}(T(\{a_{m,i,j}\}_{m,i,j}))$ provides a height three tower for $\{\tilde{\Xi}_4, \tilde{\Omega}\}$, a contradiction.

Step 3: $\{\Xi_H, \Omega_f \cap \Omega_{\mathrm{SA}}, \Lambda_1\} \in \Sigma_3^A$. To construct a height three tower for $A \in \Omega_f \cap \Omega_{\mathrm{SA}}$, if $n_2 < n_3$ set $\Gamma_{n_3,n_2,n_1}(A) = 0$. Otherwise, consider the set

$$\mathcal{A}_{n_3,n_2,n_1}(A) = \{\{U_i\}_{i\in I} : I \text{ is finite }, S_{n_1,n_2}(A) \subset \cup_{i\in I} U_i, U_i \in \cup_{n_3\leq l\leq n_2} \rho_l\}$$

where $S_{n_1,n_2}(A)$ is the union of all $S \in \rho_{n_2}$ with $S \subset [-n_1, n_1]$ and such that the algorithm discussed in Lemma 8.1.7 outputs yes for the interior of S and input parameter n_1. We then define

$$h_{n_3,n_2,n_1}(A,d) = \inf \left\{ \sum_i \mathrm{diam}(U_i)^d : \{U_i\} \in \mathcal{A}_{n_3,n_2,n_1}(A) \right\}.$$

If $S_{n_1,n_2}(A)$ is empty then we interpret the infinum as 0. There are only finitely many sets to check and hence the infinum is a minimisation problem over finitely many coverings (see §8.4.2 for a discussion of efficient implementation). It follows that $h_{n_3,n_2,n_1}(A,d)$ defines a general algorithm computable in finitely many arithmetic operations and comparisons. Furthermore, it is easy to see that

$$\lim_{n_1\to\infty} h_{n_3,n_2,n_1}(A,d) = \inf \left\{ \sum_i \mathrm{diam}(U_i)^d : \{U_i\} \in \mathcal{C}_{n_3,n_2}(A) \right\} =: h_{n_3,n_2}(A,d)$$

from below (since we are covering larger sets as n_1 increases), where

$$\mathcal{C}_{n_3,n_2}(A) = \left\{\{U_i\}_{i\in I} : I \text{ is finite }, \mathrm{Sp}(A) \cap \mathbb{D}_{n_2}^c \subset \cup_{i\in I} U_i, U_i \in \cup_{n_3\leq l\leq n_2} \rho_l \right\}$$

and $\mathbb{D}_k := 1/2^k \cdot \mathbb{Z}$ denotes the dyadic rationals of resolution k. We now use the property that $\mathcal{A}_k(F)$ consists of collections of finite coverings. As $n_2 \to \infty$, $h_{n_3,n_2}(A,d)$ is non-increasing (since we take infinum over a larger class of coverings and the sets $\mathrm{Sp}(A) \cap \mathbb{D}^c_{n_2}$ decrease) and hence converges to some number. Clearly

$$\lim_{n_2\to\infty} h_{n_3,n_2}(A,d) =: h_{n_3}(A,d) \geq \tilde{\mathcal{H}}^d_{n_3}(\mathrm{Sp}(A)\cap\mathbb{D}^c).$$

For $\epsilon > 0$ let $l \in \mathbb{N}$ and $\{U_i\} \in \mathcal{A}_{n_3}(\mathrm{Sp}(A)\cap\mathbb{D}^c_l)\}$ with

$$\sum_i \mathrm{diam}(U_i)^d \leq \epsilon + \tilde{\mathcal{H}}^d_{n_3}(\mathrm{Sp}(A)\cap\mathbb{D}^c_l).$$

For large enough n_2, $\{U_i\} \in \mathcal{C}_{n_3,n_2}(A)$ and hence since $\epsilon > 0$ was arbitrary,

$$h_{n_3}(A,d) \leq \tilde{\mathcal{H}}^d_{n_3}(\mathrm{Sp}(A)\cap\mathbb{D}^c_l)$$

for all l. For a fixed A and d, $h_{n_3}(A,d)$ is non-decreasing in n_3 and hence converges to a function of d, $h(A,d)$ (possibly taking infinite values). Furthermore,

$$\tilde{\mathcal{H}}^d(\mathrm{Sp}(A)\cap\mathbb{D}^c) \leq h(A,d) \leq \tilde{\mathcal{H}}^d(\mathrm{Sp}(A)\cap\mathbb{D}^c_l).$$

Since the set $\mathrm{Sp}(A)\cap\mathbb{D}$ is countable, its Hausdorff dimension is zero. Using sub-additivity of Hausdorff dimension and Theorem 8.3.1,

$$\begin{aligned}\dim_H(\mathrm{Sp}(A)) &\leq \dim_H(\mathrm{Sp}(A)\cap\mathbb{D}^c)\\ &\leq \dim_H(\mathrm{cl}(\mathrm{Sp}(A)\cap\mathbb{D}^c)) = \dim_{H'}(\mathrm{Sp}(A)\cap\mathbb{D}^c)\\ &\leq \dim_H(\mathrm{cl}(\mathrm{Sp}(A)\cap\mathbb{D}^c_l)) = \dim_{H'}(\mathrm{Sp}(A)\cap\mathbb{D}^c_l)\\ &\leq \dim_H(\mathrm{Sp}(A)).\end{aligned}$$

It follows that $h(A,d) = 0$ if $d > \dim_H(\mathrm{Sp}(A))$ and that $h(A,d) = \infty$ if $d < \dim_H(\mathrm{Sp}(A))$. Define

$$\Gamma_{n_3,n_2,n_1}(A) = \sup_{j=1,\dots,2^{n_3}} \left\{\frac{j}{2^{n_3}} : h_{n_3,n_2,n_1}(A,k/2^{n_3}) + \frac{1}{n_2} > \frac{1}{2} \text{ for } k = 1,\dots,j\right\},$$

where in this case we define the maximum over the empty set to be 0.

Consider $n_2 \geq n_3$. Since $h_{n_3,n_2,n_1}(A,d) \uparrow h_{n_3,n_2}(A,d)$, it is clear that

$$\lim_{n_1\to\infty}\Gamma_{n_3,n_2,n_1}(A) = \sup_{j=1,\dots,2^{n_3}} \left\{\frac{j}{2^{n_3}} : h_{n_3,n_2}(A,k/2^{n_3}) + \frac{1}{n_2} > \frac{1}{2} \text{ for } k = 1,\dots,j\right\} =: \Gamma_{n_3,n_2}(A).$$

If $h_{n_3}(A,d) \geq 1/2$, then $h_{n_3,n_2}(A,d) + 1/n_2 > 1/2$ for all n_2, otherwise we must have $h_{n_3,n_2}(A,d) + 1/n_2 < 1/2$ eventually. Hence

$$\lim_{n_2\to\infty}\Gamma_{n_3,n_2}(A) = \sup_{j=1,\dots,2^{n_3}} \left\{\frac{j}{2^{n_3}} : h_{n_3}(A,k/2^{n_3}) \geq \frac{1}{2} \text{ for } k = 1,\dots,j\right\} =: \Gamma_{n_3}(A).$$

Using the monotonicity of $h_{n_3}(A,d)$ in d and the proven properties of the limit function h, it follows that

$$\lim_{n_3\to\infty}\Gamma_{n_3}(A) = \dim_H(\mathrm{Sp}(A)).$$

The fact that h_{n_3} is non-decreasing in n_3, the set $\{1/2^{n_3}, 2/2^{n_3}, \dots, 1\}$ refines itself and the stated monotonicity show that convergence is monotonic from below and hence we get the Σ^A_3 classification.

Step 4: $\{\Xi_H, \Omega_{\mathrm{SA}}, \Lambda_1\} \in \Sigma^A_4$ and $\{\Xi_H, \Omega_{\mathrm{SA}}, \Lambda_2\} \in \Sigma^A_3$. The first of these can be proven as in step 3 by replacing (n_1, n_2, n_3) by (n_2, n_3, n_4) and the set $S_{n_2,n_1}(A)$ by the set $S_{n_3,n_2,n_1}(A)$ given by the union of all $S \in \rho_{n_3}$ with $S \subset [-n_2, n_2]$ and such that the Σ_2^A tower of algorithms discussed in Lemma 8.1.7 outputs yes for the interior of S and input parameters (n_2, n_1). To prove $\{\Xi_H, \Omega_{\mathrm{SA}}, \Lambda_2\} \in \Sigma_3^A$ we use exactly the same construction as in step 3 now using the Σ_1^A algorithm (which uses Λ_2) given by Lemma 8.1.7. □

8.4 Numerical Examples

We demonstrate that whilst some of the problems considered in this chapter require more than one limit to solve, the towers of algorithms constructed in this chapter are usable and can be efficiently implemented for large scale computations. Exactly the same comments can be made as in §7.6. The algorithms have desirable convergence properties, converging monotonically or being eventually constant, as captured by the Σ/Π classification. Generically, this monotonicity holds in all of the limits, and not just the final limit: many of the towers undergo *oscillation phenomena* where each subsequent limit is monotone but in the opposite sense/direction than the limit beforehand. We can take advantage of this when analysing the algorithms numerically, and this can be useful for creating ansatz for stopping criteria. The algorithms also highlight suitable information that lowers the SCI classification to Σ_1/Π_1. Other advantages for the algorithms based on approximating the resolvent norm include locality, numerical stability and speed/parallelisation.

8.4.1 Numerical examples for Lebesgue measure

Our first set of examples tests the towers of algorithms constructed for Lebesgue measure. We consider one example where the solution is analytically known and then one where nothing is currently known.

Almost Mathieu operator

We begin testing the algorithms on the almost Mathieu operator, which was studied in §6.4 of Chapter 6. For the benefit of the reader, we recall that the operator acts on $l^2(\mathbb{Z})$ via

$$(H_\alpha x)_n = x_{n-1} + x_{n+1} + 2\lambda \cos(2\pi n\alpha + \nu)x_n.$$

For irrational α, the spectrum of H_α does not depend on ν and [AK06]

$$\mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}(H_\alpha)) = 4\,|1 - |\lambda||\,. \tag{8.4.1}$$

We consider the case $\alpha = (\sqrt{5}-1)/2$ and without loss of generality set $\nu = 0$. Figure 8.1 shows the output of the algorithm, computing $\mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}(H_\alpha))$ and $\mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}_\epsilon(H_\alpha))$ for a range of values of ϵ. We chose values of $n = 5000$ (corresponding to 10003×10001 matrices for resolvent estimates), a grid spacing of $1/128$ and a resolution in `DistSpec` of order $1/1000$. One can clearly see that the estimates for $\mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}_\epsilon(H_\alpha))$ are decreasing to $\mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}(H_\alpha))$, which is well-estimated by `LebSpec` (Method 1).

We also compare Method 1 with the naive estimate provided by finite section estimates $\mathrm{Sp}(P_nH_\alpha P_n)$, where P_n is the orthogonal projection onto $\mathrm{span}\{e_k : |k| \leq n\}$. As expected, this gives too coarse an estimate of the Lebesgue measure, overestimating the true value, particularly when the Lebesgue measure is close to zero. `LebSpec` and `LebPseudoSpec` estimate the distance to the spectrum directly, allowing us to produce covering estimates that are tailor-made to the spectrum of the operator at hand. Other advantages include locality, numerical stability, speed/parallelisation, and guaranteed convergence.

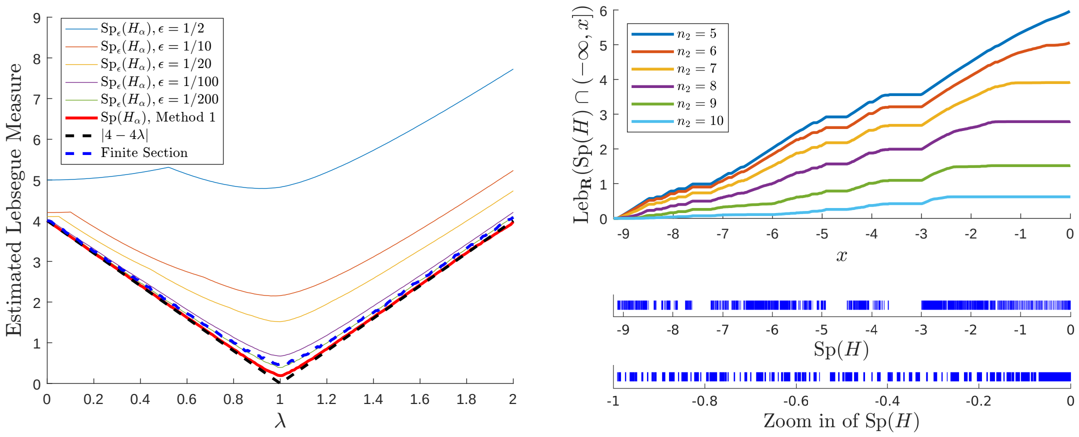


Figure 8.1: Left: Output of algorithm to compute $\mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}_\epsilon(H_\alpha))$ as well as the direct algorithm for $\mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}(H_\alpha))$ from §8.1.1 (Method 1). Note that we gain convergence to the true value as $\epsilon \downarrow 0$. Right: Estimates for $\mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}(H) \cap (-\infty, x])$ obtained by letting $n_1 = 10^5$ and selecting different n_2. The estimate above -3 appears to be well-resolved.

Graphical Laplacian on Penrose tile

We now consider the transport Hamiltonian H on a Penrose tile discussed in §3.4.1 of Chapter 3. An obvious problem of a height two tower Γ_{n_2,n_1} is that apriori we do not know, for a given input A, a choice of subsequence $n_2(n_1)$ such that $\Gamma_{n_2(n_1),n_1}(A)$ converges. There are numerous 'stopping criteria' for such scenarios (but, in general, the SCI classification shows that given such a criterion, there will always be an operator for which the subsequence choice fails). In our case, note that, for the height two tower in §8.1.1, we may assume without loss of generality that $\Gamma_{n_2,n_1}(A)$ is decreasing in n_2 but increasing in n_1. This suggests setting n_1 as computationally large as feasibly possible, then choosing a suitable cut-off, or maxima N, for n_2 and seeing if we appear to gain convergence for $n_2 \leq N$. We set $n_1 = 10^5$ and look at the average estimated error of the output. This was 0.0016 for a grid spacing of 10^{-5} so we shall consider grid refinements of spacing $1/32, 1/64, ..., 1/1024$ corresponding to $n_2 = 5, 6, ..., 10$. Figure 8.1 (right) shows the output as a cumulative Lebesgue measure, that is, an estimate of $\mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}(H) \cap (-\infty, x])$ for a given x, along with the computed spectrum (for a grid spacing of 10^{-5}). The figure suggests that we have not reached required convergence in n_1 to take n_2 any larger. However, there is strong evidence that the part of the spectrum closest to 0 is resolved by the algorithm and has Lebesgue measure zero. We shall see more evidence for this in §8.4.2.

8.4.2 Numerical examples for fractal dimensions

We begin with the box-counting dimension and denote by $\tilde{\Gamma}_n$ the Σ_1^A algorithm for the spectrum from Chapter 3. The caveat in the tower of algorithms used to compute the box-counting dimension is that convergence can, at best, only be expected to be logarithmic in the following sense. We expect that the error in approximating $\log(N_{1/2^{n_2}}(\mathrm{Sp}(A)))/\log(2^{n_2})$ (recall that $N_\delta(\mathrm{F})$ is the number of closed boxes of side length $\delta > 0$ required to cover F) via the first limit is roughly order $\mathcal{O}(1/n_2)$. This can only be reached in the worst case for $d_{\mathrm{H}}(\tilde{\Gamma}_{n_1}(A), \mathrm{Sp}(A)) = \mathcal{O}(1/2^{n_2})$ meaning that we have to resolve the spectrum to order $\exp(-1/\epsilon)$ to approximate the box-counting dimension to order ϵ. This is a problem shared by all

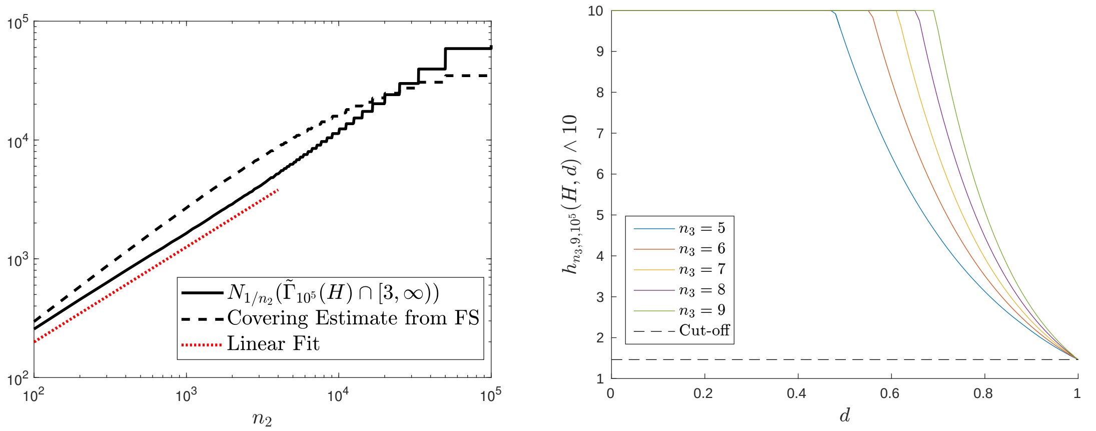


Figure 8.2: Left: A plot of $N_{1/n_2}(\tilde{\Gamma}_{10^5}(H) \cap [-3,\infty))$ against n_2. We found a scaling region with estimated box-counting dimension ≈ 0.8. Note that for large $n_2 \gtrsim 5000$, scalings are not resolved by $\tilde{\Gamma}_{10^5}$ (we can predict when this happens using the Σ_1^A property of $\tilde{\Gamma}_n$). We have also shown the approximation using finite sections (square $10^5 \times 10^5$ matrix truncations), as a dashed line, which overestimate the size of coverings, cannot detect the fractal structure, and break down for smaller n_2. Right: $h_{n_3,9,10^5}(H,d) \wedge 10$ to show a range of d where the estimates to the Hausdorff measures of $\mathrm{Sp}(H) \cap [-3,\infty)$ rapidly increase. These curves increase with n_3 consistent with the theory. This supports that the Hausdorff dimension may be close to 0.8. The 'cut-off' is a lower bound for the estimates given by $J/2^{n_2}$, with J being the number of intervals of length 2^{-n_2} that need to be covered from the estimate of $\mathrm{Sp}(H) \cap [-3,\infty)$.

methods that use the definition of box-counting dimension directly with an estimate of the spectrum. In reality, it is much better to assume that one has the stronger asymptotic condition $N_\delta \sim 1/\delta^d$, as $\delta \to 0$. We do this for the operator H from §8.4.1, for which the fractal dimension of $\mathrm{Sp}(H)$ is unknown.

In Figure 8.2, we plot $N_{1/n_2}(\tilde{\Gamma}_{10^5}(H) \cap [-3,\infty))$ against n_2. We also show a linear fit of slope 0.8. The error control provided by the algorithm $\tilde{\Gamma}_n$ allows us to deduce the region where the fit holds, corresponding to a reliable resolution of the spectrum. In other words, we can ensure that n_2 is not too large, so that the spacings of the coverings are not smaller than the numerically resolved spectrum. As expected, when n_2 is too large we see the effect of the grid spacing and the unresolved spectrum (by choosing larger n_1, we can take n_2 larger). The figure suggests that the spectrum above -3 is fractal with box-counting dimension ≈ 0.8 and hence has Lebesgue measure zero, in agreement with the findings in Figure 8.1.

Figure 8.2 shows what happens when one performs the same experiment but with finite section replacing $\tilde{\Gamma}_n$. First, for small n_2, using finite section produces an overestimate of the size of the covering and the corresponding slope of the graph due to spectral pollution. In other words, finite section prevents us from detecting the fractal spectrum. Second, the covering estimate via finite section breaks down at smaller n_2 and it is impossible to predict suitable values of n_2 so that the spacings of the coverings do not go beyond the resolution of the computed spectrum. Together, these issues highlight why finite section is unsuitable in general for approximating fractal dimensions and why the new algorithms are needed.

Finally, we investigate the Hausdorff dimension. An efficient way to compute a minimal covering is to use binary trees. We take $n_1 = 10^5$ and use the error bounds to estimate the resolution obtained which corresponds to n_2. The height three tower can be written as

$$\Gamma_{n_3,n_2,n_1}(A) = \sup_{j=1,\dots,2^{n_3}} \left\{ \frac{j}{2^{n_3}} : h_{n_3,n_2,n_1}(A, k/2^{n_3}) + \frac{1}{n_2} > \frac{1}{2} \text{ for } k = 1, \dots, j \right\},$$

where h_{n_3,n_2,n_1} is an analogue of $\mathcal{H}_\delta^d$ (see §8.3). Figure 8.2 (right) shows $h_{n_3,9,10^5}(H,d)$ for various d and restricted to estimating $\mathrm{Sp}(H) \cap [-3,\infty)$. The figure is consistent with the estimates increasing in n_3. There appears to be a region around 0.8 where the estimates begin to rapidly increase. Both algorithms support the possibility that the spectrum above -3 is fractal and hence has Lebesgue measure zero.

Chapter 9

Data-driven Computations of Spectral Properties of Koopman Operators

In this chapter, based on [CT21, CAS22], we consider spectral problems that arise in data-driven study of dynamical systems. We consider autonomous dynamical systems whose state $\boldsymbol{x}$ evolves over a state-space $\Omega \subseteq \mathbb{R}^d$ in discrete time-steps according to a function $F : \Omega \to \Omega$. In other words,

$$\boldsymbol{x}_{n+1} = F(\boldsymbol{x}_n), \qquad n \geq 0, \tag{9.0.1}$$

where $\boldsymbol{x}_0$ is a given initial condition. Such a dynamical system forms a trajectory of iterates $\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ in Ω. We are interested in answering questions about the system's behavior by analyzing such trajectories. The interaction between numerical analysis and dynamical systems theory has stimulated remarkable growth in the subject since the 1960s [Kal63, Lor63, Eps69, SH98]. With the arrival of big data [HTT$^+$09], modern statistical learning [HTF09], and machine learning [MRT18], data-driven algorithms are now becoming increasingly important in understanding dynamical systems [SL09, BK19].

9.1 Koopman Operators and associated Challenges

A classical viewpoint to analyze dynamical systems that originates in the seminal work of Poincaré [Poi99] is to study fixed points and periodic orbits, as well as stable and unstable manifolds. Two fundamental challenges with Poincaré's geometric state-space viewpoint are:

- **Non-linear dynamics:** To understand the stability of fixed points of non-linear dynamical systems, one typically forms local models centered at these fixed points. Such models allow one to predict long-time dynamics in small neighbourhoods of fixed points and attracting manifolds. However, they do not provide reasonable predictions for all initial conditions. A global understanding of non-linear dynamics in state-space remains largely qualitative [BMM12].

- **Unknown dynamics:** For many applications, a system's dynamics may be too complicated to describe analytically, or we may have incomplete knowledge of its evolution. Typically, we can only acquire several sequences of iterates of (9.0.1) starting at different values of $\boldsymbol{x}_0$. This means that constructing local models can be impossible. In this chapter, we focus on data-driven approaches to learning and analyzing the dynamical system with trajectories of iterates from (9.0.1).

Koopman operator theory, which dates back to Koopman and von Neumann [KvN32, Koo31], is an alternative viewpoint from which to analyse a dynamical system, that uses the space of scalar observable functions [Mez21]. Its increasing popularity has led to the term "Koopmanism" [BMM12], as well as thousands of articles over the last decade. A reason for the recent attention is its use in data-driven methods for studying dynamical systems (see [BBKK21] for a review and the history). Some popular applications include fluid dynamics, epidemiology, neuroscience, finance, robotics, power grids, and molecular dynamics.

Let $g : \Omega \to \mathbb{C}$ be a function that one can use to indirectly measure the state of the dynamical system in (9.0.1). Such a function g is known as an observable. One typically works in the Hilbert space $L^2(\Omega, \omega)$ of observables for a positive measure ω on Ω.[1] We consider the Koopman operator $\mathcal{K} : \mathcal{D}(\mathcal{K}) \to L^2(\Omega, \omega)$, where $\mathcal{D}(\mathcal{K}) \subseteq L^2(\Omega, \omega)$, given by

$$[\mathcal{K}g](\boldsymbol{x}) = (g \circ F)(\boldsymbol{x}), \qquad \boldsymbol{x} \in \Omega, \qquad g \in \mathcal{D}(\mathcal{K}), \tag{9.1.1}$$

where the equality is understood in the $L^2(\Omega, \omega)$ sense. $\mathcal{K}$ is a linear operator, regardless of whether the dynamics are linear or non-linear. Hence, the behaviour of the dynamical system (9.0.1) is determined by the spectral information of $\mathcal{K}$ (e.g., see (9.3.4)). However, since $\mathcal{K}$ is an infinite-dimensional operator, its spectral information can be far more complicated than that of a finite matrix. For example, $\mathcal{K}$ can have both discrete[2] and continuous spectra.

Computing the spectral properties of $\mathcal{K}$ is an active area of research - see [CT21] for further discussion. However, remaining challenges (some of which we have met in previous chapters) include:

- **Continuous spectra.**
- **Spectral pollution.**
- **Lack of (non-trivial) finite-dimensional invariant subspaces.**
- **Strong non-linearities and high-dimensional state-space.**

The goal of this chapter is to show how these challenges can be overcome. We have not framed theorems in terms of the SCI hierarchy. However, the reader will be able to see its presence in some of the theorems. Currently, it is an open problem to prove *lower bounds* for the classification of computational problems associated with Koopman operators.

We assume that we have access to discrete time snapshots of this system, i.e., a finite set of M pairs of measurements

$$\{\boldsymbol{x}^{(j)}, \boldsymbol{y}^{(j)}\}_{j=1}^{M} \quad \text{such that} \quad \boldsymbol{y}^{(j)} = F(\boldsymbol{x}^{(j)}), \quad j = 1, \ldots, M, \tag{9.1.2}$$

where the operator F evolves the system along one discrete time unit. For example, these snapshots could be measurements of unsteady velocities across M discrete spatial grid points taken via Particle Image Velocimetry (PIV). Suitable data could be collected from one long time trajectory, corresponding to $\boldsymbol{x}^{(j)} = F^{j-1}(\boldsymbol{x}_0)$, or from multiple shorter trajectories.

[1] We do not assume that this measure is invariant, and the most common choice of ω is the standard Lebesgue measure. This choice is natural for Hamiltonian systems for which the Koopman operator is unitary on $L^2(\Omega, \omega)$. For other systems, we can select ω according to the region where we wish to study the dynamics, such as a Gaussian measure.

[2] Throughout this chapter we use the term "discrete spectra" to mean the eigenvalues of $\mathcal{K}$, also known as the point spectrum. This also includes embedded eigenvalues, in contrast to the usual definition of the discrete spectrum.

9.2 Residual Dynamic Mode Decomposition (ResDMD)

We develop an algorithm, Residual DMD (ResDMD), that approximates the associated Koopman operator of the dynamics. Our approach allows for Koopman operators $\mathcal{K}$ that have no finite-dimensional invariant subspace. The key difference between ResDMD and other DMD algorithms (such as EDMD) is that we construct Galerkin approximations for not only $\mathcal{K}$, but also $\mathcal{K}^*\mathcal{K}$. This difference allows us to have rigorous convergence guarantees for ResDMD when recovering the spectral information of $\mathcal{K}$ and computing spectra and pseudospectra. In particular, we avoid spectral pollution.

9.2.1 Extended DMD (EDMD) and a new matrix for computing residuals

Before discussing our ResDMD approach, we describe EDMD. EDMD constructs a matrix $K_{\mathrm{EDMD}} \in \mathbb{C}^{N_K \times N_K}$ that approximates the action of $\mathcal{K}$ from the snapshot data. The original version of EDMD assumes that $\{\boldsymbol{x}^{(j)}\}_{j=1}^M \subset \Omega$ are drawn independently according to ω [WKR15]. Here, we describe EDMD for arbitrary initial states and use $\{\boldsymbol{x}^{(j)}\}_{j=1}^M$ as quadrature nodes.

Given a dictionary $\{\psi_1, \dots, \psi_{N_K}\} \subset \mathcal{D}(\mathcal{K})$ of observables, EDMD selects a matrix K_{EDMD} that approximates $\mathcal{K}$ on the subspace $V_{N_K} = \mathrm{span}\{\psi_1, \dots, \psi_{N_K}\}$, i.e.,

$$[\mathcal{K}\psi_j](\boldsymbol{x}) = \psi_j(F(\boldsymbol{x})) \approx \sum_{i=1}^{N_K} (K_{\mathrm{EDMD}})_{ij} \psi_i(\boldsymbol{x})$$

for $1 \leq j \leq N_K$. Define the vector-valued feature map $\Psi(\boldsymbol{x}) = \begin{bmatrix} \psi_1(\boldsymbol{x}) & \cdots & \psi_{N_K}(\boldsymbol{x}) \end{bmatrix} \in \mathbb{C}^{1 \times N_K}$. Then any $g \in V_{N_K}$ can be written as $g(\boldsymbol{x}) = \sum_{j=1}^{N_K} \psi_j(\boldsymbol{x}) g_j = \Psi(\boldsymbol{x})\, \boldsymbol{g}$ for some vector $\boldsymbol{g} \in \mathbb{C}^{N_K}$. It follows that

$$[\mathcal{K}g](\boldsymbol{x}) = \Psi(F(\boldsymbol{x}))\, \boldsymbol{g} = \Psi(\boldsymbol{x})(K_{\mathrm{EDMD}}\, \boldsymbol{g}) + \underbrace{\left(\sum_{j=1}^{N_K} \psi_j(F(\boldsymbol{x})) g_j - \Psi(\boldsymbol{x})(K_{\mathrm{EDMD}}\, \boldsymbol{g}) \right)}_{r(\boldsymbol{g}, \boldsymbol{x})}.$$

Typically, V_{N_K} is not an invariant subspace of $\mathcal{K}$ so there is no choice of K_{EDMD} that makes $r(\boldsymbol{g}, \boldsymbol{x})$ zero for all $g \in V_N$ and $\boldsymbol{x} \in \Omega$. Instead, it is natural to select K_{EDMD} as a solution of

$$\mathrm{argmin}_{B \in \mathbb{C}^{N_K \times N_K}} \left\{ \int_\Omega \max_{\boldsymbol{g} \in \mathbb{C}^{N_K}, \|\boldsymbol{g}\|=1} |r(\boldsymbol{g}, \boldsymbol{x})|^2 \, d\omega(\boldsymbol{x}) = \int_\Omega \|\Psi(F(\boldsymbol{x})) - \Psi(\boldsymbol{x})B\|_{\ell^2}^2 \, d\omega(\boldsymbol{x}) \right\}. \quad (9.2.1)$$

In practice, one cannot directly evaluate the integral in (9.2.1). Instead, we approximate it via a quadrature rule with nodes $\{\boldsymbol{x}^{(j)}\}_{j=1}^M$ and weights $\{w_j\}_{j=1}^M$. The discretised version of (9.2.1) is therefore the following weighted least-squares problem:

$$\mathrm{argmin}_{B \in \mathbb{C}^{N_K \times N_K}} \sum_{j=1}^M w_j \left\| \Psi(\boldsymbol{y}^{(j)}) - \Psi(\boldsymbol{x}^{(j)})B \right\|_{\ell^2}^2. \quad (9.2.2)$$

A solution to (9.2.2) can be written down explicitly as $K_{\mathrm{EDMD}} = (\Psi_X^* W \Psi_X)^\dagger (\Psi_X^* W \Psi_Y)$, where '†' denotes the pseudoinverse and $W = \mathrm{diag}(w_1, \dots, w_M)$. Here, Ψ_X and Ψ_Y are the $M \times N_K$ matrices

$$\Psi_X = \begin{bmatrix} \Psi(\boldsymbol{x}^{(1)})^\top & \cdots & \Psi(\boldsymbol{x}^{(M)})^\top \end{bmatrix}^\top, \quad \Psi_Y = \begin{bmatrix} \Psi(\boldsymbol{y}^{(1)})^\top & \cdots & \Psi(\boldsymbol{y}^{(M)})^\top \end{bmatrix}^\top. \quad (9.2.3)$$

By reducing the size of the dictionary if necessary, we may assume without loss of generality that $\Psi_X^* W \Psi_X$ is invertible. In practice, one may also consider regularisation through truncated singular value decompositions. Since $\Psi_X^* W \Psi_X = \sum_{j=1}^M w_j \Psi(\boldsymbol{x}^{(j)})^* \Psi(\boldsymbol{x}^{(j)})$ and $\Psi_X^* W \Psi_Y = \sum_{j=1}^M w_j \Psi(\boldsymbol{x}^{(j)})^* \Psi(\boldsymbol{y}^{(j)})$, if the quadrature converges then

$$\lim_{M\to\infty}[\Psi_X^* W\Psi_X]_{jk} = \langle\psi_k,\psi_j\rangle \quad \text{and} \quad \lim_{M\to\infty}[\Psi_X^* W\Psi_Y]_{jk} = \langle\mathcal{K}\psi_k,\psi_j\rangle,$$

where $\langle\cdot,\cdot\rangle$ is the inner product associated with $L^2(\Omega,\omega)$. Thus, EDMD can be viewed as a Galerkin method in the large data limit as $M\to\infty$. Let $P_{V_{N_K}}$ denote the orthogonal projection onto V_{N_K}. In the large data limit, K_{EDMD} approaches a matrix representation of $P_{V_{N_K}}\mathcal{K}P_{V_{N_K}}$ and the EDMD eigenvalues approach the spectrum of $P_{V_{N_K}}\mathcal{K}P_{V_{N_K}}$. Thus, approximating $\mathrm{Sp}(\mathcal{K})$ by the eigenvalues of K_{EDMD} is closely related to the finite section method. Since the finite section method can suffer from spectral pollution, spectral pollution is also a concern for EDMD and it is important to have an independent way to measure the accuracy of a candidate eigenvalue-eigenvector pair.

9.2.2 Measuring the accuracy of candidate eigenvalue-eigenvector pairs

Suppose that we have a candidate eigenvalue-eigenvector pair (λ,g) of $\mathcal{K}$, where $\lambda\in\mathbb{C}$ and $g=\Psi\boldsymbol{g}\in V_{N_K}$. One way to measure the accuracy of (λ,g) is by estimating the squared relative residual

$$\begin{aligned}\frac{\int_\Omega |[\mathcal{K}g](\boldsymbol{x})-\lambda g(\boldsymbol{x})|^2\, d\omega(\boldsymbol{x})}{\int_\Omega |g(\boldsymbol{x})|^2\, d\omega(\boldsymbol{x})} &= \frac{\langle(\mathcal{K}-\lambda)g,(\mathcal{K}-\lambda)g\rangle}{\langle g,g\rangle} \qquad (9.2.4)\\ &= \frac{\sum_{j,k=1}^{N_K}\overline{g_j}g_k\left[\langle\mathcal{K}\psi_k,\mathcal{K}\psi_j\rangle-\lambda\langle\psi_k,\mathcal{K}\psi_j\rangle-\overline{\lambda}\langle\mathcal{K}\psi_k,\psi_j\rangle+|\lambda|^2\langle\psi_k,\psi_j\rangle\right]}{\sum_{j,k=1}^{N_K}\overline{g_j}g_k\langle\psi_k,\psi_j\rangle}.\end{aligned}$$

If $\mathcal{K}$ is a normal operator, then the minimum of (9.2.4) over all normalised $g\in\mathcal{D}(\mathcal{K})$ is exactly the square distance of λ to the spectrum of $\mathcal{K}$; otherwise, for non-normal $\mathcal{K}$ the residual can still provide a measure of accuracy. One can also use the residual to bound the distance between g and the eigenspace associated with λ, assuming λ is a point in the discrete spectrum of $\mathcal{K}$.

We approximate the residual in (9.2.4) by

$$\mathrm{res}(\lambda,g)^2 = \frac{\sum_{j,k=1}^{N_K}\overline{g_j}g_k\left[(\Psi_Y^*W\Psi_Y)_{jk}-\lambda(\Psi_Y^*W\Psi_X)_{jk}-\overline{\lambda}(\Psi_X^*W\Psi_Y)_{jk}+|\lambda|^2(\Psi_X^*W\Psi_X)_{jk}\right]}{\sum_{j,k=1}^{N_K}\overline{g_j}g_k(\Psi_X^*W\Psi_X)_{jk}}. \qquad (9.2.5)$$

All the terms in this residual can be computed using the snapshot data. Note that, as well as the matrices found in EDMD, (9.2.5) has the *additional matrix* $\Psi_Y^*W\Psi_Y$. Moreover, under certain conditions, we have $\lim_{M\to\infty}\mathrm{res}(\lambda,g)^2=\int_\Omega |[\mathcal{K}g](\boldsymbol{x})-\lambda g(\boldsymbol{x})|^2\, d\omega(\boldsymbol{x})/\int_\Omega |g(\boldsymbol{x})|^2\, d\omega(\boldsymbol{x})$ for any $g\in V_{N_K}$. In particular, we often have $\lim_{M\to\infty}[\Psi_Y^*W\Psi_Y]_{jk}=\langle\mathcal{K}\psi_k,\mathcal{K}\psi_j\rangle$ and then $\Psi_Y^*W\Psi_Y$ formally corresponds to a Galerkin approximation of $\mathcal{K}^*\mathcal{K}$ as $M\to\infty$.

9.2.3 Convergence of quadrature

There are typically three scenarios that ensure that

$$\begin{aligned}\lim_{M\to\infty}[\Psi_X^*W\Psi_X]_{jk} &= \langle\psi_k,\psi_j\rangle,\\ \lim_{M\to\infty}[\Psi_X^*W\Psi_Y]_{jk} &= \langle\mathcal{K}\psi_k,\psi_j\rangle, \qquad (9.2.6)\\ \lim_{M\to\infty}[\Psi_Y^*W\Psi_Y]_{jk} &= \langle\mathcal{K}\psi_k,\mathcal{K}\psi_j\rangle.\end{aligned}$$

- Suppose that $\{\boldsymbol{x}^{(j)}\}_{j=1}^M$ are selected so that they are an M-point quadrature rule with weights $\{w_j\}_{j=1}^M$. Integrals and inner products can then be approximated with numerical integration by evaluating functions at the data points. High-order quadrature rules can lead to fast rates of convergence if the

Algorithm 1 : ResDMD for computing eigenpairs without spectral pollution.

Input: Snapshot data $\{\boldsymbol{x}^{(j)},\boldsymbol{y}^{(j)}\}_{j=1}^{M}$ (such that $\boldsymbol{y}^{(j)} = F(\boldsymbol{x}^{(j)})$), quadrature weights $\{w_j\}_{j=1}^{M}$, a dictionary of observables $\{\psi_j\}_{j=1}^{N_K}$ and an accuracy goal $\epsilon > 0$.

```
1: Compute $\Psi_X^* W \Psi_X$, $\Psi_X^* W \Psi_Y$, and $\Psi_Y^* W \Psi_Y$.
2: Solve $(\Psi_X^* W \Psi_Y)\boldsymbol{g} = \lambda(\Psi_X^* W \Psi_X)\boldsymbol{g}$ for eigenpairs $\{(\lambda_j, g_{(j)} = \Psi \boldsymbol{g}_j)\}$.
3: Compute $\mathrm{res}(\lambda_j, g_{(j)})$ for all $j$ (see (9.2.5)) and discard if $\mathrm{res}(\lambda_j, g_{(j)}) > \epsilon$.
```

Output: A collection of accurate eigenpairs $\{(\lambda_j, \boldsymbol{g}_j) : \mathrm{res}(\lambda_j, g_{(j)}) \leq \epsilon\}$.

dictionary functions and F are sufficiently regular. If the integrands are analytic in a neighbourhood of Ω, then we can often select a quadrature rule that even converges exponentially as $M \to \infty$. For example, if Ω is unbounded then we can use quadrature rules such as the trapezoidal rule and if Ω is a bounded simple domain then one can use Gaussian quadrature. When the state-space dimension d is moderately large we can use sparse grids and a kernelized approach for large d.

- If ω is a probability measure and the initial points $\{\boldsymbol{x}^{(j)}\}_{j=1}^{M}$ are drawn independently and at random according to ω, the strong law of large numbers shows that $\lim_{M\to\infty}[\Psi_X^* W \Psi_X]_{jk} = \langle \psi_k, \psi_j \rangle$ and $\lim_{M\to\infty}[\Psi_X^* W \Psi_Y]_{jk} = \langle \mathcal{K}\psi_k, \psi_j \rangle$ holds with probability one [KKS16, Section 3.4] provided that ω is not supported on a zero level set that is a linear combination of the dictionary [KM18, Section 4]. This is with the quadrature weights $w_j = 1/M$ and the convergence is typically at a Monte Carlo rate of $\mathcal{O}(M^{-1/2})$ [Caf98]. It is a practical approach if the state-space dimension is large. One could also consider quasi-Monte Carlo integration, which can achieve a faster rate of $\mathcal{O}(M^{-1})$ (up to logarithmic factors) under suitable conditions [Caf98]. This argument is straightforward to adapt to show the convergence $\lim_{M\to\infty}[\Psi_Y^* W \Psi_Y]_{jk} = \langle \mathcal{K}\psi_k, \mathcal{K}\psi_j \rangle$.

- For a single long trajectory, if the dynamical system is ergodic, then one can use Birkhoff's Ergodic Theorem to show that $\lim_{M\to\infty}[\Psi_X^* W \Psi_X]_{jk} = \langle \psi_k, \psi_j \rangle$ and $\lim_{M\to\infty}[\Psi_X^* W \Psi_Y]_{jk} = \langle \mathcal{K}\psi_k, \psi_j \rangle$ [KM18]. One chooses $w_j = 1/M$ but the convergence rate is problem dependent [Kac96]. This argument is straightforward to adapt to show $\lim_{M\to\infty}[\Psi_Y^* W \Psi_Y]_{jk} = \langle \mathcal{K}\psi_k, \mathcal{K}\psi_j \rangle$.

The scenario depends on the type of data that is collected. Typically for experiments, it is the later two that are most relevant. However, if one is entirely free to select the initial conditions of the trajectory data, and d is not too large, then we recommend picking them based on a high-order quadrature rule.

9.2.4 Convergence theorems

We now present our first ResDMD algorithm that computes the residual using the snapshot data to avoid spectral pollution. As is usually done, the algorithm assumes that K_{EDMD} is diagonalisable. First, we compute the three matrices $\Psi_X^* W \Psi_X$, $\Psi_X^* W \Psi_Y$, and $\Psi_Y^* W \Psi_Y$. Then, we find the eigenvalues and eigenvectors of K_{EDMD}, i.e., we solve $(\Psi_X^* W \Psi_X)^{\dagger}(\Psi_X^* W \Psi_Y)\boldsymbol{g} = \lambda \boldsymbol{g}$. One can solve this eigenproblem directly, but it is often numerically more stable to solve the generalised eigenproblem $(\Psi_X^* W \Psi_Y)\boldsymbol{g} = \lambda(\Psi_X^* W \Psi_X)\boldsymbol{g}$. Afterward, to avoid spectral pollution, we discard computed eigenpairs with a larger relative residual than an accuracy goal of $\epsilon > 0$.

Algorithm 1 summarises the procedure and is a simple modification of EDMD, as the only difference is a clean-up where spurious eigenpairs are discarded based on their residual. This clean-up avoids spectral

Algorithm 2 : ResDMD for estimating $\mathrm{Sp}_\epsilon(\mathcal{K})$.

Input: Snapshot data $\{\boldsymbol{x}^{(j)},\boldsymbol{y}^{(j)}\}_{j=1}^M$ (such that $\boldsymbol{y}^{(j)}=F(\boldsymbol{x}^{(j)})$), quadrature weights $\{w_j\}_{j=1}^M$, a dictionary of observables $\{\psi_j\}_{j=1}^{N_K}$, an accuracy goal $\epsilon>0$, and a grid $z_1,\dots,z_k\in\mathbb{C}$.

```
1: Compute $\Psi_X^* W\Psi_X$, $\Psi_X^* W\Psi_Y$, and $\Psi_Y^* W\Psi_Y$.
2: For each $z_j$, compute $\tau_j=\min_{\boldsymbol{g}\in\mathbb{C}^{N_K}}\mathrm{res}(z_j,\Psi\boldsymbol{g})$ (see (9.2.5)), which is a generalised SVD problem,
   and the corresponding singular vectors $\boldsymbol{g}_j$.
```

Output: Estimate of $\mathrm{Sp}_\epsilon(\mathcal{K})$: $\{z_j:\tau_j<\epsilon\}$, and approximate eigenfunctions: $\{\boldsymbol{g}_j:\tau_j<\epsilon\}$.

pollution and also removes eigenpairs that are inaccurate because of numerical errors associated with non-normal operators, up to the relative tolerance ϵ. The following result makes this precise.

Theorem 9.2.1. *Let $\mathcal{K}$ be the associated Koopman operator of* (9.0.1) *from which snapshot data is collected. Let Λ_M denote the eigenvalues in the output of Algorithm 1. Then, assuming* (9.2.6)*,*

$$\limsup_{M\to\infty}\max_{\lambda\in\Lambda_M}\|(\mathcal{K}-\lambda)^{-1}\|^{-1}\leq\epsilon.$$

Exercise: Prove Theorem 9.2.1.

Theorem 9.2.1 tells us that, in the large data limit, ResDMD computes eigenvalues inside $\mathrm{Sp}_\epsilon(\mathcal{K})$ and hence, avoids spectral pollution and returns reasonable eigenvalues. Despite this, Algorithm 1 may not approximate the whole $\mathrm{Sp}_\epsilon(\mathcal{K})$, even as $M\to\infty$ and $N_K\to\infty$. This is because the eigenvalues of K_{EDMD} may not approximate the whole spectrum of $\mathcal{K}$. For example, consider the shift operator, which is unitary. [DRAW PICTURE ON BOARD] Suppose our dictionary consists of the functions $\psi_j(k)=\delta_{k,q(j)}$, where $q:\mathbb{N}\to\mathbb{Z}$ is an enumeration of $\mathbb{Z}$. Then, in the large data limit, K_{EDMD} corresponds to a finite section of the shift operator and has spectrum $\{0\}$, whereas $\mathrm{Sp}(\mathcal{K})=\mathbb{T}$. Hence, for $\epsilon<1$, the output of Algorithm 1 is the empty set. This issue is known as *spectral inclusion*.

To overcome this issue, [CT21] developed ways to compute spectra and pseudospectra. For example, Algorithm 2 computes practical approximations of $\mathrm{Sp}_\epsilon(\mathcal{K})$ with rigorous convergence guarantees. Assuming (9.2.6), the output is guaranteed to be inside the $\mathrm{Sp}_\epsilon(\mathcal{K})$. Algorithm 2 also computes observables g with $\mathrm{res}(\lambda,g)<\epsilon$, which are known as ϵ-approximate eigenfunctions.

Exercise: Using Algorithm 2, develop an algorithm that converges to the so-called approximate point pseudospectrum,

$$\mathrm{Sp}_{\epsilon,\mathrm{ap}}(\mathcal{K}):=\mathrm{cl}\left(\{\lambda\in\mathbb{C}:\sigma_{\inf}(\mathcal{K}-\lambda)<\epsilon\}\right),$$

as $N_K\to\infty$.

9.2.5 Dealing with large state-space dimension

When d is large, it can be impractical to store or form the matrix K_{EDMD}, since the initial value of N_K is very large. We consider two common methods to overcome this issue:

(i) **DMD:** In this case, the dictionary consists of all monomials over Ω with $\psi_j(\boldsymbol{x})=e_j^*\boldsymbol{x}$. It is standard to form a low-rank approximation of $\sqrt{W}\Psi_X$ via a truncated singular value decomposition (SVD) as

$$\sqrt{W}\Psi_X\approx U_r\Sigma_rV_r^*. \tag{9.2.7}$$

Here, $\Sigma_r \in \mathbb{C}^{r\times r}$ is diagonal with strictly positive diagonal entries, and $V_r \in \mathbb{C}^{N_K\times r}$ and $U_r \in \mathbb{C}^{M\times r}$ have $V_r^* V_r = U_r^* U_r = I_r$. We then form the matrix

$$\tilde{K}_{\text{EDMD}} = (\sqrt{W}\Psi_X V_r)^\dagger \sqrt{W}\Psi_Y V_r = \Sigma_r^{-1} U_r^* \sqrt{W}\Psi_Y V_r = V_r^* K_{\text{EDMD}} V_r \in \mathbb{C}^{r\times r}. \tag{9.2.8}$$

Note that to fit into our Galerkin framework, this matrix is the transpose of the DMD matrix that is commonly computed in the literature.

(ii) **Kernelized EDMD (kEDMD):** kEDMD [WRK15] aims to make EDMD practical for large d. Supposing that Ψ_X is of full rank, kEDMD constructs a matrix with an identical formula to (9.2.8) with $r = M$, for which we have the equivalent form

$$\tilde{K}_{\text{EDMD}} = (\Sigma_M^\dagger U_M^*)(\sqrt{W}\Psi_Y\Psi_X^*\sqrt{W})(U_M\Sigma_M^\dagger). \tag{9.2.9}$$

Suitable matrices U_M and Σ_M can be recovered from the eigenvalue decomposition

$$\sqrt{W}\Psi_X\Psi_X^*\sqrt{W} = U_M\Sigma_M^2 U_M^*.$$

Moreover, both matrices $\sqrt{W}\Psi_X\Psi_X^*\sqrt{W}$ and $\sqrt{W}\Psi_Y\Psi_X^*\sqrt{W}$ can be computed using inner products. kEDMD applies the kernel trick to compute the inner products in an implicitly defined reproducing Hilbert space $\mathcal{H}$ with inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ [Sch01]. A positive-definite kernel function $\mathcal{S}:\Omega\times\Omega\to\mathbb{R}$ induces a feature map $\varphi:\mathbb{R}^d\to\mathcal{H}$ so that $\langle\varphi(\boldsymbol{x}),\varphi(\boldsymbol{y})\rangle_{\mathcal{H}} = \mathcal{S}(\boldsymbol{x},\boldsymbol{y})$. This leads to a choice of (typically non-linear) dictionary $\Psi(\boldsymbol{x})$ so that $\Psi(\boldsymbol{x})\Psi(\boldsymbol{y})^* = \langle\varphi(\boldsymbol{x}),\varphi(\boldsymbol{y})\rangle_{\mathcal{H}} = \mathcal{S}(\boldsymbol{x},\boldsymbol{y})$. Often $\mathcal{S}$ can be evaluated in $\mathcal{O}(d)$ operations, meaning that $\tilde{K}_{\text{EDMD}}$ is constructed in $\mathcal{O}(dM^2)$ operations.

In either of these two cases, the approximation of $\mathcal{K}$ is equivalent to using a new dictionary with feature map $\Psi(\boldsymbol{x})V_r \in \mathbb{C}^{1\times r}$. In the case of DMD, it is beneficial to use the mathematically equivalent choice $\Psi(\boldsymbol{x})V_r\Sigma_r^{-1}$, which is numerically better conditioned. To see why, note that $\sqrt{W}\Psi_X V_r\Sigma_r^{-1} \approx U_r$ and U_r has orthonormal columns.

The problem of vanishing residual estimates

Proposition 9.2.2. *Suppose that $\sqrt{W}\Psi_X V_r$ has full row rank, so that $r = M$, and that $\boldsymbol{v}\in\mathbb{C}^M$ is an eigenvector of $\tilde{K}_{\text{EDMD}}$ with eigenvalue λ. Then* $\text{res}(\lambda, g) = 0$.

Exercise: Prove Proposition 9.2.2.

Similarly, if r is too large, $\text{res}(\lambda, g)$ will be a bad approximation of the true residual. In other words, the regime $r\sim M$ prevents the large data convergence ($M\to\infty$) of the quadrature rule, which holds for a *fixed basis* and hence a fixed basis size. In turn, this prevents us from being able to apply the results of Section 9.2.4. We next discuss how to overcome this issue by using two sets of snapshot data; these could arise from two independent tests of the same system, or by partitioning the measured data into two groups.

Using two subsets of the snapshot data

A simple remedy to avoid the problem in Section 9.2.5 is to consider *two sets of snapshot data*. We consider an initial set $\{\tilde{\boldsymbol{x}}^{(j)},\tilde{\boldsymbol{y}}^{(j)}\}_{j=1}^{M'}$, which we use to form our dictionary. We then apply ResDMD to the

Algorithm 3 : **ResDMD with DMD selected observables.**

```
Input: Snapshot data $\{\tilde{\boldsymbol{x}}^{(j)},\tilde{\boldsymbol{y}}^{(j)}\}_{j=1}^{M'}$ and $\{\hat{\boldsymbol{x}}^{(j)},\hat{\boldsymbol{y}}^{(j)}\}_{j=1}^{M''}$, positive integer $N_K \leq M'$.
1: Set $\Psi_{\mathrm{DMD}}(\boldsymbol{x}) = \begin{bmatrix} e_1^*\boldsymbol{x} & \cdots & e_d^*\boldsymbol{x}\end{bmatrix}$.
2: Compute a truncated SVD
       $\frac{1}{\sqrt{M'}}\begin{pmatrix}\Psi_{\mathrm{DMD}}(\tilde{\boldsymbol{x}}^{(1)})^\top & \cdots & \Psi_{\mathrm{DMD}}(\tilde{\boldsymbol{x}}^{(M')})^\top\end{pmatrix}^\top \approx U_{N_K}\Sigma_N V_{N_K}^*.$
3: Apply Algorithms 1 and 2 with the matrices
       $\Psi_X = \begin{pmatrix}\Psi_{\mathrm{DMD}}(\hat{\boldsymbol{x}}^{(1)}) \\ \vdots \\ \Psi_{\mathrm{DMD}}(\hat{\boldsymbol{x}}^{(M'')})\end{pmatrix} V_{N_K}\Sigma_{N_K}^\dagger, \quad \Psi_Y = \begin{pmatrix}\Psi_{\mathrm{DMD}}(\hat{\boldsymbol{y}}^{(1)}) \\ \vdots \\ \Psi_{\mathrm{DMD}}(\hat{\boldsymbol{y}}^{(M'')})\end{pmatrix} V_{N_K}\Sigma_{N_K}^\dagger. \qquad (9.2.10)$
Output: Spectral properties of Koopman operator according to Algorithms 1 and 2.
```

Algorithm 4 : **ResDMD with kEDMD selected observables.**

```
Input: Snapshot data $\{\tilde{\boldsymbol{x}}^{(j)},\tilde{\boldsymbol{y}}^{(j)}\}_{j=1}^{M'}$ and $\{\hat{\boldsymbol{x}}^{(j)},\hat{\boldsymbol{y}}^{(j)}\}_{j=1}^{M''}$, positive-definite kernel function $\mathcal{S}:\Omega\times\Omega\rightarrow\mathbb{R}$, and positive integer $N_K \leq M'$.
1: Apply kEDMD to $\{\tilde{\boldsymbol{x}}^{(m)},\tilde{\boldsymbol{y}}^{(m)}\}_{m=1}^{M'}$ with kernel $\mathcal{S}$ to compute the matrices $\tilde{K}_{\mathrm{EDMD}}$, $U_{M'}$ and $\Sigma_{M'}$ using the kernel trick.
2: Compute the dominant $N_K$ eigenvalues of $\tilde{K}_{\mathrm{EDMD}}$ and stack the corresponding eigenvectors column-by-column into $Z \in \mathbb{C}^{M'\times N_K}$.
3: Apply a QR decomposition to orthogonalise $Z$ to $Q = \begin{bmatrix}Q_1 & \cdots & Q_{N_K}\end{bmatrix} \in \mathbb{C}^{M'\times N_K}$.
4: Apply Algorithms 1 and 2 with $\{\hat{\boldsymbol{x}}^{(m)},\hat{\boldsymbol{y}}^{(m)}\}_{m=1}^{M''}$ and the dictionary $\{\psi_j\}_{j=1}^{N_K}$, where
       $\psi_j(\boldsymbol{x}) = \begin{bmatrix}\mathcal{S}(\boldsymbol{x},\tilde{\boldsymbol{x}}^{(1)}) & \mathcal{S}(\boldsymbol{x},\tilde{\boldsymbol{x}}^{(2)}) & \cdots & \mathcal{S}(\boldsymbol{x},\tilde{\boldsymbol{x}}^{(M')})\end{bmatrix}(U_{M'}\Sigma_{M'}^\dagger)Q_j, \qquad 1 \leq j \leq N_K.$
Output: Spectral properties of Koopman operator according to Algorithms 1 and 2.
```

computed dictionary with a second set of snapshot data $\{\hat{\boldsymbol{x}}^{(j)},\hat{\boldsymbol{y}}^{(j)}\}_{j=1}^{M''}$, allowing us to prove convergence as $M'' \rightarrow \infty$.

Exactly how to acquire a second set of snapshot data depends on the problem and method of data collection. Given snapshot data with random and independent $\{\boldsymbol{x}^{(j)}\}$, one can simply split up the snapshot data into two parts. For initial conditions that are distributed according to a high-order quadrature rule, if one already has access to M' snapshots then one must typically go back to the original dynamical system and request M'' further snapshots. For ergodic sampling along a trajectory, we can let $\{\tilde{\boldsymbol{x}}^{(j)},\tilde{\boldsymbol{y}}^{(j)}\}_{j=1}^{M'}$ correspond to the initial $M'+1$ points of the trajectory ($\tilde{\boldsymbol{x}}^{(j)} = F^{j-1}(\boldsymbol{x}_0)$ for $j = 1,\ldots,M'$) and let $\{\hat{\boldsymbol{x}}^{(j)},\hat{\boldsymbol{y}}^{(j)}\}_{j=1}^{M''}$ correspond to the initial $M''+1$ points of the trajectory ($\hat{\boldsymbol{x}}^{(j)} = F^{j-1}(\boldsymbol{x}_0)$ for $j = 1,\ldots,M''$).

In the case of DMD, the two stage process is summarised in Algorithm 3. Often a suitable choice of N_K can be obtained by studying the decay of the singular values of the data matrix.

In the case of kEDMD, the two stage process is summarised in Algorithm 4. The choice of kernel $\mathcal{S}$ determines the dictionary and the best choice depends on the application. In the following experiments, we use the Laplacian kernel $\mathcal{S}(\boldsymbol{x},\boldsymbol{y}) = \exp\left(-\gamma\|\boldsymbol{x}-\boldsymbol{y}\|\right)$, where γ is the reciprocal of the average ℓ^2-norm of the snapshot data after it is shifted to have mean zero.

We can now apply the theory of Section 9.2.4 in the limit $M'' \rightarrow \infty$. It is well-known that the eigenvalues computed by DMD and kEDMD may suffer from spectral pollution. However, and crucially in our setting, we do not directly use these methods to compute spectral properties of $\mathcal{K}$. Instead, we are only using them to select a reasonable dictionary of size N_K, after which our rigorous ResDMD algorithms can be used. Moreover, we use $\{\hat{\boldsymbol{x}}^{(m)}, \hat{\boldsymbol{y}}^{(m)}\}_{m=1}^{M''}$ to check the quality of the constructed dictionary. By studying the residuals and using the error control in ResDMD, we can tell a-posteriori whether the dictionary is satisfactory and whether N_K is sufficiently large.

Finally, it is worth pointing out that the above choices of dictionaries are certainly not the only choices. ResDMD can be applied to any suitable choice. For example, one could use other data-driven methods such as diffusion kernels [GKKS18] or trained neural networks [LDBK17, MFF20].

9.3 Spectral Measures for measure-preserving Systems

In the following, we will use the setup outlined at the start of Chapter 4.

9.3.1 The setup and Koopman mode decompositions

Suppose that the associated dynamics is measure-preserving so that $\omega(E) = \omega(\{\boldsymbol{x} : F(\boldsymbol{x}) \in E\})$ for any Borel measurable subset $E \subset \Omega$. Equivalently, this means that the Koopman operator $\mathcal{K}$ associated with the dynamical system in (9.0.1) is an isometry, i.e., $\|\mathcal{K}g\| = \|g\|$ for all observables $g \in \mathcal{D}(\mathcal{K}) = L^2(\Omega, \omega)$. Dynamical systems such as Hamiltonian flows, geodesic flows on Riemannian manifolds, Bernoulli schemes in probability theory, and ergodic systems are all measuring-preserving. Moreover, many dynamical systems become measure-preserving in the long-run.

As discussed in Chapter 4, spectral measures provide a way of diagonalising normal operators, including self-adjoint and unitary operators, even in the presence of continuous spectra. Unfortunately, a Koopman operator that is an isometry does not necessarily commute with its adjoint. Therefore, we must consider its unitary extension before defining a spectral measure and Koopman mode decomposition.

Unitary extensions of isometries

Given a Koopman operator $\mathcal{K}$ of a measure-preserving dynamical system, we use the concept of unitary extension to formally construct a related normal operator $\mathcal{K}'$. That is, suppose that $\mathcal{K} : L^2(\Omega, \omega) \to L^2(\Omega, \omega)$ is an isometry, then there exists a unitary extension $\mathcal{K}'$ defined on an extended Hilbert space $\mathcal{H}'$ with $L^2(\Omega, \omega) \subset \mathcal{H}'$ [NFBK10, Proposition I.2.3].[3] Even though such an extension is not unique, it allows us to understand the spectral information of $\mathcal{K}$ by considering $\mathcal{K}'$, which is a normal operator. If F is invertible and measure-preserving, $\mathcal{K}$ is unitary and we can simply take $\mathcal{K}' = \mathcal{K}$ and $\mathcal{H}' = L^2(\Omega, \omega)$.

Spectral measures of a Koopman operator

Given an observable $g \in L^2(\Omega, \omega) \subset \mathcal{H}'$ of interest such that $\|g\| = 1$, the spectral measure of $\mathcal{K}'$ with respect to g is a scalar measure defined as $\mu_g(U) := \langle E^{\mathcal{K}'}(U)g, g\rangle$, where $U \subset \mathbb{T}$ is a Borel measurable set [RS80]. For plotting and visualisation, it is more convenient to equivalently consider the corresponding probability measures ν_g defined on the periodic interval $[-\pi, \pi]_{\text{per}}$ after a change of variables $\lambda = \exp(i\theta)$

[3] To see how to extend $\mathcal{K}$ to a unitary operator $\mathcal{K}'$, consider the Wold–von Neumann decomposition [NFBK10, Theorem I.1.1]. This decomposition states that $\mathcal{K}$ can be written as $\mathcal{K} = (\oplus_{\alpha \in I} S_\alpha) \oplus U$ for some index set I, where S_α is the unilateral shift on a Hilbert space $\mathcal{H}_\alpha$ and U is a unitary operator. Since one can extend any unilateral shift to a unitary bilateral shift, one can extend $\mathcal{K}$ to a unitary operator $\mathcal{K}'$.

so that $d\mu_g(\lambda) = d\nu_g(\theta)$. We use the notation $\int_{[-\pi,\pi]_{\mathrm{per}}}$ to denote integration along the periodic interval $[-\pi,\pi]_{\mathrm{per}}$ as $\int_{-\pi}^{\pi}$ is ambiguous since spectral measures can have atoms at $\pm\pi$. The precise choice of g is up to the practitioner: smooth g makes ν_g easier to compute but tends to blur out the spectral information of $\mathcal{K}$, whereas ν_g is more challenging to compute for non-smooth g but can give better resolution of the underlying dynamics. In other situations the application itself dictates that a particular g is of interest.

To compute ν_g, we start by noting that the Fourier coefficients of ν_g are given by

$$\widehat{\nu_g}(n) := \frac{1}{2\pi}\int_{[-\pi,\pi]_{\mathrm{per}}} e^{-in\theta}\, d\nu_g(\theta) = \frac{1}{2\pi}\int_{\mathbb{T}} \lambda^{-n}\, d\mu_g(\lambda) = \frac{1}{2\pi}\langle \mathcal{K}'^{-n} g, g\rangle, \qquad n \in \mathbb{Z}. \tag{9.3.1}$$

Since $\mathcal{K}'$ is a unitary operator its inverse is its adjoint and thus, the Fourier coefficients of ν_g can be expressed in terms of correlations $\langle \mathcal{K}^n g, g\rangle$ and $\langle g, \mathcal{K}^n g\rangle$. That is, for $g \in L^2(\Omega,\omega)$,

$$\widehat{\nu_g}(n) = \frac{1}{2\pi}\langle \mathcal{K}^{-n} g, g\rangle, \quad n < 0, \qquad \widehat{\nu_g}(n) = \frac{1}{2\pi}\langle g, \mathcal{K}^{n} g\rangle, \quad n \geq 0. \tag{9.3.2}$$

Since $g \in L^2(\Omega,\omega)$ and (9.3.2) only depend on correlations with $\mathcal{K}$, and ν_g is determined by its Fourier coefficients, we find that ν_g is independent of the choice of unitary extension $\mathcal{K}'$. Henceforth, we can safely dispense with the extension $\mathcal{K}'$, and call ν_g the spectral measure of $\mathcal{K}$ with respect to g.

From (9.3.2), we find that $\widehat{\nu_g}(-n) = \overline{\widehat{\nu_g}(n)}$ for $n \in \mathbb{Z}$, which tells us that ν_g is completely determined by the forward-time dynamical *autocorrelations* $\langle g, \mathcal{K}^n g\rangle$ with $n \geq 0$. Equivalently, the spectral measure of $\mathcal{K}$ with respect to almost every $g \in L^2(\Omega,\omega)$ is a signature for the forward-time dynamics of (9.0.1). This is because ν_g completely determines $\mathcal{K}$ when g is cyclic, i.e., when the closure of $\mathrm{span}\{g, \mathcal{K}g, \mathcal{K}^2 g, \ldots\}$ is $L^2(\Omega,\omega)$, and almost every g is cyclic. If g is not cyclic, then ν_g only determines the action of $\mathcal{K}$ on the closure of $\mathrm{span}\{g, \mathcal{K}g, \mathcal{K}^2 g, \ldots\}$, which can still be useful if one is interested in particular observables.

Continuous and discrete parts of spectra, and Koopman mode decompositions

Of particular importance to dynamical systems is Lebesgue's decomposition of ν_g:

$$d\nu_g(y) = \underbrace{\sum_{\lambda=\exp(i\theta)\in \mathrm{Sp_p}(\mathcal{K})} \langle \mathcal{P}_\lambda g, g\rangle\, \delta(y-\theta) dy}_{\text{discrete part}} + \underbrace{\rho_g(y)\, dy + d\nu_g^{(\mathrm{sc})}(y)}_{\text{continuous part}}. \tag{9.3.3}$$

The discrete (or atomic) part of ν_g is a sum of Dirac delta distributions, supported on $\mathrm{Sp_p}(\mathcal{K})$, the set of eigenvalues of $\mathcal{K}$.[4] The coefficient of each δ in the sum is $\langle \mathcal{P}_\lambda g, g\rangle = \|\mathcal{P}_\lambda g\|^2$, where $\mathcal{P}_\lambda$ is the orthogonal spectral projector associated with the eigenvalue λ. The continuous part of ν_g consists of a part that is absolutely continuous with respect to the Lebesgue measure, with Radon–Nikodym derivative $\rho_g \in L^1([-\pi,\pi]_{\mathrm{per}})$, and a singular continuous component $\nu_g^{(\mathrm{sc})}$. The decomposition in (9.3.3) provides important information on the evolution of dynamical systems. For example, suppose that there is no singular continuous spectrum, then any $g \in L^2(\Omega,\omega)$ can be written as

$$g = \sum_{\lambda\in \mathrm{Sp_p}(\mathcal{K})} c_\lambda \varphi_\lambda + \int_{[-\pi,\pi]_{\mathrm{per}}} \phi_{\theta,g}\, d\theta,$$

where the φ_λ are the eigenfunctions of $\mathcal{K}$, c_λ are expansion coefficients and $\phi_{\theta,g}$ is a "continuously parametrised" collection of eigenfunctions.[5] Then, one obtains the Koopman mode decomposition [Mez05]

$$g(\boldsymbol{x}_n) = [\mathcal{K}^n g](\boldsymbol{x}_0) = \sum_{\lambda\in \mathrm{Sp_p}(\mathcal{K})} c_\lambda \lambda^n \varphi_\lambda(\boldsymbol{x}_0) + \int_{[-\pi,\pi]_{\mathrm{per}}} e^{in\theta} \phi_{\theta,g}(\boldsymbol{x}_0)\, d\theta. \tag{9.3.4}$$

[4] After mapping to the periodic interval, the discrete part of ν_g is supported on the closure of $\mathrm{Sp_p}(\mathcal{K}')$. However, we can always choose the extension $\mathcal{K}'$ so that $\mathrm{Sp_p}(\mathcal{K}') = \mathrm{Sp_p}(\mathcal{K})$ with the same eigenspaces.

[5] To be precise, $\phi_{\theta,g}\, d\theta$ is the absolutely continuous component of $dE^{\mathcal{K}'}(\theta) g$ and $\rho_g(\theta) = \langle \phi_{\theta,g}, g\rangle$.

Often, one can also characterise a dynamical system in terms of these decompositions. For example, suppose F is measure-preserving and bijective, and ω is a probability measure. In that case, the dynamical system is: (1) ergodic if and only if $\lambda = 1$ is a simple eigenvalue of $\mathcal{K}$, (2) weakly mixing if and only if $\lambda = 1$ is a simple eigenvalue of $\mathcal{K}$ and there are no other eigenvalues, and (3) mixing if $\lambda = 1$ is a simple eigenvalue of $\mathcal{K}$ and $\mathcal{K}$ has absolutely continuous spectrum on $\mathrm{span}\{1\}^{\perp}$ [Hal17]. Different spectral types also have interpretations in the context of fluid mechanics [Mez13], and weakly autonomous transport where the Koopman operator has singular continuous spectra [Zas02].

9.3.2 General framework for computations

To develop convergence theory, we consider convolution with kernels. We form an approximation to ν_g by convolution. That is, we define

$$\nu_g^{\epsilon}(\theta_0) = \int_{[-\pi,\pi]_{\mathrm{per}}} K_{\epsilon}(\theta_0 - \theta) d\nu_g(\theta),$$

where K_ϵ are a family of integrable functions $\{K_\epsilon : 0 < \epsilon \leq 1\}$ satisfying certain properties (see Definition 9.3.1) so that ν_g^{ϵ} converges to ν_g in some sense. The most famous example of K_ϵ is the Poisson kernel for the unit disc given by

$$K_{\epsilon}(\theta) = \frac{1}{2\pi} \frac{(1+\epsilon)^2 - 1}{1 + (1+\epsilon)^2 - 2(1+\epsilon)\cos(\theta)}, \tag{9.3.5}$$

in polar coordinates with $r = (1+\epsilon)^{-1}$. The Poisson kernel is a first-order kernel because, up to a logarithmic factor, it leads to a first-order algebraic rate of convergence of ν_g^{ϵ} to ν_g. We now give the following general definition of an mth order kernel, and justify their name by showing that they lead to an mth order rate of convergence of ν_g^{ϵ} to ν_g in a weak and pointwise sense (see Section 9.3.2).

Definition 9.3.1 (mth order periodic kernel). *Let $\{K_\epsilon : 0 < \epsilon \leq 1\}$ be a family of integrable functions on the periodic interval $[-\pi,\pi]_{\mathrm{per}}$. We say that $\{K_\epsilon\}$ is an mth order kernel for $[-\pi,\pi]_{\mathrm{per}}$ if*

(i) (Normalised) $\int_{[-\pi,\pi]_{\mathrm{per}}} K_\epsilon(\theta)\, d\theta = 1$.

(ii) (Approximately vanishing moments) There exists a constant C_K such that

$$\left| \int_{[-\pi,\pi]_{\mathrm{per}}} \theta^n K_\epsilon(\theta)\, d\theta \right| \leq C_K \epsilon^m \log(\epsilon^{-1}), \quad \textit{for any integer } 1 \leq n \leq m-1. \tag{9.3.6}$$

(iii) (Decay away from 0) For any $\theta \in [-\pi,\pi]$ and $0 < \epsilon \leq 1$,

$$|K_\epsilon(\theta)| \leq \frac{C_K \epsilon^m}{(\epsilon + |\theta|)^{m+1}}. \tag{9.3.7}$$

The conditions in Definition 9.3.1 are mostly technical assumptions that allow one to prove appropriate convergence rates of ν_g^{ϵ} to ν_g. For pointwise convergence, property (iii) is required to apply a local cut-off argument away from singular parts of the measure. Properties (i) and (ii) are used to show that terms vanish in a local Taylor series expansion of the Radon–Nikodym derivative, and the remainder is bounded by (iii). For weak convergence, we apply similar arguments to the test function by Fubini's theorem.

Pointwise convergence

For a point $\theta_0 \in [-\pi,\pi]$, the value of the approximate spectral measure $\nu_g^\epsilon(\theta_0)$ converges to the Radon–Nikodym derivative, $\rho_g(\theta_0)$ provided that ν_g is absolutely continuous in an interval containing θ_0 (without this separation condition it still converges for almost every θ_0). The precise rate of convergence depends on the smoothness of ρ_g in a small interval I containing θ_0. In particular, we write $\rho_g \in \mathcal{C}^{n,\alpha}(I)$ if ρ_g is n-times continuously differentiable on I and the nth derivative is Hölder continuous with parameter $0 \leq \alpha < 1$. For $h_1 \in \mathcal{C}^{0,\alpha}(I)$ and $h_2 \in \mathcal{C}^{k,\alpha}(I)$ we define the seminorm and norm, respectively, as

$$|h_1|_{\mathcal{C}^{0,\alpha}(I)} = \sup_{x\neq y\in I} \frac{|h_1(x)-h_1(y)|}{|x-y|^\alpha}, \quad \|h_2\|_{\mathcal{C}^{k,\alpha}(I)} = |h_2^{(k)}|_{\mathcal{C}^{0,\alpha}(I)} + \max_{0\leq j\leq k} \|h_2^{(j)}\|_{\infty,I}.$$

We state the following pointwise convergence theorem for general complex-valued measures ν as we apply it to measures corresponding to test functions to prove Theorem 9.3.3. The choice $\nu = \nu_g$ with $\|\nu_g\| = 1$ in Theorem 9.3.2 gives pointwise convergence of spectral measures.

Theorem 9.3.2 (Pointwise convergence). *Let $\{K_\epsilon\}$ be an mth order kernel for $[-\pi,\pi]_{\mathrm{per}}$ and let ν be a complex-valued measure on $[-\pi,\pi]_{\mathrm{per}}$ with finite total variation $\|\nu\|$. Suppose that for some $\theta_0 \in [-\pi,\pi]$ and $\eta \in (0,\pi)$, ν is absolutely continuous on $I = (\theta_0-\eta,\theta_0+\eta)$ with Radon–Nikodym derivative $\rho \in \mathcal{C}^{n,\alpha}(I)$ ($\alpha \in [0,1)$). Then the following hold for any $0 \leq \epsilon < 1$:*

(i) If $n+\alpha < m$, then

$$\left|\rho(\theta_0) - \int_{[-\pi,\pi]_{\mathrm{per}}} K_\epsilon(\theta_0-\theta)\,d\nu(\theta)\right| \lesssim C_K\big(\|\nu\| + \|\rho\|_{\mathcal{C}^{n,\alpha}(I)}\big)\left(\epsilon^{n+\alpha} + \frac{\epsilon^m}{(\epsilon+\eta)^{m+1}}\right)\big(1+\eta^{-n-\alpha}\big). \tag{9.3.8}$$

(ii) If $n+\alpha \geq m$, then

$$\left|\rho(\theta_0) - \int_{[-\pi,\pi]_{\mathrm{per}}} K_\epsilon(\theta_0-\theta)\,d\nu(\theta)\right| \lesssim C_K\big(\|\nu\| + \|\rho\|_{\mathcal{C}^{m}(I)}\big)\left(\epsilon^{m}\log(\epsilon^{-1}) + \frac{\epsilon^m}{(\epsilon+\eta)^{m+1}}\right)\big(1+\eta^{-m}\big). \tag{9.3.9}$$

Here, '$\lesssim$' means that the inequality holds up to a constant that only depends on $n+\alpha$ and m.

Proof. By periodicity, we can assume without loss of generality that $\theta_0 = 0$. First, we decompose ρ into two parts $\rho = \rho_1 + \rho_2$, where $\rho_1 \in \mathcal{C}^{n,\alpha}(I)$ is compactly supported on I and ρ_2 vanishes on $(-\eta/2, +\eta/2)$. Using (9.3.7), we have

$$\left|\rho(0) - \int_{[-\pi,\pi]_{\mathrm{per}}} K_\epsilon(-\theta)d\nu(\theta)\right| \leq \left|\rho_1(0) - \int_{[-\pi,\pi]_{\mathrm{per}}} K_\epsilon(-\theta)\rho_1(\theta)\,d\theta\right| + \int_{|\theta|>\eta/2} \frac{C_K\epsilon^m\,d|\nu^{\mathrm{r}}|(\theta)}{(\epsilon+\eta/2)^{m+1}}, \tag{9.3.10}$$

where $d\nu^{\mathrm{r}}(\theta) = d\nu(\theta) - \rho_1(\theta)\,d\theta$. The second term on the right-hand side of (9.3.10) is bounded by $\hat{C}_1 C_K(\|\nu\| + \|\rho_1\|_{L^\infty(I)})\epsilon^m(\epsilon+\eta)^{-(m+1)}$ for some constant $\hat{C}_1$ independent of all parameters. To bound the first term, we expand ρ_1 using Taylor's theorem:

$$\rho_1(\theta) = \sum_{j=0}^{k-1} \frac{\rho_1^{(j)}(0)}{j!}\theta^j + \frac{\rho_1^{(k)}(\xi_\theta)}{k!}\theta^k, \qquad k = \min(n,m), \tag{9.3.11}$$

where $|\xi_\theta| \leq |\theta|$. We now consider the two cases of the theorem separately.

Case (i): $\mathbf{n+\alpha < m}$. In this case, $k = n$ and we can select ρ_1 so that,

$$\|\rho_1\|_{\mathcal{C}^{n,\alpha}(I)} \leq C(n,\alpha)\|\rho\|_{\mathcal{C}^{n,\alpha}(I)}\left(1+\eta^{-n-\alpha}\right), \quad \|\rho_1\|_{L^\infty(I)} \leq C(n,\alpha)\|\rho\|_{\mathcal{C}^{n,\alpha}(I)} \tag{9.3.12}$$

for some universal constant $C(n,\alpha)$ that only depends on n and α. Existence of such a decomposition follows from standard arguments with cut-off functions. Using (9.3.11), part (ii) of Definition 9.3.1 and the first bound of (9.3.12), we obtain

$$\begin{aligned}\left|\rho_1(0)-\int_{[-\pi,\pi]_{\mathrm{per}}}K_\epsilon(-\theta)\rho_1(\theta)\,d\theta\right| \leq &\hat{C}_2C_K\|\rho\|_{\mathcal{C}^{n,\alpha}(I)}\epsilon^m\log(\epsilon^{-1})\left(1+\eta^{-n-\alpha}\right)\\ &+\left|\int_{[-\pi,\pi]_{\mathrm{per}}}K_\epsilon(-\theta)\frac{\rho_1^{(n)}(\xi_\theta)-\rho_1^{(n)}(0)}{n!}\theta^n\,d\theta\right|,\end{aligned} \tag{9.3.13}$$

for some constant $\hat{C}_2$ independent of ϵ, η and ν (or ρ,ρ_1,ρ_2). Note that we have added the factor of $\rho_1^{(n)}(0)$ into the integrand by a second application of part (ii) of Definition 9.3.1 and the fact that $n<m$. The Hölder continuity of $\rho_1^{(n)}$ implies that $|\rho_1^{(n)}(\xi_\theta)-\rho_1^{(n)}(0)|\leq C(n,\alpha)\|\rho\|_{\mathcal{C}^{n,\alpha}(I)}\left(1+\eta^{-n-\alpha}\right)\theta^\alpha$. Using this bound in the integrand on the right-hand side of (9.3.13) and (9.3.7), we obtain

$$\left|\rho_1(0)-\int_{[-\pi,\pi]_{\mathrm{per}}}K_\epsilon(-\theta)\rho_1(\theta)\,d\theta\right| \leq \hat{C}_3C_K\|\rho\|_{\mathcal{C}^{n,\alpha}(I)}\left(\epsilon^m\log(\epsilon^{-1})+\epsilon^{n+\alpha}\int_0^{\pi/\epsilon}\frac{\tau^{n+\alpha}d\tau}{(1+\tau)^{m+1}}\right)\left(1+\eta^{-n-\alpha}\right),$$

for some constant $\hat{C}_3$ independent of ϵ, η and ν (or ρ,ρ_1,ρ_2). Since $m>n+\alpha$, the integral in brackets converges as $\epsilon\downarrow 0$ and the bound in (9.3.8) now follows.

Case (ii): $\mathbf{n}+\alpha\geq\mathbf{m}$. In this case, $k=m$ and we can select ρ_1 such that

$$\|\rho_1\|_{\mathcal{C}^m(I)}\leq C(m)\|\rho\|_{\mathcal{C}^m(I)}\left(1+\eta^{-m}\right),$$

for some universal constant $C(m)$ that only depends on m. Again, existence of such a decomposition follows from standard arguments with cut-off functions. Using (9.3.11) and applying (9.3.6) to the powers θ_j for $j<m$ and (9.3.7) to the θ^m term, we obtain

$$\left|\rho_1(0)-\int_{[-\pi,\pi]_{\mathrm{per}}}K_\epsilon(-\theta)\rho_1(\theta)\,d\theta\right| \leq \hat{C}_2C_K\|\rho\|_{\mathcal{C}^m(I)}\left(\epsilon^m\log(\epsilon^{-1})+\epsilon^m\int_0^{\pi/\epsilon}\frac{\tau^m d\tau}{(1+\tau)^{m+1}}\right)\left(1+\eta^{-m}\right),$$

for some constant $\hat{C}_2$ independent of ϵ, η and ν (or ρ,ρ_1,ρ_2). The bound in (9.3.9) now follows. □

Weak convergence

We now turn to proving weak convergence.

Theorem 9.3.3 (Weak convergence). *Let $\{K_\epsilon\}$ be an mth order kernel for $[-\pi,\pi]_{\mathrm{per}}$, $\phi\in\mathcal{C}^{n,\alpha}([-\pi,\pi]_{\mathrm{per}})$, and let ν_g be a spectral measure on the periodic interval $[-\pi,\pi]_{\mathrm{per}}$. Then*

$$\left|\int_{[-\pi,\pi]_{\mathrm{per}}}\phi(\theta)\nu_g^\epsilon(\theta)\,d\theta-\int_{[-\pi,\pi]_{\mathrm{per}}}\phi(\theta)\,d\nu_g(\theta)\right|\lesssim C_K\|\phi\|_{\mathcal{C}^{n,\alpha}([-\pi,\pi]_{\mathrm{per}})}\left(\epsilon^{n+\alpha}+\epsilon^m\log(\epsilon^{-1})\right), \tag{9.3.14}$$

where '$\lesssim$' means that the inequality holds up to a constant that only depends on $n+\alpha$ and m.

Proof. Let $\tilde{K}_\epsilon(\theta)=K_\epsilon(-\theta)$, then it is easily seen that $\{\tilde{K}_\epsilon\}$ is an mth order kernel for $[-\pi,\pi]_{\mathrm{per}}$. Fubini's theorem allows us to exchange the order of integration to see that

$$\int_{[-\pi,\pi]_{\mathrm{per}}}\phi(\theta)\nu_g^\epsilon(\theta)\,d\theta=\int_{[-\pi,\pi]_{\mathrm{per}}}\phi(\theta)[K_\epsilon*\nu_g](\theta)\,d\theta=\int_{[-\pi,\pi]_{\mathrm{per}}}[\tilde{K}_\epsilon*\phi](\theta)\,d\nu_g(\theta).$$

We can now apply Theorem 9.3.2 to the absolutely continuous measure with Radon–Nikodym derivative ϕ and the kernel $\tilde{K}_\epsilon$ (e.g., with $\eta = \pi/2$) to see that

$$\left|[\tilde{K}_\epsilon * \phi](\theta) - \phi(\theta)\right| \leq C_1 C_K \|\phi\|_{\mathcal{C}^{n,\alpha}([-\pi,\pi]_{\mathrm{per}})} \left(\epsilon^{n+\alpha} + \epsilon^m \log(\epsilon^{-1})\right),$$

for some constant C_1 depending on n, α and m. Since ν_g is a probability measure, (9.3.14) follows. □

The high-order convergence in Theorem 9.3.3 does not require any regularity assumptions on ν_g. Moreover, though not covered by the theorem, for any mth order kernel and continuous periodic function ϕ, weak convergence still holds.

Recovery of the atomic parts of the spectral measure

Finally, we consider the recovery of the atomic parts of spectral measures or, equivalently, $\mathrm{Sp}_{\mathrm{p}}(\mathcal{K})$ - the set of eigenvalues of $\mathcal{K}$ (see (9.3.3)). This convergence is achieved by *rescaling* the smoothed approximation $K_\epsilon * \nu_g$.

Theorem 9.3.4 (Recovery of atoms)**.** *Let $\{K_\epsilon\}$ be an mth order kernel for $[-\pi,\pi]_{\mathrm{per}}$ that satisfies*

$$\limsup_{\epsilon\downarrow 0} \frac{\epsilon^{-1}}{|K_\epsilon(0)|} < \infty,$$

and let ν_g be a spectral measure on $[-\pi,\pi]_{\mathrm{per}}$. Then, for any $\theta_0 \in [-\pi,\pi]_{\mathrm{per}}$,

$$\nu_g(\{\theta_0\}) = \lim_{\epsilon\downarrow 0} \frac{1}{K_\epsilon(0)}[K_\epsilon * \nu_g](\theta_0). \tag{9.3.15}$$

Proof. By periodicity, we may assume without loss of generality that $\theta_0 = 0$. Let $\nu_g' = \nu_g - \nu_g(\{0\})\delta_0$, then

$$\frac{1}{K_\epsilon(0)}[K_\epsilon * \nu_g](0) = \nu_g(\{0\}) + \frac{1}{K_\epsilon(0)}[K_\epsilon * \nu_g'](0). \tag{9.3.16}$$

Consider the function $K_\epsilon(-\theta)/K_\epsilon(0)$, which is uniformly bounded for sufficiently small ϵ using (9.3.7) and the assumption $\limsup_{\epsilon\downarrow 0} \frac{\epsilon^{-1}}{|K_\epsilon(0)|} < \infty$. Since $\lim_{\epsilon\downarrow 0} K_\epsilon(-\theta)/K_\epsilon(0) = 0$ for any $\theta \neq 0$ and $\nu_g'(\{0\}) = 0$,

$$\lim_{\epsilon\downarrow 0} \frac{1}{K_\epsilon(0)}[K_\epsilon * \nu_g'](0) = \lim_{\epsilon\downarrow 0} \int_{[-\pi,\pi]_{\mathrm{per}}} \frac{K_\epsilon(-\theta)}{K_\epsilon(0)} d\nu_g' = 0,$$

where we used the dominated convergence theorem. Using (9.3.16), the theorem now follows. □

The condition that $\limsup_{\epsilon\downarrow 0} \frac{\epsilon^{-1}}{|K_\epsilon(0)|} < \infty$ is a technical condition that is satisfied by all the kernels constructed in this chapter. A condition such as this is required to recover the atomic part of ν_g, as it says that K_ϵ must become localised around 0 sufficiently quickly as $\epsilon \to 0$.

9.3.3 Computation from autocorrelations

We now suppose that one has already computed the autocorrelations $\langle g, \mathcal{K}^n g\rangle$ for $0 \leq n \leq N$ and would like to recover a smoothed approximation of ν_g. Since the Fourier coefficients of ν_g are given by autocorrelations (see (9.3.1)), the task is similar to Fourier recovery [GS97b, AH12]. We are particularly interested in approaches with good convergence properties as $N \to \infty$, as this reduces the number of computed autocorrelations and the sample size M required for good recovery of the spectral measure.

Motivated by the classical task of recovering a continuous function by its partial Fourier series, we start by considering the "windowing trick" from sampling theory. That is, we define a smoothed approximation to ν_g as

$$\nu_{g,N}(\theta) = \sum_{n=-N}^{N} \varphi\Big(\frac{n}{N}\Big)\, \widehat{\nu_g}(n)e^{in\theta} = \frac{1}{2\pi}\sum_{n=-N}^{-1} \varphi\Big(\frac{n}{N}\Big)\, \overline{\langle g, \mathcal{K}^{-n}g\rangle}e^{in\theta} + \frac{1}{2\pi}\sum_{n=0}^{N} \varphi\Big(\frac{n}{N}\Big)\, \langle g, \mathcal{K}^{n}g\rangle e^{in\theta}. \tag{9.3.17}$$

The function $\varphi : [-1,1] \to \mathbb{R}$ is often called a filter function. The idea of φ is that $\varphi(x)$ is close to 1 when x is close to 0, and φ tapers to 0 near $x = \pm 1$. By carefully tapering φ, the partial sum in (9.3.17) converges to ν_g as $N \to \infty$. For fast pointwise or weak convergence of $\nu_{g,N}$ to ν_g, it is desirable for φ to be an even function that smoothly tapers from 1 to 0.

One of the simplest filters is the hat function $\varphi_{\text{hat}}(x) = 1 - |x|$, for which (9.3.17) corresponds to the classical Cesàro summation of Fourier series. With this choice of φ, $\nu_{g,N}(\theta)$ is the convolution of ν_g with the famous Fejèr kernel, $F_N(\theta) = \sum_{n=-N}^{N}(1 - |n|/N)e^{in\theta}$. Other filter functions can provide a faster rate of convergence than $\varphi_{\text{hat}}(x) = 1 - |x|$, including the cosine and fourth-order filters [GS97b]:

$$\varphi_{\cos}(x) = \frac{1}{2}(1 - \cos(\pi x)), \qquad \varphi_{\text{four}}(x) = 1 - x^4(-20|x|^3 + 70x^2 - 84|x| + 35).$$

For the recovery of measures, we find that a particularly good choice is

$$\varphi_{\text{bump}}(x) = \exp\left(-\frac{2}{1-|x|}\exp\left(-\frac{c}{x^4}\right)\right), \qquad c \approx 0.109550455106347, \tag{9.3.18}$$

where the value of c is selected so that $\varphi_{\text{bump}}(1/2) = 1/2$. This filter can lead to arbitrary high orders of convergence with errors between $\nu_{g,N}$ and ν_g that go to zero faster than any polynomial in N^{-1}. A further useful property is that $\nu_{g,N}$ localises any singular behavior of ν_g.[6]

Algorithm 5 summarises our computational framework for recovering a smoothed version of ν_g from autocorrelations of the trajectory data. It is easy to verify that $\nu_{g,N} = K_\epsilon * \nu_g$ with

$$K_\epsilon(\theta) = \frac{1}{2\pi}\sum_{n=-N}^{N} \varphi\left(\frac{n}{N}\right) e^{in\theta}, \quad N = \lfloor \epsilon^{-1} \rfloor.$$

The properties of an mth order kernel can be translated to properties of a filter and we can therefore use the convergence theory of Section 9.3.2.

Proposition 9.3.5. *Let $m \in \mathbb{N}$ and suppose that φ is an even continuous function that is compactly supported on $[-1,1]$ such that (a) $\varphi \in \mathcal{C}^{m-1}([-1,1])$, (b) $\varphi(0) = 1$ and $\varphi^{(n)}(0) = 0$ for any integer $1 \le n \le m-1$, (c) $\varphi^{(n)}(1) = 0$ for any integer $0 \le n \le m-1$, and (d) $\varphi|_{[0,1]} \in \mathcal{C}^{m+1}([0,1])$. Then,*

$$K_\epsilon(\theta) = \frac{1}{2\pi}\sum_{n=-N}^{N} \varphi\left(\frac{n}{N}\right) e^{in\theta}, \qquad N = \lfloor \epsilon^{-1} \rfloor \tag{9.3.19}$$

is an mth order kernel for $[-\pi,\pi]_{\text{per}}$.

Exercise: Prove Proposition 9.3.5 using the Poisson summation formula.

Therefore, it can be verified that: φ_{hat}, $\varphi_{\cos}$, and φ_{four} induce first-order, second-order and fourth-order kernels in (9.3.19), respectively. Similarly, φ_{bump} induces a kernel that is mth order for any $m \in \mathbb{N}$. For example, up to a logarithmic factor, the rate of convergence for φ_{four} is $\mathcal{O}(\epsilon^4)$ as $\epsilon \to 0$ (resp. $\mathcal{O}(N^{-4})$ as $N \to \infty$) in a weak and pointwise sense.

[6] This because the kernel associated with φ_{bump} (see Proposition 9.3.5) is highly localised due to the smoothness of φ_{bump}.

Algorithm 5 A computational framework for recovering an approximation of the spectral measure ν_g associated with a Koopman operator that is an isometry.

Input: Trajectory data, a filter φ, and an observable $g \in L^2(\Omega, \omega)$.

```
1: Approximate the autocorrelations $a_n = \frac{1}{2\pi}\langle g, \mathcal{K}^n g\rangle$ for $0 \leq n \leq N$. (The precise value of $N$ and the
   approach depends on the trajectory data (see Section 9.2.3).)
2: Set $a_{-n} = \overline{a_n}$ for $1 \leq n \leq N$.
```

Output: The function $\nu_{g,N}(\theta) = \sum_{n=-N}^{N} \varphi\left(\frac{n}{N}\right) a_n e^{in\theta}$ that can be evaluated for any $\theta \in [-\pi, \pi]_{\text{per}}$.

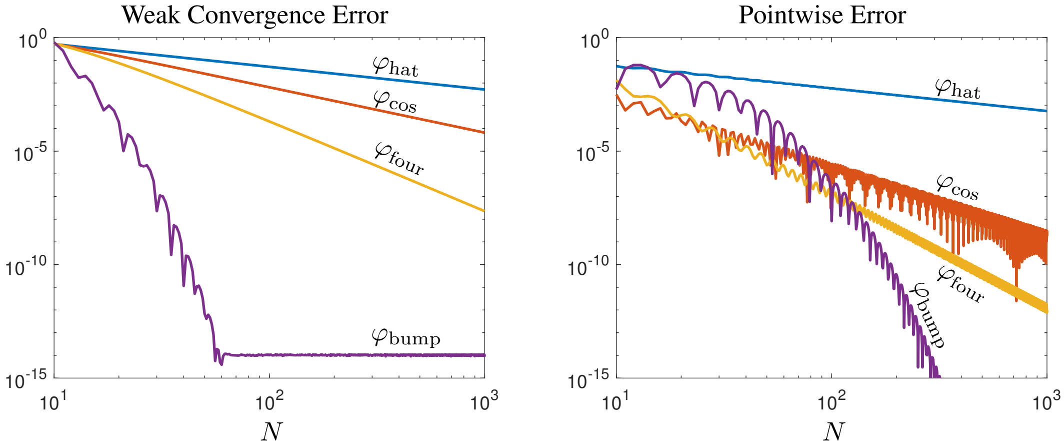


Figure 9.1: Relative errors between $\nu_{g,N}$ and ν_g for the shift operator computed with filters φ_{hat} (blue), φ_{cos} (red), φ_{four} (yellow), and φ_{bump} (purple). Left: Relative error between $\nu_{g,N}$ to ν_g in the sense of weak convergence for the test function $\phi(\theta) = \cos(5\theta)/(2 + \cos(\theta))$. Right: Relative error between $\nu_{g,N}$ to ρ_g at $\theta = 0$, illustrating pointwise convergence.

As an example, consider the shift operator with state-space $\Omega = \mathbb{Z}$ (and counting measure ω) given by

$$x_{n+1} = F(x_n), \qquad F(x) = x + 1.$$

We seek to compute the spectral measure ν_g with respect to $g \in L^2(\mathbb{Z}, \omega) = \ell^2(\mathbb{Z})$, where $\ell^2(\mathbb{Z})$ is the space of square summable doubly infinite vectors. This example is a building block of many dynamical systems, such as Bernoulli shifts, with so-called Lebesgue spectrum [AA68, Chapter 2]. We consider the observable $g(k) = C\sin(k)/k$, where $C \approx 0.564189583547756$ is a normalisation constant so that $\|g\| = 1$. For this example, ν_g is absolutely continuous but ρ_g has discontinuities at $\theta = \pm 1$. Figure 9.1 shows the weak convergence (left) and pointwise convergence (right) for various filters.

9.3.4 Computation using ResDMD

We now develop rational kernels that allow us to compute smoothed approximations of spectral measures from the ResDMD matrices $\Psi_X^* W \Psi_X$, $\Psi_X^* W \Psi_Y$, and $\Psi_Y^* W \Psi_Y$. Moreover, these matrices can be reused to computed spectral measures with respect to different observable functions g.

The following lemma will be used to build mth order rational kernels. It provides sufficient conditions for a family of integrable functions to be an mth order kernel.

Lemma 9.3.6. *Let $\{K_\epsilon : \epsilon \in (0, 1]\}$ be a family of integrable functions on the periodic interval $[-\pi, \pi]_{\text{per}}$*

that integrate to 1*. Suppose that there exists a constant C such that for any integer n with $0 < n \leq m-1$,*

$$\left|\int_{[-\pi,\pi]_{\rm per}} K_\epsilon(-\theta)e^{in\theta}d\theta - 1\right| \leq C\epsilon^m\log(\epsilon^{-1}), \tag{9.3.20}$$

and such that

$$|K_\epsilon(\theta)| \leq \frac{C\epsilon^m}{(\epsilon+|\theta|)^{m+1}}, \tag{9.3.21}$$

for any $\theta \in [-\pi,\pi]_{\rm per}$ and $\epsilon \in (0,1]$. Then $\{K_\epsilon\}$ is an mth order kernel for $[-\pi,\pi]_{\rm per}$.

Exercise: Prove Lemma 9.3.6.

We begin by considering a unitary extension $\mathcal{K}'$ of $\mathcal{K}$, which is defined on a Hilbert space $\mathcal{H}'$ that is an extension of $L^2(\Omega,\omega)$. Let $z \in \mathbb{C}$ with $|z| > 1$ and $g \in L^2(\Omega,\omega)$. Since $\|\mathcal{K}\| = 1 < |z|$ for $z \notin \mathrm{Sp}(\mathcal{K})$, $(\mathcal{K}'-z)^{-1}g = (\mathcal{K}-z)^{-1}g$ and

$$\langle(\mathcal{K}-z)^{-1}g, \mathcal{K}^*g\rangle = \langle\mathcal{K}'(\mathcal{K}'-z)^{-1}g, g\rangle_{\mathcal{H}'} = \int_{\mathbb{T}} \frac{\lambda}{\lambda - z}\,d\mu_g(\lambda) = \int_{[-\pi,\pi]_{\rm per}} \frac{e^{i\theta}}{e^{i\theta}-z}\,d\nu_g(\theta), \tag{9.3.22}$$

where the last equality follows from a change-of-variables. If $z \neq 0$ with $|z| < 1$, then z may be in $\mathrm{Sp}(\mathcal{K})$ since $\mathcal{K}$ is not necessarily unitary. However, since $|\overline{z}^{-1}| > 1$, $\overline{z}^{-1} \notin \mathrm{Sp}(\mathcal{K})$ and hence $(\mathcal{K}'-\overline{z}^{-1})^{-1}g = (\mathcal{K}-\overline{z}^{-1})^{-1}g$. Since ν_g is a real-valued measure, we find that

$$\langle g, (\mathcal{K}-\overline{z}^{-1})^{-1}g\rangle = \langle g, (\mathcal{K}'-\overline{z}^{-1})^{-1}g\rangle_{\mathcal{H}'} = \overline{\int_{[-\pi,\pi]_{\rm per}} \frac{d\nu_g(\theta)}{e^{i\theta}-\overline{z}^{-1}}} = -z\int_{[-\pi,\pi]_{\rm per}} \frac{e^{i\theta}d\nu_g(\theta)}{e^{i\theta}-z}. \tag{9.3.23}$$

The leftmost and rightmost sides of (9.3.22) and (9.3.23) are independent of $\mathcal{K}'$, so we can safely dispense with the extension and have an expression for a generalised Cauchy transform of ν_g, i.e.,

$$\mathtt{C}_{\nu_g}(z) = \frac{1}{2\pi}\int_{[-\pi,\pi]_{\rm per}} \frac{e^{i\theta}d\nu_g(\theta)}{e^{i\theta}-z} = \frac{1}{2\pi}\begin{cases}\langle(\mathcal{K}-z)^{-1}g, \mathcal{K}^*g\rangle, & \text{if } |z|>1,\\ -z^{-1}\langle g, (\mathcal{K}-\overline{z}^{-1})^{-1}g\rangle, & \text{if } z\neq 0 \text{ with } |z|<1.\end{cases} \tag{9.3.24}$$

The importance of (9.3.24) is that it relates $\mathtt{C}_{\nu_g}$ to the resolvent operator $(\mathcal{K}-z)^{-1}$ for $|z| > 1$. Below, we show how to compute the resolvent operator from snapshot data for $|z| > 1$. Since $|z| > 1$, we can provide convergence results and stability results even when we replace $\mathcal{K}$ by a discretisation.

To recover ν_g from $\mathtt{C}_{\nu_g}$, a derivation motivated by the Sokhotski–Plemelj formula shows that

$$\mathtt{C}_{\nu_g}\left(e^{i\theta_0}(1+\epsilon)^{-1}\right) - \mathtt{C}_{\nu_g}\left(e^{i\theta_0}(1+\epsilon)\right) = \int_{[-\pi,\pi]_{\rm per}} K_\epsilon(\theta_0-\theta)\,d\nu_g(\theta), \tag{9.3.25}$$

where K_ϵ is the Poisson kernel for the unit disc (see (9.3.5)). The Poisson kernel for the unit disc is a first-order kernel. We can generalise the Sokhotski–Plemelj-like formula in (9.3.25) to develop high-order rational kernels. Let $\{z_j\}_{j=1}^m$ be distinct points with positive real part and consider the rational function:

$$K_\epsilon(\theta) = \frac{e^{-i\theta}}{2\pi}\sum_{j=1}^m\left[\frac{c_j}{e^{-i\theta}-(1+\epsilon\overline{z_j})^{-1}} - \frac{d_j}{e^{-i\theta}-(1+\epsilon z_j)}\right]. \tag{9.3.26}$$

A short derivation using (9.3.24) shows that

$$\begin{aligned}\int_{[-\pi,\pi]_{\rm per}} K_\epsilon(\theta_0-\theta)\,d\nu_g(\theta) &= \sum_{j=1}^m\left[c_j\mathtt{C}_{\nu_g}\left(e^{i\theta_0}(1+\epsilon\overline{z_j})^{-1}\right) - d_j\mathtt{C}_{\nu_g}\left(e^{i\theta_0}(1+\epsilon z_j)\right)\right]\\ &= \frac{-1}{2\pi}\sum_{j=1}^m\left[c_je^{-i\theta_0}(1+\epsilon\overline{z_j})\langle g, (\mathcal{K}-e^{i\theta_0}(1+\epsilon z_j))^{-1}g\rangle + d_j\langle(\mathcal{K}-e^{i\theta_0}(1+\epsilon z_j))^{-1}g, \mathcal{K}^*g\rangle\right].\end{aligned} \tag{9.3.27}$$

It follows that we can compute the convolution $[K_\epsilon * \nu_g](\theta_0)$ by evaluating the resolvent at the m points $\{e^{i\theta_0}(1+\epsilon z_j)\}_{j=1}^m$. We use rational kernels because they allow us to compute smoothed approximations of the spectral measure by applying the resolvent operator to functions.

However, for (9.3.27) to be a good approximation of ν_g, we must carefully select the points z_j and the coefficients $\{c_j, d_j\}$ in (9.3.26). In particular, we would like $\{K_\epsilon\}$ to be an mth order kernel. First, we define $\zeta_j(\epsilon)$ by the relationship $1+\epsilon\zeta_j(\epsilon) = (1+\epsilon\overline{z_j})^{-1}$ and use Cauchy's Residue Theorem to show that for any integer $n \geq 1$,

$$\begin{aligned}\int_{[-\pi,\pi]_{\mathrm{per}}} K_\epsilon(-\theta)e^{in\theta}\,d\theta &= \frac{1}{2\pi i}\int_{\mathbb{T}}\left[\sum_{j=1}^m \frac{c_j}{\lambda-(1+\epsilon\overline{z_j})^{-1}} - \sum_{j=1}^m \frac{d_j}{\lambda-(1+\epsilon z_j)}\right]\lambda^n\,d\lambda\\ &= \sum_{j=1}^m c_j(1+\epsilon\overline{z_j})^{-n} = \left(\sum_{j=1}^m c_j\right) + \sum_{k=1}^n \epsilon^k\binom{n}{k}\sum_{j=1}^m c_j\zeta_j(\epsilon)^k.\end{aligned}$$

It follows that condition (9.3.20) in Lemma 9.3.6 holds if

$$\begin{pmatrix} 1 & \dots & 1\\ \zeta_1(\epsilon) & \dots & \zeta_m(\epsilon)\\ \vdots & \ddots & \vdots\\ \zeta_1(\epsilon)^{m-1} & \dots & \zeta_m(\epsilon)^{m-1}\end{pmatrix}\begin{pmatrix} c_1(\epsilon)\\ c_2(\epsilon)\\ \vdots\\ c_m(\epsilon)\end{pmatrix} = \begin{pmatrix}1\\0\\ \vdots\\ 0\end{pmatrix}. \tag{9.3.28}$$

Note also that, if this holds, the coefficients $c_j = c_j(\epsilon)$ remain bounded as $\epsilon \downarrow 0$. To ensure that the decay condition in (9.3.21) is satisfied, let $\omega = (e^{-i\theta}-1)/\epsilon$. The kernel in (9.3.26) can then be re-written as

$$K_\epsilon(\theta) = \frac{\epsilon^{-1}e^{-i\theta}}{2\pi}\sum_{j=1}^m\left[\frac{c_j}{\omega-\zeta_j(\epsilon)} - \frac{d_j}{\omega - z_j}\right]. \tag{9.3.29}$$

Therefore, we have

$$\begin{aligned}\omega K_\epsilon(\theta) &= \frac{\epsilon^{-1}e^{-i\theta}}{2\pi}\sum_{j=1}^m\left[c_j + \frac{c_j\zeta_j(\epsilon)}{\omega-\zeta_j(\epsilon)} - d_j - \frac{d_j z_j}{\omega-z_j}\right]\\ &= \frac{\epsilon^{-1}e^{-i\theta}}{2\pi}\sum_{j=1}^m(c_j-d_j) + \frac{\epsilon^{-1}e^{-i\theta}}{2\pi}\sum_{j=1}^m\left[\frac{c_j\zeta_j(\epsilon)}{\omega-\zeta_j(\epsilon)} - \frac{d_j z_j}{\omega-z_j}\right].\end{aligned}$$

By repeating the same argument m times, we arrive at

$$\omega^m K_\epsilon(\theta) = \frac{\epsilon^{-1}e^{-i\theta}}{2\pi}\left[\sum_{k=0}^{m-1}\omega^{m-1-k}\sum_{j=1}^m(c_j\zeta_j(\epsilon)^k - d_j z_j^k) + \sum_{j=1}^m\left(\frac{c_j\zeta_j(\epsilon)^m}{\omega-\zeta_j(\epsilon)} - \frac{d_j z_j^m}{\omega-z_j}\right)\right]. \tag{9.3.30}$$

This means that we should select the d_k's so that

$$\begin{pmatrix} 1 & \dots & 1\\ z_1 & \dots & z_m\\ \vdots & \ddots & \vdots\\ z_1^{m-1} & \dots & z_m^{m-1}\end{pmatrix}\begin{pmatrix} d_1\\ d_2\\ \vdots\\ d_m\end{pmatrix} = \begin{pmatrix}1\\0\\ \vdots\\ 0\end{pmatrix}. \tag{9.3.31}$$

We conclude that if the coefficients $\{c_j\}_{j=1}^m$ and $\{d_j\}_{j=1}^m$ satisfy (9.3.28) and (9.3.31), respectively, then

$$\sum_{k=0}^{m-1}\omega^{m-1-k}\sum_{j=1}^m(c_j\zeta_j(\epsilon)^k - d_j z_j^k) = 0, \qquad \left|\sum_{j=1}^m\left(\frac{c_j\zeta_j(\epsilon)^m}{\omega-\zeta_j(\epsilon)} - \frac{d_j z_j^m}{\omega-z_j}\right)\right| \lesssim |\omega|^{-1}.$$

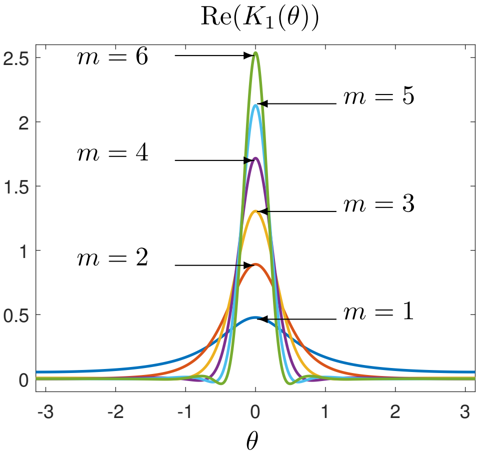


Figure 9.2: The mth-order kernels (9.3.26) constructed with the choice (9.3.32) for $\epsilon = 1$ and $1 \leq m \leq 6$.

m	$\{d_1, \dots, d_{\lceil m/2 \rceil}\}$	$\{c_1(\epsilon), \dots, c_{\lceil m/2 \rceil}(\epsilon)\}$, $\epsilon = 0.1$
2	$\left\{\frac{1-3i}{2}\right\}$	$\left\{\frac{3+10i}{6}\right\}$
3	$\{-2-i, 5\}$	$\left\{\frac{-202+79i}{80}, \frac{121}{20}\right\}$
4	$\left\{\frac{-39+65i}{24}, \frac{17-85i}{8}\right\}$	$\left\{\frac{-1165710-2944643i}{750000}, \frac{513570+3570527i}{250000}\right\}$
5	$\left\{\frac{15+10i}{4}, \frac{-39-13i}{2}, \frac{65}{2}\right\}$	$\left\{\frac{4052283-1460282i}{648000}, \frac{-2393157+486551i}{81000}, \frac{190333}{4000}\right\}$
6	$\left\{\frac{725-1015i}{192}, \frac{-2775+6475i}{192}, \frac{1073-7511i}{96}\right\}$	$\left\{\frac{24883929805+81589072062i}{8067360000}, \frac{-19967590755-93596942182i}{1613472000}, \frac{7898770397+102424504746i}{806736000}\right\}$

Table 9.1: Coefficients in the rational kernels in (9.3.26) for $1 \leq m \leq 6$, the choice (9.3.32), and $\epsilon = 0.1$. We give the first $\lceil m/2 \rceil$ coefficients as $c_{m+1-j} = \overline{c_j}$ and $d_{m+1-j} = \overline{d_j}$.

By (9.3.30), this means that $|\omega|^m\,|K_\epsilon(\theta)| \lesssim \epsilon^{-1}|\omega|^{-1}$. Moreover, since $|K_\epsilon(\theta)| \lesssim \epsilon^{-1}$ we see that $|K_\epsilon(\theta)| \lesssim \min\{\epsilon^{-1}, \epsilon^m|\theta|^{-(m+1)}\} \lesssim \epsilon^m(\epsilon + |\theta|)^{-(m+1)}$. Using Lemma 9.3.6, we have proved the following proposition.

Proposition 9.3.7. *Let $\{z_j\}_{j=1}^m$ be distinct points with positive real part and let K_ϵ be given by* (9.3.26)*. Then, $\{K_\epsilon\}$ is an mth order kernel for $[-\pi,\pi]_{\mathrm{per}}$ if the coefficients $\{c_j\}_{j=1}^m$ and $\{d_j\}_{j=1}^m$ satisfy* (9.3.28) *and* (9.3.31)*, respectively.*

The choice of rational kernel

We are free to choose the points $\{z_j\}_{j=1}^m$ in (9.3.7) subject to $\mathrm{Re}(z_j) > 0$, after which the linear systems (9.3.28) and (9.3.31) provide suitable $\{c_j\}_{j=1}^m$ (dependent on ϵ) and $\{d_j\}_{j=1}^m$ (independent of ϵ). As a natural extension of the Poisson kernel in (9.3.5), we select the points $\{z_j\}_{j=1}^m$ as

$$z_j = 1 + \left(\frac{2j}{m+1} - 1\right)i, \qquad 1 \leq j \leq m. \tag{9.3.32}$$

The kernels that we have developed are typically not real-valued. Since ν_g is a probability measure and hence real-valued, we often gain better accuracy for a particular ϵ by considering the kernel $\mathrm{Re}(K_\epsilon)$, and this is what we do throughout this chapter. The convolution with $\mathrm{Re}(K_\epsilon)$ can be computed by taking the real part of the right-hand side of (9.3.27). The first six kernels with the choice $\epsilon = 1$ are shown in Figure 9.2. The exact coefficients $\{c_j, d_j\}_{j=1}^m$ for $\epsilon = 0.1$ are shown in Table 9.1 for the first six kernels.

An algorithm for evaluating a smoothed spectral measure

To evaluate $[K_\epsilon * \nu_g]$ at a single point $\theta_0 \in [-\pi, \pi]_{\text{per}}$, we use the setup of ResDMD so that we can obtain rigorous a posteriori error bounds on the computed resolvents, allowing us to adaptively select the dictionary size N_K based on the smoothing parameter ϵ. Since $\mathcal{K}$ is an isometry, we only need to compute $(\mathcal{K} - z)^{-1}$ for $|z| > 1 = \|\mathcal{K}\|$ and so we can achieve our goal with standard Galerkin truncations of $\mathcal{K}$.

Theorem 9.3.8. *Suppose that $\mathcal{K}$ is an isometry and $\lambda \in \mathbb{C}$ with $|\lambda| > 1$. Let $\psi_1, \psi_2, \dots$ be a dictionary of observables and $V_{N_K} = \operatorname{span}\{\psi_1, \dots, \psi_{N_K}\}$, so that $\cup_{N_K \in \mathbb{N}} V_{N_K}$ is dense in $L^2(\Omega, \omega)$. Then, for any sequence of observables $g_{N_K} \in V_{N_K}$ such that $\lim_{N_K \to \infty} g_{N_K} = g \in L^2(\Omega, \omega)$,*

$$\lim_{N_K \to \infty} \left(P_{V_{N_K}} \mathcal{K} P_{V_{N_K}} - \lambda I_{N_K}\right)^{-1} g_{N_K} = (\mathcal{K} - \lambda)^{-1} g,$$

where $P_{V_{N_K}}$ is the orthogonal projection operator onto V_{N_K} and I_{N_K} is the $N_K \times N_K$ identity matrix.

Exercise: Prove Theorem 9.3.8 using a Neumann series argument.

We now apply Theorem 9.3.8 to evaluate $[K_\epsilon * \nu_g]$ at θ_0. Recall from (9.3.27) that there are two types of inner products to compute: (i) $\langle g, (\mathcal{K} - \lambda)^{-1} g\rangle$ and (ii) $\langle (\mathcal{K} - \lambda)^{-1} g, \mathcal{K}^* g\rangle$ for some observable g. We form a sequence of observables $g_{N_K} \in V_{N_K}$ by setting $g_{N_K} = P_{V_{N_K}} g$, which can be approximately computed from snapshot data as

$$\tilde{g}_{N_K} = \sum_{j=1}^{N_K} \boldsymbol{a}_j \psi_j, \qquad \boldsymbol{a} = (\Psi_X^* W \Psi_X)^{-1} \Psi_X^* W \begin{bmatrix} g(\boldsymbol{x}_0^{(1)}) \\ \vdots \\ g(\boldsymbol{x}_0^{(M)}) \end{bmatrix} \in \mathbb{C}^{N_K}. \tag{9.3.33}$$

Under suitable conditions, such as those already discussed in Section 9.2.3, $\lim_{M \to \infty} \tilde{g}_{N_K} = g_{N_K}$. Since

$$\left(P_{V_{N_K}} \mathcal{K} P_{V_{N_K}} - \lambda I_{N_K}\right)^{-1} g_{N_K} = \lim_{M \to \infty} \sum_{j=1}^{N_K} \left[(\Psi_X^* W \Psi_Y - \lambda \Psi_X^* W \Psi_X)^{-1} \Psi_X^* W \Psi_X \boldsymbol{a}\right]_j \psi_j,$$

it follows that our two types of inner products satisfy

$$\left\langle g_{N_K}, \left(P_{V_{N_K}} \mathcal{K} P_{V_{N_K}} - \lambda I_{N_K}\right)^{-1} g_{N_K} \right\rangle = \lim_{M \to \infty} \overline{\boldsymbol{a}^* \Psi_X^* W \Psi_X (\Psi_X^* W \Psi_Y - \lambda(\Psi_X^* W \Psi_X))^{-1} \Psi_X^* W \Psi_X \boldsymbol{a}}, \tag{9.3.34}$$

$$\left\langle \left(P_{V_{N_K}} \mathcal{K} P_{V_{N_K}} - \lambda I_{N_K}\right)^{-1} g_{N_K}, \mathcal{K}^* g_{N_K} \right\rangle = \lim_{M \to \infty} \boldsymbol{a}^* \Psi_X^* W \Psi_Y (\Psi_X^* W \Psi_Y - \lambda \Psi_X^* W \Psi_X)^{-1} \Psi_X^* W \Psi_X \boldsymbol{a}. \tag{9.3.35}$$

For a given value of M, the right-hand side of (9.3.34) and (9.3.35) can then be substituted into (9.3.27) to evaluate $[K_\epsilon * \nu_g](\theta_0)$. Often we can estimate the error between these computed inner products and the limiting value as $M \to \infty$ by comparing the computations for different M or by using a priori knowledge of the convergence rates. N_K can be adaptively chosen (by approximating the error in the large data limit and adding observables to the dictionary if required) so that the left-hand sides of (9.3.34) and (9.3.35) approximate the inner products $\langle g, (\mathcal{K} - \lambda)^{-1} g\rangle$ and $\langle (\mathcal{K} - \lambda)^{-1} g, \mathcal{K}^* g\rangle$, respectively, to a desired accuracy. Thus, for a given smoothing parameter ϵ, we have a principled way of selecting (a) the sample size M and (b) the truncation size N_K to ensure that our approximations of the inner products in (9.3.27) are accurate.

Algorithm 6 A computational framework for evaluating an approximate spectral measure with respect to $g \in L^2(\Omega, \omega)$ at $\{\theta_k\}_{k=1}^P \subset [-\pi, \pi]_{\mathrm{per}}$ of an isometry $\mathcal{K}$ using snapshot data.

```
Input: Snapshot data $\{\boldsymbol{x}^{(j)}, \boldsymbol{y}^{(j)}\}_{j=1}^M$ (such that $\boldsymbol{y}^{(j)} = F(\boldsymbol{x}^{(j)})$), quadrature weights $\{w_j\}_{j=1}^M$, a dictionary of observables $\{\psi_j\}_{j=1}^{N_K}$, $m \in \mathbb{N}$, smoothing parameter $0 < \epsilon < 1$ (accuracy goal is $\epsilon^m$), distinct points $\{z_j\}_{j=1}^m \subset \mathbb{C}$ with $\mathrm{Re}(z_j) > 0$ (recommended choice is (9.3.32)), and evaluation points $\{\theta_k\}_{k=1}^P \subset [-\pi, \pi]_{\mathrm{per}}$.
Solve (9.3.28) and (9.3.31) for $c_1(\epsilon), \dots, c_m(\epsilon) \in \mathbb{C}$ and $d_1, \dots, d_m \in \mathbb{C}$, respectively.
Compute $\Psi_X^* W \Psi_X$ and $\Psi_X^* W \Psi_Y$, where $\Psi_X$ and $\Psi_Y$ are given in (9.2.3).
Compute a generalised Schur decomposition of $\Psi_X^* W \Psi_Y$ and $\Psi_X^* W \Psi_X$, i.e., $\Psi_X^* W \Psi_Y = QSZ^*$ and $\Psi_X^* W \Psi_X = QTZ^*$, where $Q, Z$ are unitary and $S, T$ are upper triangular.
Compute $\boldsymbol{a}$ in (9.3.33) and $v_1 = TZ^*\boldsymbol{a}$, $v_2 = T^*Q^*\boldsymbol{a}$, and $v_3 = S^*Q^*\boldsymbol{a}$.
For $k = 1, \dots, P$
    Compute $I_j = (S - e^{i\theta_k}(1 + \epsilon z_j)T)^{-1} v_1$ for $1 \leq j \leq m$.
    Compute $\nu_g^\epsilon(\theta_k) = \frac{-1}{2\pi} \sum_{j=1}^m \mathrm{Re}\big[c_j(\epsilon) e^{-i\theta_k}(1 + \epsilon \overline{z_j})(I_j^* v_2) + d_j(v_3^* I_j)\big]$ .
end for
Output: Values of the approximate spectral measure, i.e., $\{\nu_g^\epsilon(\theta_k)\}_{k=1}^P$.
```

In general, the cost of point evaluation of $[K_\epsilon * \nu_g]$ using these formulas is $\mathcal{O}(N_K^3)$ operations as it requires m solutions of $N_K \times N_K$ dense linear systems.

To evaluate $[K_\epsilon * \nu_g]$ at $\theta_1, \dots, \theta_P \in [-\pi, \pi]_{\mathrm{per}}$, one can be more computationally efficient than independently computing each of the inner products in (9.3.34) and (9.3.35) for each θ_k for $1 \leq k \leq P$. Instead, one can compute a generalised Schur decomposition and use it to speed up the evaluation. Let $\Psi_X^* W \Psi_Y = QSZ^*$ and $\Psi_X^* W \Psi_X = QTZ^*$ be a generalised Schur decomposition, where Q and Z are unitary matrices and S and T are upper-triangular matrices. With this decomposition in hand,

$$\boldsymbol{a}^* \Psi_X^* W \Psi_X (\Psi_X^* W \Psi_Y - \lambda(\Psi_X^* W \Psi_X))^{-1} \Psi_X^* W \Psi_X \boldsymbol{a} = \boldsymbol{a}^* QT(S - \lambda T)^{-1} TZ^* \boldsymbol{a},$$
$$\boldsymbol{a}^* \Psi_X^* W \Psi_Y (\Psi_X^* W \Psi_Y - \lambda \Psi_X^* W \Psi_X)^{-1} \Psi_X^* W \Psi_X \boldsymbol{a} = \boldsymbol{a}^* QS(S - \lambda T)^{-1} TZ^* \boldsymbol{a}.$$

Now, after computing the generalised Schur decomposition costing $\mathcal{O}(N_K^3)$ operations, each evaluation requires solving $N_K \times N_K$ upper-triangular linear systems in $\mathcal{O}(N_K^2)$ operations. Additional computational savings can be realised if one is willing to do each evaluation at $\theta_1, \dots, \theta_P$ in parallel. We summarise the evaluation scheme in Algorithm 6.

9.4 Numerical Examples

9.4.1 Non-linear pendulum ($d = 2$)

Let $\boldsymbol{x} = (x_1, x_2) = (\theta, \dot{\theta})$ be the state variables governed by the following equations of motion:

$$\dot{x_1} = x_2, \quad \dot{x_2} = -\sin(x_1), \qquad \text{with} \quad \Omega = [-\pi, \pi]_{\mathrm{per}} \times \mathbb{R}, \tag{9.4.1}$$

where ω is the standard Lebesgue measure. We consider the corresponding discrete-time dynamical system by sampling with a time-step $\Delta_t = 0.5$. For the dictionary of observables $\psi_1, \dots, \psi_{N_K}$, we use a hyperbolic

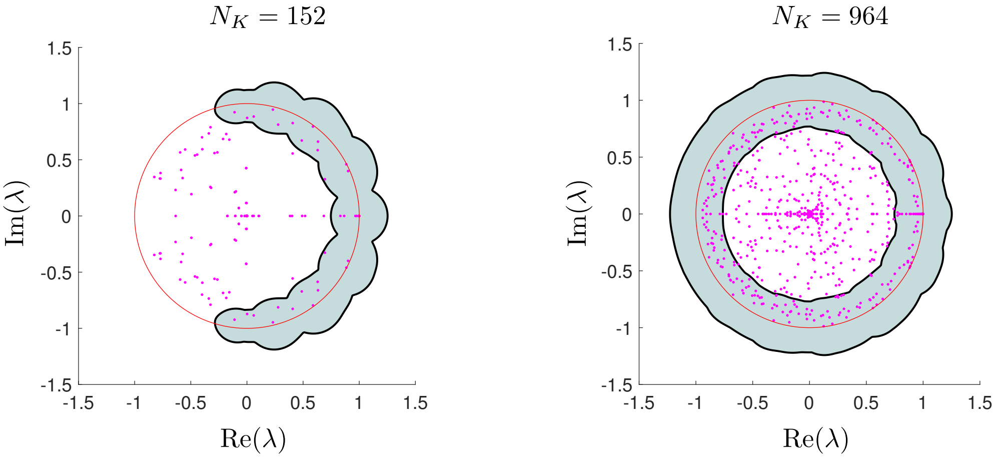


Figure 9.3: The ϵ-pseudospectra for the non-linear pendulum and $\epsilon = 0.25$ (shaded region) computed using Algorithm 2 with discretisation sizes N_K. Discretisation sizes correspond to a hyperbolic cross approximation. The computed ϵ-pseudospectra converge as $N_K \to \infty$. The unit circle (red line) is shown with the EDMD eigenvalues (magenta dots), many of which are spurious. ResDMD removes spurious eigenvalues by computing pseudospectra.

cross approximation with the standard Fourier basis (in $x_1 \in [-\pi, \pi]_{\text{per}}$) and Hermite functions (in $x_2 \in \mathbb{R}$). We use the trapezoidal quadrature rule to compute $\Psi_X^* W \Psi_X$, $\Psi_X^* W \Psi_Y$, and $\Psi_Y^* W \Psi_Y$, where Ψ_X and Ψ_Y are given in (9.2.3). To simulate the collection of trajectory data, we compute trajectories starting at each initial condition using the `ode45` command in MATLAB. We stress that we only use `ode45` as a black-box integrator - all of our algorithms in this chapter are purely data driven.

The system is Hamiltonian and hence the Koopman operator is unitary. It follows that $\mathrm{Sp}_\epsilon(\mathcal{K}) = \{z \in \mathbb{C} : \mathrm{dist}(z, \mathbb{T}) \leq \epsilon\}$. Figure 9.3 shows the computed pseudospectrum for $\epsilon = 0.25$. The algorithm uses a discretisation size of N_K to compute a set guaranteed to be inside $\mathrm{Sp}_\epsilon(\mathcal{K})$ (i.e., no spectral pollution), that also converges as $N_K \to \infty$. We also show the corresponding EDMD eigenvalues. Some of these EDMD eigenvalues are reliable, but the majority are not, demonstrating severe spectral pollution. Note that this spectral pollution has nothing to do with any stability issues, but instead is due to the discretisation of the infinite-dimensional operator $\mathcal{K}$ by a finite matrix. Using the ϵ-pseudospectrum for different ϵ, we can detect exactly which of these eigenvalues are reliable. Using Algorithm 2 and $N_K = 964$, we also compute some approximate eigenfunctions corresponding to $\lambda = \exp(0.4932i)$, $\lambda = \exp(0.9765i)$, $\lambda = \exp(1.4452i)$, and $\lambda = \exp(1.8951i)$ (see Figure 9.4). As λ moves further from 1, we typically see increased oscillations in the approximate eigenfunctions.

The Koopman operator associated with (9.4.1) has a continuous spectrum. We now compute spectral measures from autocorrelations using Algorithm 5 and consider a corresponding discrete-time system by sampling (9.4.1) with a time-step of $\Delta_t = 1$. We look at the following observable that involves non-trivial dynamics in each coordinate:

$$g(x_1, x_2) = C(1 + i\sin(x_1))(1 - \sqrt{2}x_2)e^{-x_2^2/2},$$

where $C \approx 0.24466788518668$ is a normalisation constant. Figure 9.5 shows high resolution approximations of the spectral measure ν_g for $N = 100$ and $N = 1000$. The spectral measure is purely continuous (no atoms) away from $\theta = 0$, consistent with the general theory of integrable Hamiltonian systems with one degree of freedom [Mez20]. Note that the constant function 1 is not in $L^2([0, 2\pi]_{\text{per}} \times \mathbb{R})$ and hence cannot be an eigenfunction. We confirmed this by using Theorem 9.3.4 for larger N and observing that the

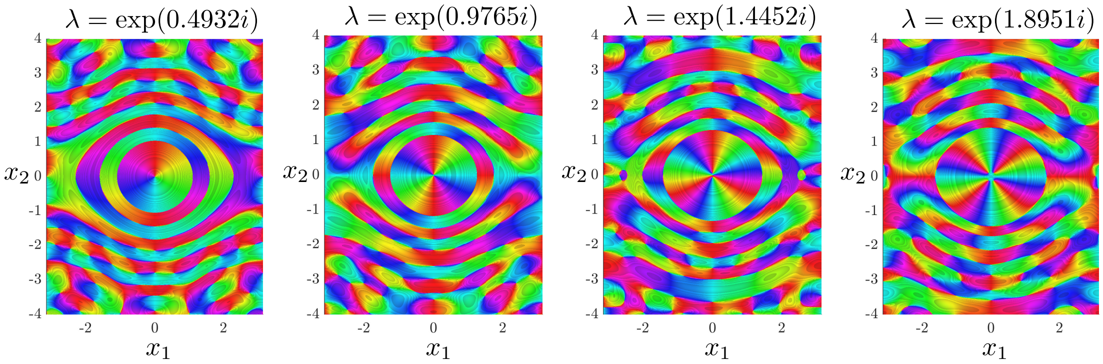


Figure 9.4: The approximate eigenfunctions of the non-linear pendulum visualised as phase portraits, where the color illustrates the complex argument of the eigenfunction. We also plot lines of constant modulus as shadowed steps. All of these approximate eigenfunctions have residuals at most $\epsilon = 0.05$ as judged by (9.2.5), which can be made smaller by increasing N_K.

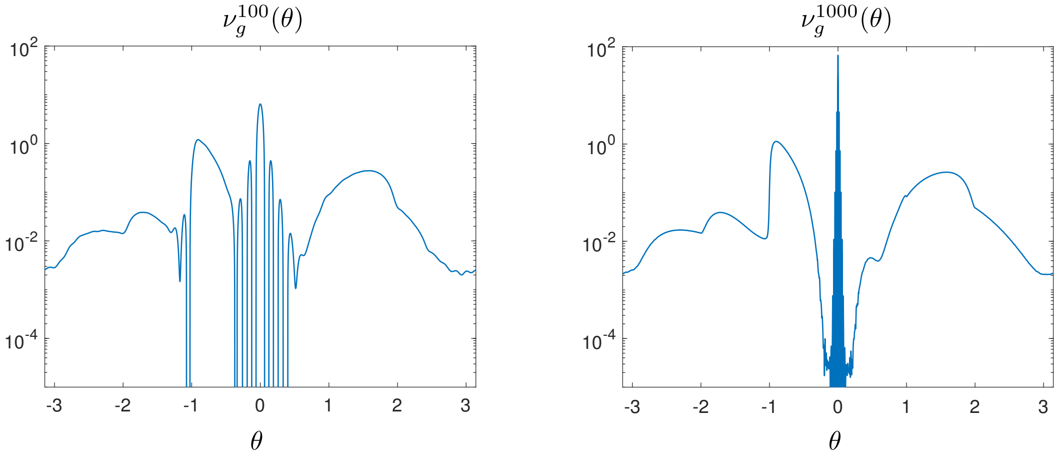


Figure 9.5: Computed spectral measure using Algorithm 5 with (9.3.18) for the non-linear pendulum.

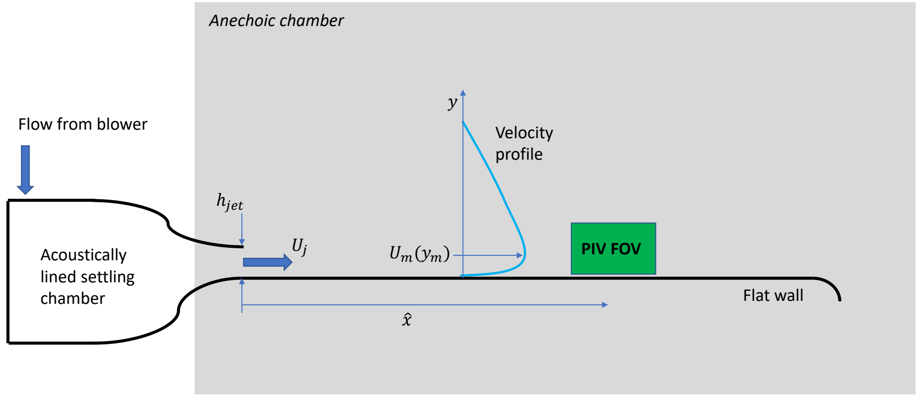


Figure 9.6: Schematic of the TR-PIV experiments conducted in the Wall Jet Wind Tunnel of Virginia Tech.

peak at $\theta = 0$ seen in Figure 9.5 does not grow as fast as $\propto N$.

9.4.2 Turbulent wall-jet boundary layer flow ($d = 102,300$)

We now consider a turbulent wall-jet boundary layer flow [Ger15, GAE$^+$00, KRH$^+$19]. For this example, we assess the performance of the ResDMD algorithm on a set of time-resolved (TR) particle image velocimetry (PIV) data. We consider the boundary layer generated by a thin jet ($h_{jet} = 12.7$mm) injecting air onto a smooth flat wall. This case is challenging for regular DMD approaches due to multiple turbulent scales expected within the boundary layer. This section demonstrates the use of ResDMD for a high Reynolds number, turbulent, complex flow field.

Experiments using TR-PIV are performed at the Wall Jet Wind Tunnel of Virginia Tech, as schematically shown in Figure 9.6. A two-dimensional two-component TR-PIV system is used to capture the wall-jet flow and the streamwise origin of the field-of-view (FOV) is $\hat{x}$ =1282.7mm downstream of the wall-jet nozzle. We use a jet velocity of U_j =50m/s, corresponding to a jet Reynolds number of $\mathrm{Re}_{jet} = h_{jet}U_j/\nu = 63.5 \times 10^3$. The length and height of the FOV is approximately 75mm × 45mm, and the spatial resolution of the velocity vector field is 0.25mm. The high-speed cameras are operated in a double frame mode, with a rate of 12,000 frame pairs per second, resulting in a fine temporal resolution of 0.083ms.

The flow consists of two main regions. Within the region bounded by the wall and the peak in the velocity profile, the flow exhibits the properties of a zero pressure gradient turbulent boundary layer. Above this fluid portion, the flow is dominated by a two-dimensional shear layer consisting of rather large, energetic flow structures. While the peak in the velocity profile is $y_m \approx$ 18mm from the wall in our case, the overall thickness of the wall-jet flow is on the order of 200mm. Clearly, the PIV experiments must compromise between a good spatial resolution or capturing the entire flow field. In our case, the FOV was not tall enough to capture the entire wall-jet flow field. For this reason, the standard DMD algorithm under-predicts the energies corresponding to the shear-layer portion of the wall-jet flow as the corresponding length scales fall outside of the limits of the FOV.

We collect snapshot data of the velocity field from *two separate realisations of the experiment*. We use

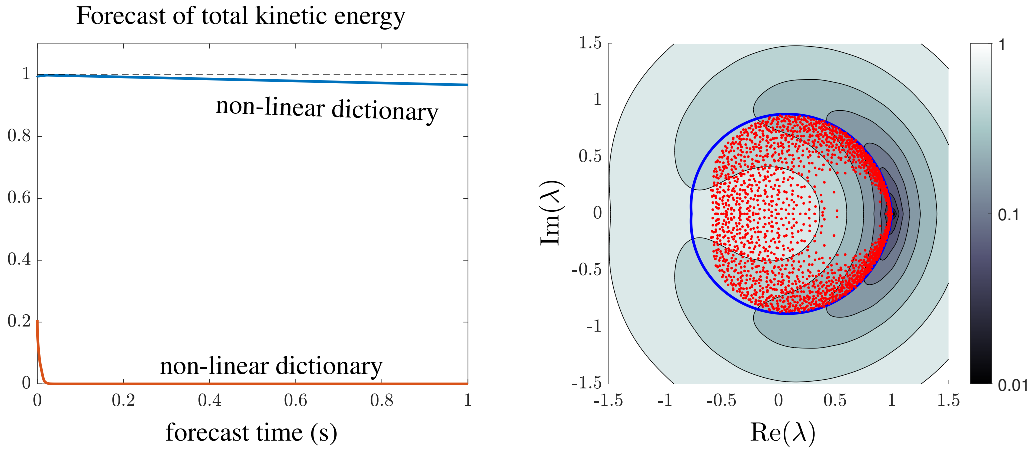


Figure 9.7: Left: Forecast of total kinetic energy (normalised by the time average of the kinetic energy), averaged over the 12000 initial conditions. Values closer to 1 correspond to better predictions. Right: Pseudospectral contours computed using Algorithm 2 for the wall-jet example, using a non-linear dictionary. The eigenvalues of the finite Galerkin matrix K_{EDMD} are shown as red dots. The shape of the contours reflect the transient modes. The blue curve corresponds to a fit $r = \exp(-c|\theta|)$ of these contours and the boundary of the eigenvalues, and represents successive powers of modes.

the first experiment to generate data $\{\tilde{\boldsymbol{x}}^{(j)}, \tilde{\boldsymbol{y}}^{(j)}\}_{j=1}^{M'}$ with $M' = 2000$, corresponding to 121 boundary layer turnover times. This data is used to select our dictionary of functions. We use the second experiment to generate data $\{\hat{\boldsymbol{x}}^{(j)}, \hat{\boldsymbol{y}}^{(j)}\}_{j=1}^{M''}$ with $M'' = 12000$ (a single trajectory of one second of physical flow time and 728 boundary layer turnover times), which we use to generate the ResDMD matrices, as outlined in Section 9.2.5. To demonstrate the need for non-linear functions in our dictionary, we compute the Koopman mode decomposition of the total kinetic energy of the domain. Using this decomposition, we compute forecasts of the total energy from a given initial condition of the system. Figure 9.7 (left) shows the results, where we average over the 12000 initial conditions in the data set and normalise by the true time-averaged kinetic energy. We use Algorithms 3 and 4 with $N_K = 2000$, which we refer to as a linear dictionary and non-linear dictionary, respectively. The importance of including non-linear functions in the dictionary is clear, and corresponds to a much better approximation of $\mathcal{K}$'s spectral content near 0. For the rest of this section, we therefore only use the non-linear dictionary. Figure 9.7 (right) shows pseudospectral contours computed using Algorithm 2. The contours appear to be centered around a curve of the form $r = \exp(-c|\theta|)$ (shown as blue in the plot), corresponding to successive powers of transient modes. This is reflected in the eigenvalues of the finite $N_K \times N_K$ Galerkin matrix K_{EDMD}, shown as red dots, some of which correspond to spectral pollution. The eigenvalues of non-normal matrices can be severely unstable to perturbations, particularly for large N_K, so we checked the computation of the eigenvalues of K_{EDMD} by comparing to extended precision and predict a bound of $\approx 10^{-10}$ on the error in Figure 9.7 (right).

To investigate the Koopman modes, we compute the ResDMD Koopman mode decomposition corresponding to Algorithm 1 with the error tolerance $\epsilon = 0.5$ to get rid of the most severe spectral pollution. The total number of modes used is 656. Figure 9.8 illustrates a range of Koopman modes which are long-lasting (left-hand column) and transient (right-hand column). Due to residual measures, we are able to accurately select physical transient modes. Within each figure, the arrows dictate the unsteady fluid structure (computed from the Koopman modes of the velocity fields), with the magnitude of the arrow indicating the local flow speed, and the colourbar denotes the Koopman mode of the velocity magnitude. The corresponding approximate eigenvalues, λ, and residual bound are provided for each mode.

The modes in the left column of Figure 9.8 illustrate the range of rolling eddies within the boundary layer, with the smaller structures containing less energy than the largest structures. Interestingly, the third mode in the left column resembles the shape of ejection-like motions within the boundary layer flow ($y/y_m < 1$) while larger-scale structures above the boundary layer ($y/y_m > 1$) are also visible. This may be interpreted as a non-linear interaction in the turbulent flow field, which is efficiently captured using the ResDMD algorithm. The transient modes in the right column of Figure 9.8 show a richer structure. Based on our analysis, we interpret these modes as transient, short-lived behaviour of turbulence. The uppermost panel may be seen as the shear layer traveling over the boundary layer ($y/y_m > 1$), with the following panel potentially seen as the breakdown of this transient structure into smaller structures. The third panel may be seen as an interaction between an ejection-type vortex and the shear layer, note the ejection-like shape of negative contours below $y/y_m = 1.5$ with a height-invariant positive island of contour at $y/y_m \approx 1.75$. Finally the bottom-most panel could be seen as a flow uplift out of the boundary layer and further turbulent streaks with counter-rotating properties.

9.4.3 MD simulation of the Adenylate Kinase enzyme ($d = 20,046$)

Molecular dynamics (MD) analyses the movement of atoms and molecules by numerically solving Newton's equations of motion for a system of interacting particles. Energies and forces between particles are typically computed using potentials. MD is arguably one of the most robust approaches for simulating macromolecular dynamics, in large part due to the availability of full atomistic detail [DDG+12]. Recently, DMD-type and Koopman techniques are making an impact in MD [NKPH+14, KNK+18, SP15, SP13]. For example, [KSM20] applies kernel EDMD to the positions of the carbon atoms in *n*-butane ($d = 12$) and shows that the EDMD eigenfunctions parametrise a dihedral angle that controls key dynamics.

Here, we study trajectory data from the dynamics of Adenylate Kinase (ADK), which is an enzyme (see Figure 9.9) that catalyses important phosphate reactions in cellular biology. ADK is a common benchmark enzyme in MD and consists of 3341 atoms split into 214 residues (specific monomers that can be thought of as parts). The trajectory data comes from an all-atom equilibrium simulation for 1.004×10^{-6}s, with a so-called CHARMM force field, that is produced on PSC Anton [SDS+09] and publicly available [BFG+]. The data consists of a single trajectory of the positions of all atoms as ADK moves. To make the system Hamiltonian, we append the data with approximations of the velocities computed using centered finite differences. This leads to $d = 6 \times 3341 = 20046$. We sample the trajectory data every 240×10^{-12}s so that $M = 4184$.

To apply the kernelized version of Algorithm 6, we subselect $M' = 2000$ initial conditions from the trajectory data. We select $N_K = 1000$ EDMD eigenfunctions and append the dictionary with the four observables of interest that are discussed below. Accuracy of the corresponding matrices in (9.2.6) is verified by comparing to smaller M'' and also computing pseudospectra with Algorithm 2.

ADK has three parts of its molecule called CORE, LID, and NMP (see Figure 9.9 (left)). The LID and NMP domains move around the stable CORE. By computing root-mean-square-fluctuations, we select the most mobile residue from the LID and NMP domains. These residues have canonical dihedral angles (ϕ, ψ) defined on the backbone atoms that determine the overall shape of the residue. Figure 9.9 (middle, left) shows the spectral measures with respect to these dihedral angles (where we have subtracted the mean

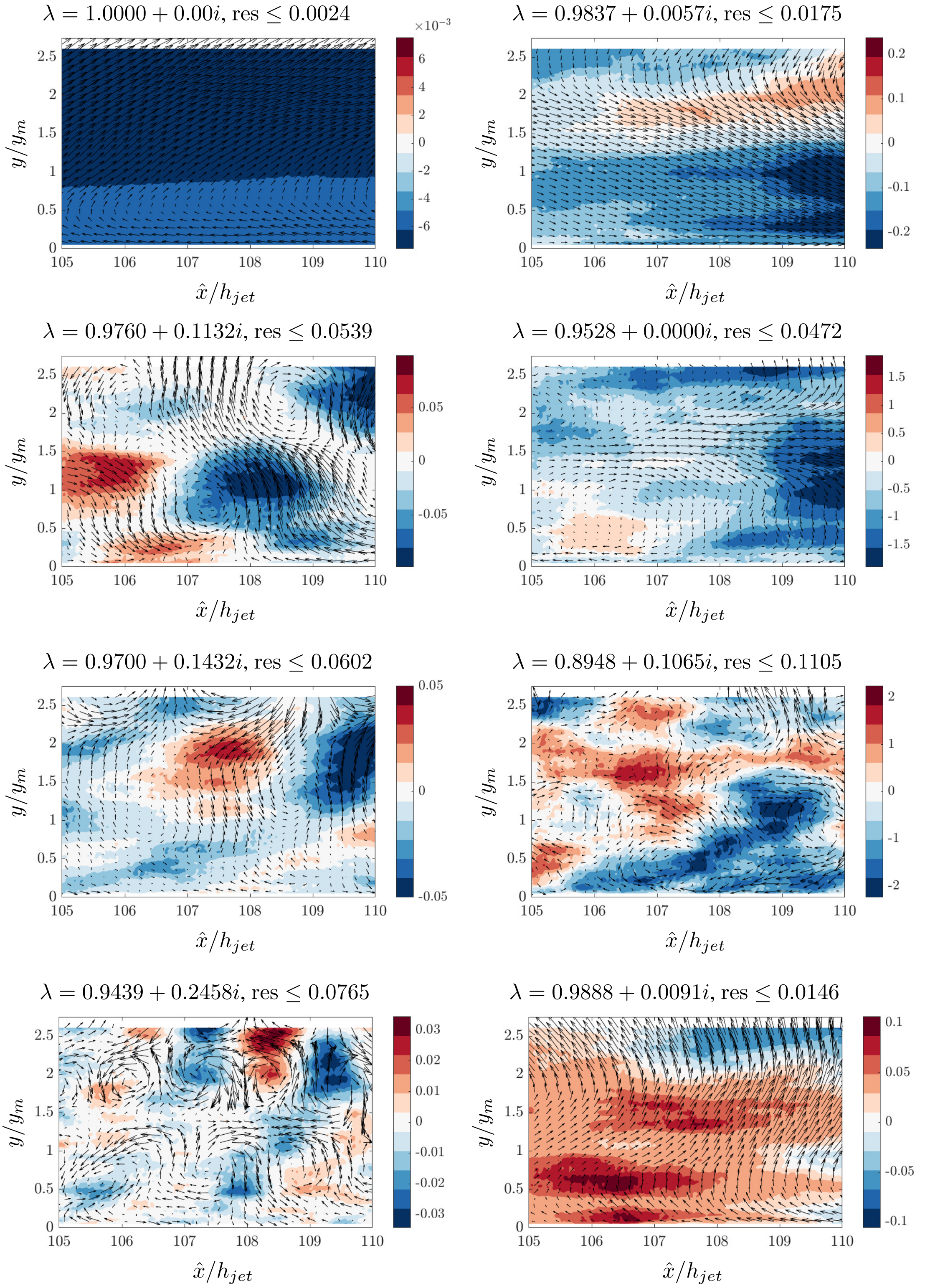


Figure 9.8: Left: A range of long-lasting modes from the ResDMD Koopman mode decomposition. Right: A range of transient modes from the ResDMD Koopman mode decomposition.The arrows dictate the unsteady fluid structure (computed from the Koopman modes of the velocity fields), with the magnitude of the arrow indicating the local flow speed, and the colourbar denotes the Koopman mode of the velocity magnitude.

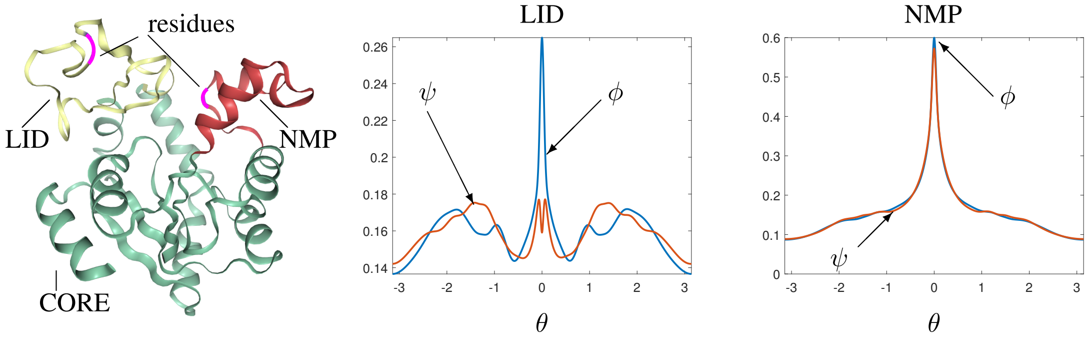


Figure 9.9: Left: Structure of ADK which has three domains: CORE (green), LID (yellow) and NMP (red). Middle and right: Spectral measures with respect to the dihedral angles of the selected residues.

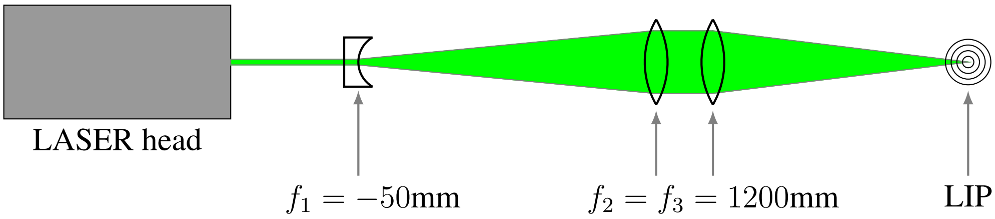


Figure 9.10: Schematic diagram of the laser beam setup to generate the laser-induced plasma.

angle value) for both selected residues. These spectral measures are computed using the sixth order rational kernel with $\epsilon = 0.1$ (see Table 9.1). The computed spectral measures are verified with higher order kernels and smaller ϵ, as well as comparison with a polynomial kernel $\mathcal{S}$. The spectral measures for the angles in the LID residue are much broader than for the NMP residue. This hints at a more complicated dynamical interaction and may have biological consequences.

9.4.4 Shockwave propagation (d varies)

The computation of residuals allows an efficient compression of the Koopman mode decomposition by discarding modes associated with spectral pollution. As our final example, we demonstrate the use of ResDMD on an acoustic example where the sound source of interest exhibits highly non-linear properties.

We investigate a near-ideal acoustic monopole source that is generated using the laser optical setup illustrated in Figure 9.10. When a high-energy laser beam is focused into a point, the air ionizes and plasma is generated due to the extremely high electromagnetic energy density (on the order of 10^{12}W/m^2). As a result of the sudden deposit of energy, the volume of air undergoes a sudden expansion that generates a shockwave. The initial propagation characteristics can be modeled using von Neumann's point strong explosion theory, which was originally developed for nuclear explosion modeling. For our ResDMD analysis, we use laser-induced plasma (LIP) sound signature data measured using an 1/8inch, Bruel & Kjaer (B&K) type 4138 microphone operated using a B&K Nexus module [SDB$^+$22]. The data from the microphone is acquired using an NI-6358 module at a sampling rate of f_s =1.25MS/s. With this apparatus, we can resolve the high-frequency nature of the LIP up to 100kHz.

The important acoustic characteristic of the LIP is that it has a short time period of initial supersonic propagation speed, which are shown as Schlieren images taken over a 15μs window in Figure 9.11. When observed from the far field, this initial supersonic propagation is observed as a non-linear characteristic

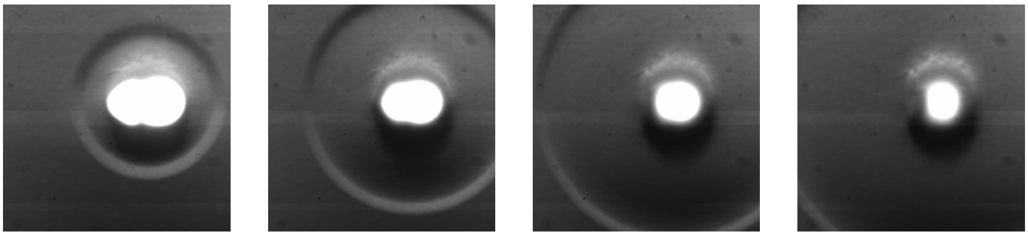

a) $t = 5\ \mu$s b) $t = 10\ \mu$s c) $t = 15\ \mu$s d) $t = 20\ \mu$s

Figure 9.11: Schlieren images of the initial laser-induced plasma illustrating the shock wave formation and propagation.

despite that the wavespeed is supersonic only in a short radius around the source, namely, until about 3–4mm from the optical focal point. During the experiments, 65 realisations of LIP are captured using microphones. Each realisation of LIP is then gated in time such that only the direct propagation path of the LIP remains in the signal. We use this gated data for our ResDMD analysis.

For a positive integer d, we take the state at time n to be

$$\boldsymbol{x}_n = \begin{pmatrix} p(n) & p(n-1) & \cdots & p(n-d+1) \end{pmatrix}^\top \in \mathbb{R}^d,$$

where p is acoustic pressure. This corresponds to time-delay embedding, which is a popular method for DMD-type algorithms. There is a further interpretation of d when we make future state predictions using the Koopman mode decomposition. The value of d corresponds to the initial time interval that we use to make future state prediction. This is shown as vertical dashed lines in the plots below.

We split the data into three parts. The first 10 realisations of LIP correspond to $\{\tilde{\boldsymbol{x}}^{(j)}, \tilde{\boldsymbol{y}}^{(j)}\}_{j=1}^{M'}$ and are used to train the dictionary. The next 50 realisations correspond to $\{\hat{\boldsymbol{x}}^{(j)}, \hat{\boldsymbol{y}}^{(j)}\}_{j=1}^{M''}$, and are used to construct the ResDMD matrices. The final 5 realisations are used to test the resulting Koopman mode decomposition. We consider two choices of dictionary. The first is a linear dictionary computed using Algorithm 3. The second is the union of the linear dictionary and the dictionary computed using Algorithm 4 with $N_K = 200$. We refer to this combined dictionary as the non-linear dictionary.

Figure 9.12 (left) shows the results of the Koopman mode decomposition, applied to the first realisation of the experiment in the test set, with $d = 10$. Namely, we approximate the state as

$$\begin{aligned} \boldsymbol{x}_n &\approx K_{\mathrm{EDMD}}^n \Psi(\boldsymbol{x}_0) V \left[V^{-1} (\sqrt{W}\Psi_X)^\dagger \sqrt{W} \begin{pmatrix} \hat{\boldsymbol{x}}^{(1)} & \cdots & \hat{\boldsymbol{x}}^{(M'')} \end{pmatrix}^\top \right] \\ &= \Psi(\boldsymbol{x}_0) V \Lambda^n \left[V^{-1} (\sqrt{W}\Psi_X)^\dagger \sqrt{W} \begin{pmatrix} \hat{\boldsymbol{x}}^{(1)} & \cdots & \hat{\boldsymbol{x}}^{(M'')} \end{pmatrix}^\top \right]. \end{aligned} \tag{9.4.2}$$

In particular, we test the Koopman mode decomposition on unseen data corresponding to the test set. The values of p to the left of the vertical dashed line correspond to $\boldsymbol{x}_0$. It is clear that the non-linear dictionary does a much better job of representing the non-linear behaviour of the system. While the linear dictionary can predict the positive pressure peak, it fails to predict both the magnitude and shape of the negative peak, and it also fails to capture the smaller, high-frequency oscillations following the fist two large oscillation. These discrepancies between the linear and non-linear dictionary-based results also pinpoint where non-linearity in the signal relies. In other words, the non-linear signature of the pressure wave relies in the negative portion of the wave. Figure 9.12 (right) plots the relative mean squared error (RMSE) averaged over the test set for different values of d. The non-linear dictionary allows an average relative L^2 error of around 6% for $d = 15$.

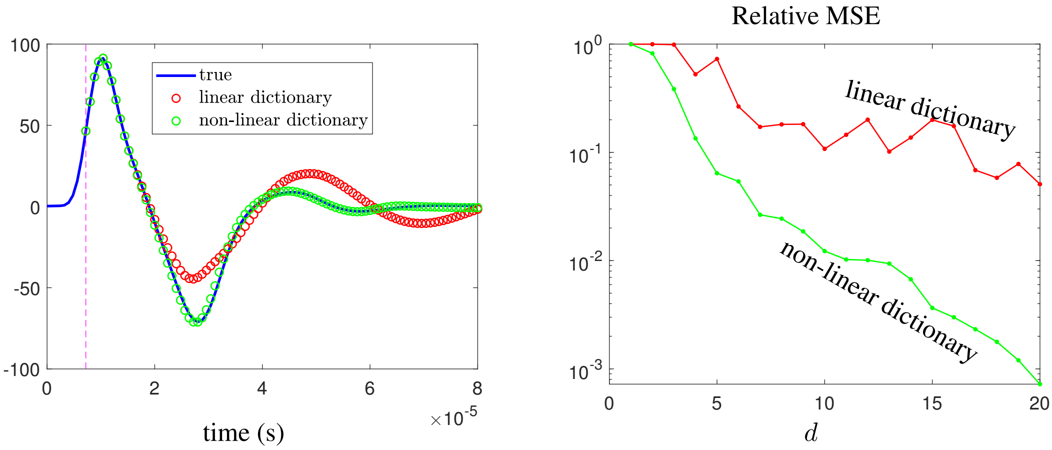


Figure 9.12: Left: Prediction using (9.4.2) on the first experiment in the test set. The values of p to the left of the vertical dashed line correspond to $\boldsymbol{x}_0$. Right: Relative mean squared error (RMSE) averaged over the test set for different values of d.

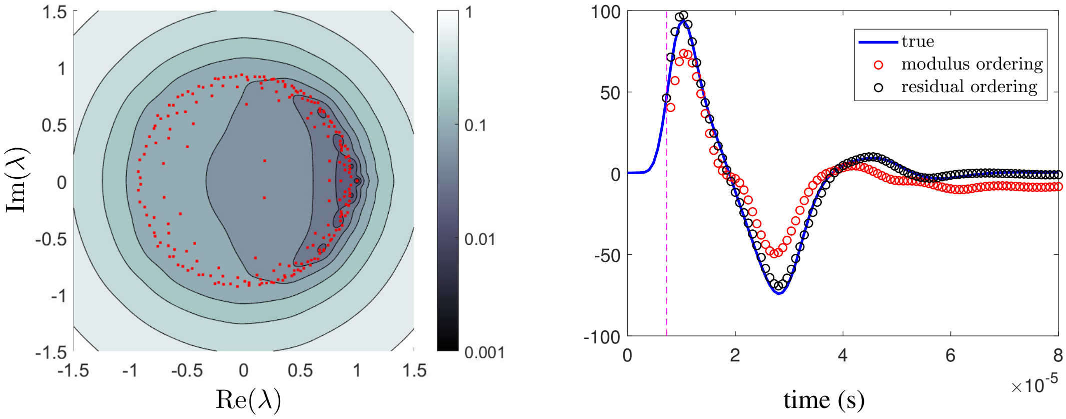


Figure 9.13: Left: Pseudospectral contours, computed using Algorithm 2 with the non-linear dictionary and $d = 10$. The eigenvalues of K_{EDMD} are shown as red dots. Right: Prediction on the first experiment in the test set. The values of p to the left of the vertical dashed line correspond to $\boldsymbol{x}_0$. For each type of ordering, we use 40 modes.

Figure 9.13 (left) shows the corresponding pseudospectral contours, computed using Algorithm 2 with $d = 10$. We can use ResDMD to compress the representation of the dynamics, by ordering the Koopman eigenvalues λ_j, eigenfunctions g_j, and modes according to their (relative) residual $\text{res}(\lambda_j, g_j)$ (defined in (9.2.5)). For a prescribed value of N', we can produce a Koopman mode decomposition of the N' eigenfunctions with the smallest residual. In Figure 9.13 (right), we compare this to a compression based on the modulus of the eigenvalues using 40 modes in each expansion. It is clear that ordering the eigenvalues by their residuals gives a much better compression of the dynamics. To investigate this further, Figure 9.14 shows the error curves of the two different compressions for various dictionary sizes and choices of d. This suggests ResDMD may be effective in the construction of reduced order models.

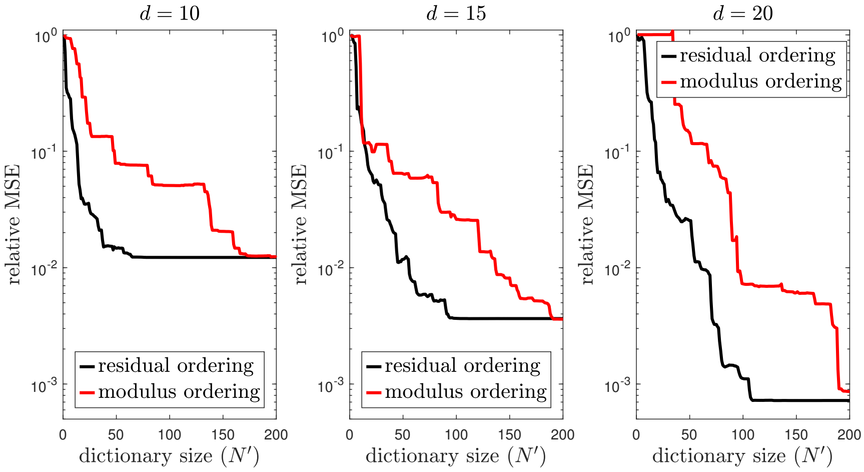


Figure 9.14: RMSE averaged over the test set for two types of compression. ‘residual ordering’ (black curves) corresponds to ordering approximate eigenvalues based on their residual. ‘modulus ordering’ (red curves) corresponds to ordering approximate eigenvalues based on their modulus.

Chapter 10

Beyond Spectral Problems

In this final chapter we discuss work beyond spectral problems. The framework provided by this course encompasses a vast number of areas in computational mathematics, including numerical solution of PDEs, the foundations of AI and neural networks, optimisation, and computer-assisted proofs. Much of the work below is ongoing. If you are interested in working on any of these projects, please do get in touch!

10.1 Numerical Solution of PDEs

10.1.1 Semigroups

The discussion in this section is based on [Col22]. Given a linear operator A on an *infinite-dimensional* separable Hilbert space $\mathcal{H}$, can we numerically compute, with *error control*, the solution of

$$u'(t) = Au(t) \text{ for } t \geq 0, \quad \text{with initial condition} \quad u(0) = u_0 \in \mathcal{H}? \tag{10.1.1}$$

The desired solution is written as $u(t) = \exp(tA)u_0$ and made rigorous through the theory of semigroups [Paz12, ABHN01]. Equation (10.1.1) arises in numerous applications and there exist many numerical methods designed to approximate $u(t)$, including but not limited to: contour methods [WT07, TW14, SST03, HHT08]; domain truncation and absorbing boundary conditions (e.g., when A represents a differential operator on an unbounded domain) [EM77, AES03, Tsy98, Sze04, A$^+$08]; Galerkin methods [Lub08b, KLY19, LL20]; Krylov methods [Gri12, GG13b, LS13]; rational approximations [CLPT93, BT79, Pal93]; and series expansions, splitting methods, and exponential integrators [Hig05, Lub08a, IKS18, HO10, MQ02, AMH11].

The majority of convergence results in the literature concern specific cases of the operator A. If A is unbounded with domain $\mathcal{D}(A)$, it is common to assume regularity on u_0 (e.g., $u_0 \in \mathcal{D}(A^\nu)$ for some $\nu > 0$) to obtain asymptotic rates of convergence. Instead, we consider the following question:

> **Q.1:** *Can we compute semigroups with error control? That is, does there exist an algorithm that when given a generator A of a strongly continuous semigroup on $\mathcal{H}$, time $t > 0$, arbitrary $u_0 \in \mathcal{H}$ and error tolerance $\epsilon > 0$, computes an approximation of* $\exp(tA)u_0$ *to accuracy ϵ in $\mathcal{H}$?*

A prototypical example of (10.1.1) is when A is a partial differential operator (PDO) on some domain. For unbounded domains, such as $\mathcal{H} = L^2(\mathbb{R}^d)$, this is a well-studied yet notoriously difficult challenge.

The methods listed above yield invaluable insight into many computational issues. However, the answer to Q.1 for unbounded domains remains largely unknown in the general case. For example, only in specific cases does one know how to truncate the domain and set appropriate boundary conditions. Even if one can prove the *existence* of suitable truncations and boundary conditions, there may not be an *algorithm* that does this (the original results of [EM79] reflect this). Moreover, difficulties are intensified in the case of irregular geometry or variable coefficients. A motivating example is the Schrödinger equation,

$$i\frac{\partial u}{\partial t} = -\Delta u + Vu, \quad u_0 \in L^2(\mathbb{R}^d), \tag{10.1.2}$$

In light of this, a second question we consider is the following:

> **Q.2:** *For $\mathcal{H} = L^2(\mathbb{R}^d)$, is there a large class of PDO generators A (more general than* (10.1.2)*) on the unbounded domain $\mathbb{R}^d$ where the answer to Q.1 is yes?*

To have any hope of answering this question, we need the semigroup to be well-behaved, i.e., (10.1.1) to be well-posed. The following are standard [Paz12].

Definition 10.1.1. *A strongly continuous semigroup (C_0-semigroup) on a Banach space X is a map $S : [0,\infty) \to \mathcal{L}(X)$ such that*

1. *$S(0) = I$*
2. *$S(s+t) = S(s)S(t), \quad \forall s,t \geq 0$*
3. *$S(t)$ converges strongly to I as $t \downarrow 0$ (i.e., $\lim_{t\downarrow 0} S(t)x = x$, for all $x \in X$).*

The infinitesimal generator A of S is defined as $Ax = \lim_{t\downarrow 0} \frac{1}{t}(S(t)-I)x$, where $\mathcal{D}(A)$ is all $x \in X$ such that the limit exists, and we write $S(t) = \exp(tA)$.

Definition 10.1.2. *A continuous function $u : [0,\infty) \to X$ is a*

1. *Classical solution of the Cauchy problem* (10.1.1) *if it is continuously differentiable, $u(t) \in \mathcal{D}(A)$ for all $t \geq 0$, and* (10.1.1) *is satisfied,*
2. *Mild solution of the Cauchy problem* (10.1.1) *if for all $t \geq 0$,*

$$\int_0^t u(s)ds \in \mathcal{D}(A) \quad \text{and} \quad A\int_0^t u(s)ds = u(t) - u_0.$$

The following theorem tells us precisely when a unique mild solution exists.

Theorem 10.1.3 (Theorem 3.1.12 of [ABHN01]). *Let A be a closed operator acting on the Banach space X. The following assertions are equivalent:*

(a) *For any $u_0 \in X$, there exists a unique mild solution of* (10.1.1).

(b) *$\rho(A) \neq \emptyset$ and for every $u_0 \in \mathcal{D}(A)$, there is a unique classical solution of* (10.1.1).

(c) *The operator A generates a C_0-semigroup S.*

When these conditions hold, the solution is given by $u(t) = S(t)u_0 = \exp(tA)u_0$.

The Hille–Yosida theorem tells us precisely when an operator A generates a strongly continuous semigroup, and thus, by Theorem 10.1.3, when (10.1.1) admits a unique solution.

Theorem 10.1.4 (Hille–Yosida theorem)**.** *A closed operator A on X generates a C_0-semigroup if and only if A is densely defined and there exists $\omega \in \mathbb{R}$, $M > 0$ with*

(1) $\{\lambda \in \mathbb{R} : \lambda > \omega\} \subset \rho(A)$.

(2) For all $\lambda > \omega$ and $n \in \mathbb{N}$, $(\lambda - \omega)^n \|R(\lambda, A)^n\| \leq M$.

Under these conditions, $\|\exp(tA)\| \leq M \exp(\omega t)$ and if $\mathrm{Re}(\lambda) > \omega$ then $\lambda \in \rho(A)$ with

$$\|R(\lambda, A)^n\| \leq \frac{M}{(\mathrm{Re}(\lambda) - \omega)^n}, \quad \textit{for all } n \in \mathbb{N}. \tag{10.1.3}$$

Example results

First, consider the canonical separable Hilbert space $l^2(\mathbb{N})$ of square summable sequences, using $e_1, e_2, \dots$ to denote the canonical orthonormal basis. Let $\mathcal{C}(l^2(\mathbb{N}))$ denote the set of closed and densely defined linear operators A such that $\mathrm{span}\{e_n : n \in \mathbb{N}\}$ forms a core of A and its adjoint A^*. If $A \in \mathcal{C}(l^2(\mathbb{N}))$, then we can associate an infinite matrix with the operator A through the inner products $A_{j,k} = \langle Ae_k, e_j \rangle$. Given $(A, u_0) \in \mathcal{C}(l^2(\mathbb{N})) \times l^2(\mathbb{N})$, we consider the following evaluation functions (recall that this is the readable input to our algorithm), denoted by Λ_1, which include the case of inexact input:

- Matrix evaluation functions: $\{f^{(1)}_{j,k,m}, f^{(2)}_{j,k,m} : j, k, m \in \mathbb{N}\}$ such that

$$|f^{(1)}_{j,k,m}(A) - \langle Ae_k, e_j \rangle| \leq 2^{-m}, \quad |f^{(2)}_{j,k,m}(A) - \langle Ae_k, Ae_j \rangle| \leq 2^{-m}, \quad \forall j, k, m \in \mathbb{N}.$$

- Coefficient and norm evaluation functions: $\{f_{j,m} : j \in \mathbb{N} \cup \{0\}, m \in \mathbb{N}\}$ such that

$$|f_{0,m}(u_0) - \langle u_0, u_0 \rangle| \leq 2^{-m}, \quad |f_{j,m}(u_0) - \langle u_0, e_j \rangle| \leq 2^{-m}, \quad \forall j, m \in \mathbb{N}. \tag{10.1.4}$$

Let Ω_{C_0} denote the set of triples (A, u_0, t) where $A \in \mathcal{C}(l^2(\mathbb{N}))$ generates a strongly continuous semigroup, $u_0 \in l^2(\mathbb{N})$ and $t > 0$. We define the set of evaluation functions for such triples to be $\Lambda_{C_0} = \Lambda_1 \cup \{M(A), \omega(A)\}$, where $M = M(A)$ and $\omega = \omega(A)$ are constants satisfying the conditions in Theorem 10.1.4 for the generator A. Finally, we consider the problem function $\Xi_{C_0} : \Omega_{C_0} \to l^2(\mathbb{N}), (A, u_0, t) \mapsto \exp(tA)u_0$. In other words, the computation of the solution of (10.1.1). The following theorem provides a positive answer to Q.1.

Theorem 10.1.5 (C_0-semigroups on $l^2(\mathbb{N})$ computed with error control)**.** *There exists an algorithm Γ using Λ_{C_0} such that for any $\epsilon > 0$ and $(A, u_0, t) \in \Omega_{C_0}$,*

$$\|\Gamma(A, u_0, t, \epsilon) - \exp(tA)u_0\| \leq \epsilon.$$

It follows that $\{\Xi_{C_0}, \Omega_{C_0}\} \in \Delta_1^A$.

We now extend the above to PDOs. Consider the closure, denoted by A, of the initial operator

$$[\tilde{A}u](x) = \sum_{k \in \mathbb{Z}^d_{\geq 0}, |k| \leq N} a_k(x) \partial^k u(x), \quad \mathcal{D}(\tilde{A}) = \{u \text{ smooth with compact support}\}. \tag{10.1.5}$$

We use multi-index notation with $|k| = \max\{|k_1|, ..., |k_d|\}$ and $\partial^k = \partial^{k_1}_{x_1} \partial^{k_2}_{x_2} ... \partial^{k_d}_{x_d}$. We assume that $\tilde{A}$ is closable and that the coefficients $a_k(x)$ are complex-valued measurable functions on $\mathbb{R}^d$. For dimension d and $r > 0$, consider the space

$$\mathcal{A}_r = \{f \in \mathrm{Meas}([-r, r]^d) : \|f\|_\infty + \mathrm{TV}_{[-r,r]^d}(f) < \infty\},$$

where $\mathrm{Meas}([-r,r]^d)$ denotes the set of measurable functions on the hypercube $[-r,r]^d$ and $\mathrm{TV}_{[-r,r]^d}$ the total variation norm in the sense of Hardy and Krause [Nie92]. This space becomes a Banach algebra when equipped with the norm [BT89]

$$\|f\|_{\mathcal{A}_r} := \left\|f|_{[-r,r]^d}\right\|_\infty + (3^d+1)\mathrm{TV}_{[-r,r]^d}(f).$$

We let Ω_{PDE} be all such (A,u_0,t) with $u_0 \in L^2(\mathbb{R}^d)$ and $t>0$, for which A generates a strongly continuous semigroup on $L^2(\mathbb{R}^d)$ and the following hold:

(1) The set of smooth, compactly supported functions forms a core of A and A^*.

(2) At most polynomial growth: There exist positive constants C_k and integers B_k such that almost everywhere on $\mathbb{R}^d$, $|a_k(x)| \leq C_k(1+|x|^{2B_k})$.

(3) Locally bounded total variation: For all $r>0$, $u_0|_{[-r,r]^d}, a_k|_{[-r,r]^d} \in \mathcal{A}_r$.

These assumptions are very mild as the class of functions with locally bounded variation includes discontinuous functions and functions with arbitrary wild oscillations at infinity. For input $(A,u_0,t)\in\Omega_{\mathrm{PDE}}$, we define Λ_{PDE} as the set of evaluation functions (where ranges of indices have been suppressed for notational convenience):

(a) Pointwise coefficient evaluations: $\{S_{k,q,m}\}$ such that for all $m\in\mathbb{N}$,

$$|S_{k,q,m}(A)-a_k(q)| \leq 2^{-m}, \quad \forall q\in\mathbb{Q}^d.$$

(b) Pointwise initial condition evaluations: $\{S_{q,m}\}$ such that for all $m\in\mathbb{N}$,

$$|S_{q,m}(u_0)-u_0(q)| \leq 2^{-m}, \quad \forall q\in\mathbb{Q}^d.$$

(c) Bounds on growth and total variation: $\{C_k,B_k\}$ such that the bound in (2) holds and positive sequences $\{b_n\}_{n\in\mathbb{N}}$ and $\{c_n\}_{n\in\mathbb{N}}$ such that for all $n\in\mathbb{N}$,

$$\max_{|k|\leq N}\|a_k\|_{\mathcal{A}_n} \leq b_n, \quad \|u_0\|_{\mathcal{A}_n}\leq c_n.$$

(d) Decay of initial condition: A positive sequence $\{d_n\}_{n\in\mathbb{N}}$, such that

$$\|u_0|_{[-n,n]^d}-u_0\|_{L^2(\mathbb{R}^d)} \leq d_n, \quad \lim_{n\to\infty} d_n = 0,$$

together with constants $M=M(A)>0$ and $\omega=\omega(A)>0$ satisfying the conditions in Theorem 10.1.4 for the generator A. We consider the problem function $\Xi_{\mathrm{PDE}}:\Omega_{\mathrm{PDE}}\to L^2(\mathbb{R}^d)$, $(A,u_0,t)\mapsto\exp(tA)u_0$. In other words, the computation of the solution of (10.1.1) for PDOs A on $L^2(\mathbb{R}^d)$. The following theorem provides a positive answer to Q.2.

Theorem 10.1.6 (PDO C_0-semigroups on $L^2(\mathbb{R}^d)$ computed with error control)**.** *There exists an algorithm* Γ *using* Λ_{PDE} *such that for any* $\epsilon>0$ *and* $(A,u_0,t)\in\Omega_{\mathrm{PDE}}$*,*

$$\|\Gamma(A,u_0,t,\epsilon)-\exp(tA)u_0\|\leq\epsilon.$$

It follows that $\{\Xi_{\mathrm{PDE}},\Omega_{\mathrm{PDE}}\}\in\Delta_1^A$.

Exercise: Assuming Theorem 10.1.5, prove Theorem 10.1.6 using the techniques of Chapter 3.

The idea of the method

The solution of (10.1.1) is, at least formally, the Bromwich complex contour integral

$$\exp(tA)u_0 = \left[\frac{-1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} e^{zt}(A-zI)^{-1}\,dz\right]u_0, \quad \text{for sufficiently large } \sigma\in\mathbb{R}, \tag{10.1.6}$$

and computing solutions of (10.1.1) is a special case of inverting an operator-valued Laplace transform. However, there are two challenges with using (10.1.6). First, the integrand need not decay (special cases when it does after a contour deformation include analytic semigroups). Second, how do we compute the inverses $(A-zI)^{-1}$? To overcome these challenges, our method combines a regularised functional calculus, suitable contour quadrature rules, and the adaptive computation of resolvents in infinite dimensions. We compute the resolvent in an adaptive manner, providing explicit error control.

Dealing with the operator A directly, as opposed to a truncation or discretisation, allows us to provide rigorous convergence results under quite general assumptions. In many problems, there is an additional practical benefit in that it is easier to bound the resolvent. In contrast, previous approaches to (10.1.1) are typically of the flavor "truncate-then-solve." A truncation/discretisation of A is adopted and methods for computing the exponential of a finite matrix are used. In rigorously answering Q.1, it is vital to adopt a "solve-then-discretise" approach.

10.1.2 Non-linear Schrödinger equations

The discussion in this subsection follows [BH20]. We consider the situation of a single particle described by a self-adjoint Schrödinger operator $H_0 = -\Delta + V : D(H_0) \subset L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ with static *pinning potential* V. Apart from the static pinning potential, we also allow the presence of an additional *control potential* V_{con} with time-dependent control function $u \in W^{1,1}_{\mathrm{pcw}}(0,T)$ (piecewise $W^{1,1}$). Thus, writing $V_{\mathrm{TD}}(t) := u(t)V_{\mathrm{con}}$ for the time-dependent potential, we cover time-dependent Schrödinger equations

The non-linear Schrödinger equations we consider are

$$\begin{aligned} &i\partial_t\psi(x,t) = H_0\psi(x,t) + V_{\mathrm{TD}}(x,t)\psi(x,t) + \nu F_\sigma(\psi(x,t)), \quad (x,t)\in\mathbb{R}\times(0,T)\\ &\qquad\psi(\cdot,0) = \varphi_0 \end{aligned} \tag{10.1.7}$$

with scattering length $\nu = 1$ and non-linearity $F_\sigma(\psi(x,t)) = |\psi(x,t)|^{\sigma-1}\psi(x,t)$ where we consider $\sigma = 3$ (cubic NLS) and $\sigma = 5$ (quintic NLS). The choice $\nu = 1$ yields a *defocussing* non-linearity and $\nu = -1$ a *focussing* one.

Numerical methods are often used to analyse if the solution of a NLS blows up in finite time or not [DS11]. While the solution to the quintic NLS in (10.1.7) exists for all times if the non-linearity is defocussing, this is no longer the case if (10.1.7) has a focussing quintic non-linearity. In greater generality, we study whether it is possible to numerically decide whether a solution to a NLS will blow up in finite time or not? We show that this is impossible in great generality.

Definition 10.1.7 (Initial state with controlled local boundedness and bounded variation and (CLBBV)). *Given an initial state* $\varphi_0 \in \mathrm{BV}_{\mathrm{loc}}(\mathbb{R}^d)$ *we say that* φ_0 *has* controlled local boundedness and bounded variation *(CLBBV) by* $\omega : \mathbb{N} \to \mathbb{N}$ *if for every* $R \in \mathbb{N}$ *then* $K = \omega(R)$ *is such that* $\|\varphi_0|_{\mathcal{C}_R(0)}\|_{L^\infty}, \mathrm{TV}(\varphi_0|_{\mathcal{C}_R(0)}) \le K$*, where* $\mathcal{C}_R(0)$ *is the closed cube of length* R *centered at zero.*[1]

[1] We emphasize that bounded total variation already implies a possibly weak L^∞ estimate by $\|\varphi_0|_{\mathcal{C}_R(0)}\|_{L^\infty} \le |\varphi_0(0)| + \mathrm{TV}(\varphi_0|_{\mathcal{C}_R(0)})$

Remark 10.1.8 (Input to the algorithms)**.** *We assume that*

$$\left\{(\varphi_0(x_k), V_{\mathrm{TD}}(x_k,t_j)) \,|\, \{x_k\}_{k\in\mathbb{N}}, \{t_j\}_{j\in\mathbb{N}} \textit{ are dense in } \mathbb{R}^d \textit{ and } [0,T], x_k, t_j \textit{ have coordinates} \in \mathbb{Q}\right\},$$

are accessible to the algorithm.

At least from a physics perspective, the most prominent example of a NLS with non-trivial blow-up dichotomy is the focussing ($\nu = -1$) cubic NLS

$$\begin{aligned} i\partial_t\psi(x,t) + \Delta\psi(x,t) &= \nu|\psi(x,t)|^2\psi(x,t), \quad (t,x) \in \mathbb{R}\times\mathbb{R}^3, \\ \psi(0,x) &= \varphi_0(x). \end{aligned} \tag{10.1.8}$$

Choose any fixed $C > 0$ and g as in Def. 10.1.7. Then, we define, for $\rho, \nu \geq 1$, the set of initial data as

$$\Omega_{\mathrm{BU}(1)} = \{\varphi_0 \in H^\rho_\nu(\mathbb{R}) \,|\, \|\varphi_0\|_{H^\rho_\nu(\mathbb{R})} \leq C \text{ and } \varphi_0 \text{ has CLBV by } g\}.$$

We consider the computational problem

$$\Xi_{\mathrm{BU}(1)} : \Omega_{\mathrm{BU}(1)} \ni \varphi_0 \mapsto \begin{cases} \text{Yes} & \text{if (10.1.8) blows up in finite time,} \\ \text{No} & \text{if (10.1.8) does not blows up in finite time} \end{cases} \in \mathcal{M}, \tag{10.1.9}$$

where $\mathcal{M} = \{\text{Yes}, \text{No}\} = \{1, 0\}$. Next, we consider the focussing ($\nu = -1$) mass-critical NLS with $\sigma = 1 + 4/d$, in particular,

$$\begin{aligned} i\partial_t\psi(x,t) + \Delta\psi(x,t) &= \nu|\psi(x,t)|^{\sigma-1}\psi(x,t), \quad (t,x) \in \mathbb{R}\times\mathbb{R}^d, \\ \psi(0,x) &= \varphi_0(x). \end{aligned} \tag{10.1.10}$$

Choose any fixed $C > 0$ and g as in Def. 10.1.7. Then, we define, for $\rho, \nu \geq 1$, the set of initial data as

$$\Omega_{\mathrm{BU}(2)} = \{\varphi_0 \in H^\rho_\nu(\mathbb{R}^d) \,|\, \|\varphi_0\|_{H^\rho_\nu(\mathbb{R}^d)} \leq C \text{ and } \varphi_0 \text{ has CLBV by } g\}.$$

We consider the computational problem

$$\Xi_{\mathrm{BU}(2)} : \Omega_{\mathrm{BU}(2)} \ni \varphi_0 \mapsto \begin{cases} \text{Yes} & \text{if (10.1.10) blows up in finite time,} \\ \text{No} & \text{if (10.1.10) does not blows up in finite time} \end{cases} \in \mathcal{M}. \tag{10.1.11}$$

Our main result on the computability of blow-ups is then Theorem 10.1.9.

Theorem 10.1.9 (Blow up cannot be decided, in fact not verified nor falsified)**.** *Consider the decision problems* $\{\Xi_{\mathrm{BU}(1)}, \Omega_{\mathrm{BU}(1)}\}$ *and* $\{\Xi_{\mathrm{BU}(2)}, \Omega_{\mathrm{BU}(2)}\}$ *defined in* (10.1.9) *and* (10.1.11) *concerning the blow up of the NLS. Then, there do not exist sequences of algorithms* $\{\Gamma^1_k\}$, $\{\Gamma^2_k\}$, *with* $\Gamma^1_k : \Omega_{\mathrm{BU}(1)} \to \mathcal{M}$ *and* $\Gamma^2_k : \Omega_{\mathrm{BU}(2)} \to \mathcal{M}$ *such that*

$$\lim_{k\to\infty} \Gamma^1_k(\varphi_0) = \Xi_{\mathrm{BU}(1)}(\varphi_0), \textit{ such that } \Gamma^1_k(\varphi_0) = \text{No} \Rightarrow \Xi_{\mathrm{BU}(1)}(\varphi_0) = \text{No},$$

$$\lim_{k\to\infty} \Gamma^2_k(\varphi_0) = \Xi_{\mathrm{BU}(2)}(\varphi_0), \textit{ such that } \Gamma^1_k(\varphi_0) = \text{Yes} \Rightarrow \Xi_{\mathrm{BU}(2)}(\varphi_0) = \text{Yes}.$$

These statements are universal independent of the computational model.

Let us consider norms $|f|_{L^2} := \|f\|_{L^2}$, $|f|_{H^1} := \|f\|^{\alpha}_{L^2}\|f\|^{1-\alpha}_{\dot{H}^1}$ for some fixed $\alpha \in (0,1)$, and $|f|_{\dot{H}^1} = \|f\|_{\dot{H}^1}$ a non-trivial function f. We then let $X \in \left\{L^2(\Omega), H^1(\Omega), \dot{H}^1(\Omega)\right\}$ and $\Omega \subset \mathbb{R}^d$ a domain. Let

$C > 0$ be given and let $\Omega_{\mathrm{BU}(X)}$ be the set of functions $v \in X \cap C(\Omega)$ with $|v|_X \leq C$ and v has controlled local bounded variation by h. We then consider the condition

$$|u_0|_X \leq |f|_X. \tag{10.1.12}$$

To define the computational problem we define for $f \in \Omega_{\mathrm{BU}(X)}$ the set

$$\begin{aligned}\Omega_{\mathrm{BU(X,f)}} &= \left\{u_0 \in \Omega_{\mathrm{BU}(X)}; |u_0|_X \neq |f|_X\right\}, \mathcal{M} = \{\text{No}, \text{Yes}\} = \{0,1\} \text{ and} \\ \Xi_{\mathrm{BU(X,f)}}(\varphi_0) &= \text{Does (10.1.12) hold?}.\end{aligned} \tag{10.1.13}$$

Proposition 10.1.10. *Given the above setup, we have that*

$$\{\Xi_{\mathrm{BU}(f,X)}, \Omega_{\mathrm{BU}(f,X)}, \mathcal{M}, \Lambda\} \notin \Sigma_1^G.$$

Proof. To show that $\{\Xi_{\mathrm{BU}(f,X)}, \Omega_{\mathrm{BU}(f,X)}, \mathcal{M}, \Lambda\} \notin \Sigma_1^G$ we argue by contradiction and assume the contrary. Let therefore $\{\Gamma_n\}$ be a sequence of general algorithms such that $\Gamma_n(\varphi_0) \to \Xi_{\mathrm{BU}(f,X)}(\varphi_0)$ as $n \to \infty$, and with $\Gamma_n(\varphi_0) = 1 \Rightarrow \Xi_{\mathrm{BU}(f,X)}(\varphi_0) = 1$. Let $\varphi_0 \in \Omega_{\mathrm{BU}(f,X)}$ denote a function satisfying (10.1.12) and note that, by the reasoning above $\Xi_{\mathrm{BU}(f,X)}(\varphi_0) = 1$. Thus, there is an $N \in \mathbb{N}$ such that $\Gamma_n(\varphi_0) = 1$ for all $n \geq N$. Choose any such $n \geq N$ and let $\mathcal{B} \subset \Omega$ be an open ball such that for all $f_j \in \Lambda_{\Gamma_n}(\varphi_0)$ we have $\omega_j \notin \mathcal{B}$. Choose a $\tilde{\varphi}_0 \in \Omega$ such that $\operatorname{supp}(\tilde{\varphi}_0) \subset \mathcal{B}$ and

$$|\tilde{\varphi}_0|_X > |f|_X. \tag{10.1.14}$$

Note that such a choice is easy to justify by using bump functions. Note that, by the choice of $\tilde{\varphi}_0$ we have that $f_j(\tilde{\varphi}_0) = f_j(\varphi_0) \quad \forall f_j \in \Lambda_{\Gamma_n}(\varphi_0)$. Hence, by assumption (iii) in (ii) in Definition 2.1.1 it follows that $1 = \Gamma_n(\varphi_0) = \Gamma_n(\tilde{\varphi}_0)$. However, by (10.1.14), it follows that $\Xi_{\mathrm{BU}(f,X)}(\tilde{\varphi}_0) = 0$, which contradicts that $\Gamma_n(\tilde{\varphi}_0) = 1 \Rightarrow \Xi_{\mathrm{BU}(f,X)}(\tilde{\varphi}_0) = 1$, and we have reached the desired contradiction. □

Proposition 10.1.11 (Mass critical NLS)**.** *Given the setup as in* (10.1.10)*, we have that*

$$\{\Xi_{\mathrm{BU}(2)}, \Omega_{\mathrm{BU}(2)}, \mathcal{M}, \Lambda\} \notin \Sigma_1^G.$$

Proof. The ground state soliton Q satisfying

$$-\Delta Q - Q|Q|^4 + Q = 0$$

for the $1d$-quintic NLS is known explicitly $Q(x) = \left(\frac{3}{\cosh^2(2x)}\right)^{1/4}$ and exists for all $d \geq 1$. For $d \geq 1$ and $\sigma = 1+4/d$, it is known [Dod15] that if $\|\varphi_0\|_{L^2} < \|Q\|_{L^2}$ then the solution to (10.1.10) exists globally and scatters whereas for $\|\varphi_0\|_{L^2} > \|Q\|_{L^2}$ there exist solutions that exist only for finite time. The statement then follows from Proposition 10.1.10. □

Showing that $\{\Xi_{\mathrm{BU}(2)}, \Omega_{\mathrm{BU}(2)}, \mathcal{M}, \Lambda\} \notin \Pi_1^G$ is in general more subtle. To see this, observe that by Sobolev's embedding in dimension one, we have $\|\varphi_0\|_{L^\infty} \leq \|\varphi_0\|_{H^1}$. This implies that if an algorithm samples a sufficiently large value of φ_0 it follows that $\|\varphi_0\|_{H^1}$ is large as well.

For our next proposition we consider a bump function

$$\chi_{\varepsilon,x_0}(x) := e^{1+\frac{\varepsilon^2}{\|x-x_0\|^2-\varepsilon^2}} \mathbb{1}_{B(x_0,\varepsilon)}(x).$$

We then have that

$$\|\chi_{\varepsilon,x_0}\|_{L^2} = \mathcal{O}(\varepsilon^d) \text{ and } \|\chi_{\varepsilon,x_0}\|_{\dot{H}^1} = \mathcal{O}(\varepsilon^{d-2}). \tag{10.1.15}$$

If we impose stronger conditions on X and the dimension, we obtain the following result:

Proposition 10.1.12. *For the setup as above, it follows that* $\{\Xi_{\mathrm{BU}(f,X)}, \Omega_{\mathrm{BU}(f,X)}, \mathcal{M}, \Lambda\} \notin \Pi_1^G$ *under the following conditions on the space* X *and the dimension* d *with open domain* $\Omega \subset \mathbb{R}^d$

- *If* $d = 1$ *and* $X \in \left\{L^2(\Omega) \cap C(\Omega), H^1(\Omega) \cap C(\Omega)\right\}$ *with* $\alpha < 1/2$.
- *If* $d = 2$ *and* $X \in \left\{L^2(\Omega) \cap C(\Omega), H^1(\Omega) \cap C(\Omega)\right\}$.
- $d \geq 3$.

Proof. We argue again by contradiction. Assuming the contrary, let $\{\Gamma_n\}$ be a sequence of general algorithms such that $\Gamma_n(\varphi_0) \to \Xi_{\mathrm{BU}(f,X)}(\varphi_0)$ as $n \to \infty$, and with $\Gamma_n(\varphi_0) = 0 \Rightarrow \Xi_{\mathrm{BU}(f,X)}(\varphi_0) = 0$. Let $\varphi_0 \in \Omega_{\mathrm{BU}(f,X)}$ be a function that does not satisfy (10.1.12). In this case $\Xi_{\mathrm{BU}(f,X)}(\varphi_0) = 0$ and hence there is an $N \in \mathbb{N}$ such that $\Gamma_n(\varphi_0) = 0$ for all $n \geq N$. Let ε be small enough such that $B(\tilde{\omega}_j, \varepsilon)$ are disjoint.

Choose any such n and choose $\tilde{\varphi}_0 := \sum \varphi_0(\omega_j)\chi_{\varepsilon,\omega_j}$ such that $\tilde{\varphi}_0$ interpolates φ_0 at the points $\tilde{\omega}_j$, where $f_j(\varphi_0) = \varphi_0(\tilde{\omega}_j)$ and $f_j \in \Lambda_{\Gamma_n}(\varphi_0)$. Let ε be sufficiently small, then by (10.1.15) it follows that $|\tilde{\varphi}_0|_X < |f|_X$. Then, as argued as above, we have $f_j(\tilde{\varphi}_0) = f_j(\varphi_0) \quad \forall f_j \in \Lambda_{\Gamma_n}(\tilde{\varphi}_0)$, and hence by by assumption (iii) in (ii) in Definition 2.1.1 it follows that $0 = \Gamma_n(\varphi_0) = \Gamma_n(\tilde{\varphi}_0)$. However, since $|\tilde{\varphi}_0|_X < |f|_X$ we have that $\Xi_{\mathrm{BU}(f,X)}(\tilde{\varphi}_0) = 1$, which contradicts that $\Gamma_n(\tilde{\varphi}_0) = 0 \Rightarrow \Xi_{\mathrm{BU}(f,X)}(\tilde{\varphi}_0) = 0$. □

We continue with our result on the cubic NLS:

Proposition 10.1.13. *Given the setup in* (10.1.8) *we have that*

$$\{\Xi_{\mathrm{BU}(1)}, \Omega_{\mathrm{BU}(1)}, \mathcal{M}, \Lambda\} \notin \Pi_1^G.$$

Proof. For (10.1.8) one has the following blow up dichotomy [HR08, HPR09]: Let $\varphi_0 \in H^1_1(\mathbb{R}^3)$ be an initial state to the focusing NLS (10.1.8) with ground state soliton Q satisfying

$$-\Delta Q - Q|Q|^2 + Q = 0.$$

- If $\|\varphi_0\|_{L^2} \|\nabla\varphi_0\|_{L^2} < \|Q\|_{L^2} \|\nabla Q\|_{L^2}$, then the solution to (10.1.8) exists globally in time in the space $H^1(\mathbb{R}^3)$.
- If $\|\varphi_0\|_{L^2} \|\nabla\varphi_0\|_{L^2} > \|Q\|_{L^2} \|\nabla Q\|_{L^2}$, then the solution to (10.1.8) blows up in finite time, i.e. the solution to (10.1.8) exists only in a maximum time interval $[0, T_{\max})$ in $H^1(\mathbb{R}^3)$. The result then follows from Proposition 10.1.12.

□

Proof of Theorem 10.1.9. Theorem 10.1.9 follows immediately from the analysis above. □

The phenomenon of undecidability is, for the blow-up dichotomy, not due to the unboundedness of the domain as the following example shows:

Example 10.1.14 (Cubic NLS on bounded domain)**.** Let $\Omega \subset \mathbb{R}^2$ be a bounded and smooth domain: Consider the cubic NLS with Dirichlet data $\varphi_0 \in H^2(\Omega) \cap H^1_0(\Omega)$

$$\begin{aligned} i\partial_t\psi(x,t) + \Delta\psi(x,t) + |\psi(x,t)|^2\psi(x,t) = 0,\ (x,t) \in \Omega \times (0,T), \\ \psi(x,t) = 0,\ (x,t) \in \partial\Omega \times (0,T), \\ \psi(x,0) = \varphi_0(x),\ x \in \Omega. \end{aligned} \tag{10.1.16}$$

This equation has a unique positive ground state to the equation

$$-Q(x) + \Delta Q(x) + |Q(x)|^2 Q(x) = 0, \quad x \in \mathbb{R}^2.$$

Then, there exists a solution with the same L^2 norm as Q that blows up in finite time [BGT03, Theorem 1], whereas [BGT03, Lemma 2.3] shows that for Dirichlet initial data $\varphi_0 \in H^2(\Omega) \cap H_0^1(\Omega)$ with $\|\varphi_0\|_{L^2(\Omega)} < \|Q\|_{L^2(\mathbb{R}^2)}$ the solution exists globally in time, see also [Wei82].

10.1.3 Future work

- **Fractional PDEs:** The above results on semigroups can be extended to certain time-fractional PDEs [CA22]. So far this has only be done in one space variable. Higher dimensions would be of interest. Further, classifying exactly which type of fractional PDEs lead to Δ_1^A classifications would be an interesting direction.

- **Practical computation:** On the practical side, it should be relatively straightforward (and very useful) to develop a finite element implementation of the above results. NB: The results themselves were proven using spectral methods.

- **Foundations of non-linear PDEs:** It may be possible to use the Δ_1^A algorithm for semigroups as part of exponential integrators to certain solve non-linear PDEs with error control. Currently classifying non-linear PDEs that lend themselves to Δ_1^A results is an open and challenging, yet fundamental, problem. Undoubtedly, this would lead to a theory as rich as that for infinite-dimensional spectral computations.

10.2 Foundations of AI and Smale's 18th Problem

The relevant paper for this is [CAH22b]. For ease of exposition, I will go through this article:
sinews.siam.org/mathematical-paradoxes-unravel-limits-of-ai

This is a hot topic. For example, some news pieces on this paradox can be found here:
https://spectrum.ieee.org/deep-neural-network
www.cam.ac.uk/news/mathematical-paradox-demonstrates-the-limits-of-ai

10.3 Optimisation

This discussion in based on [BHV21].

10.3.1 Background

Finding minimisers for linear and semidefinite programming, regularisation techniques such as basis pursuit, Lasso etc. has become a main focus over the last decades. These approaches have in many areas of mathematics, statistics, learning and data science changed the state of the art from linear to non-linear approaches, typically via obtaining minimisers of convex problems. Key examples include

(i) Linear Programming (LP)

$$z \in \operatorname*{argmin}_{x} \langle x, c \rangle \text{ subject to } Ax = y, \quad x \geq 0, \tag{10.3.1}$$

(ii) Basis Pursuit (BP)

$$z \in \operatorname*{argmin}_{x} \mathcal{J}(x) \text{ subject to } \|Ax - y\|_2 \leq \delta, \qquad \delta \in [0,1], \tag{10.3.2}$$

(iii) Unconstrained Lasso (UL)

$$z \in \operatorname*{argmin}_{x} \|Ax - y\|_2^2 + \lambda \mathcal{J}(x), \qquad \lambda \in (0,1], \tag{10.3.3}$$

(iv) Constrained Lasso (CL)

$$z \in \operatorname*{argmin}_{x} \|Ax - y\|_2 \text{ subject to } \|x\|_1 \leq \tau, \qquad \tau > 0, \tag{10.3.4}$$

(v) Semidefinite Programming (SDP)

$$Z \in \operatorname*{argmin}_{X \in \mathbb{S}^n} \langle C, X \rangle_{\mathbb{S}^n} \text{ subject to } \langle A_k, X \rangle_{\mathbb{S}^n} = b_k,\ X \succeq 0,\ k = 1, \ldots, m. \tag{10.3.5}$$

In the above notation we have

$$A \in \mathbb{R}^{m \times N}, y \in \mathbb{R}^m, c \in \mathbb{R}^N, \quad \mathcal{J}(x) = \|x\|_1 \text{ or } \mathcal{J}(x) = \|x\|_{\mathrm{TV}},$$

where the TV semi-norm is defined as $\|x\|_{\mathrm{TV}} = \sum_{j=1}^{N-1} |x_j - x_{j+1}|$. For SDP, the notation is

$$C, A_k \in \mathbb{S}^n \text{ (real } n \times n \text{ symmetric matrices)}, \quad b_k \in \mathbb{R}, \quad \langle C, X \rangle_{\mathbb{S}^n} = \operatorname{trace}(C^T X).$$

All of the problems above may have multi-valued solutions in certain cases. Whenever this occurs, the computational problem of interest is to compute any of these solutions. We use the notation

$$\Xi : \Omega \rightrightarrows \mathcal{M}, \tag{10.3.6}$$

to denote the multivalued solution map, mapping an input $\iota \in \Omega$ to a metric space $(\mathcal{M}, d_{\mathcal{M}})$, allowing measurement of error. The metric space is typically $\mathbb{R}^N$ or $\mathbb{C}^N$ equipped with the $\|\cdot\|_2$ norm, however, any metric can be considered. Even though the solution map Ξ may be multivalued, in our theory the output of an algorithm will always be single-valued. Thus, if $\Gamma : \Omega \to \mathcal{M}$ is an algorithm we measure the approximation error by

$$\operatorname{dist}_{\mathcal{M}}(\Gamma(\iota), \Xi(\iota)) = \inf_{\xi \in \Xi(\iota)} d_{\mathcal{M}}(\Gamma(\iota), \xi).$$

Remark 10.3.1 (Objective function vs minimisers). *We are primarily concerned with the problem of obtaining minimisers that are vectors and not the real-valued minimum value of the objective function. There is a very rich literature on how to compute the objective function, and, in particular, the minimum value* $f(x^*) = \min\{f(x) \,|\, x \in \mathcal{X}\}$, *for some convex function* $f : \mathbb{R}^d \to \mathbb{R}$, *convex set* $\mathcal{X} \subset \mathbb{R}^d$, *and minimiser* $x^* \in \mathcal{X}$. *The traditional problem of interest is as follows. Given* $\epsilon > 0$, *compute an* $x_\epsilon \in \mathbb{R}^d$ *such that* $f(x_\epsilon) - f(x^*) \leq \epsilon$. *Note that* $f(x_\epsilon) - f(x^*) \leq \epsilon$ *does not necessarily mean that*

$$\|x_\epsilon - x^*\| \leq \epsilon. \tag{10.3.7}$$

Our main focus is the problem of computing x_ϵ *satisfying* (10.3.7). *The motivation behind this is self-evident as there are vast areas of mathematics of information, regularisation, estimation, learning, compressed sensing and data sciences where the object of interest is the minimiser and not the minimum value.*

The question: "is LP in P?" [Kha80, GL81, Law80] was a fundamental problem whose solution, proven by L. Khachiyan – based on work by N. Shor, D. Yudin, A. Nemirovski – reached the front page of The New York Times [GLS88]. The affirmative answer has been refined several times and is now typically stated in the following form. One can solve LPs with rational inputs in runtime is bounded by

$$\mathcal{O}(n^{3.5}L^2 \cdot \log L \cdot \log\log L), \tag{10.3.8}$$

where n denotes the number of variables and L is the number of bits or digits required in the representation of the inputs. The problem, however, is that in an overwhelming number of problems in computational mathematics and scientific computing the input contains irrational numbers. This leads to the following basic question:

> *Given a class of LPs that contain irrational numbers which can be computed in polynomial time, what is the computational cost of computing a K-digit accurate approximate minimiser? Is that problem in P (solvable in polynomial time in the number of variables n)?*

Note that the estimate (10.3.8) will not answer this question as $L = \infty$ for an irrational number.

10.3.2 Inexact input and the extended model

Given that the input is inexact, the output of an algorithm will come with an error as well. The model, both in the Turing and the BSS case, where one measures the computational cost of running the algorithm in terms of the number of variables n and the error (or the number of correct digits $K = |\log(\epsilon)|$, where ϵ is the error) is well established. See, for example [BCSS98, p. 29], [GLS88, p. 34] and [Val13, p. 131]). We thus arrive at the following extension of Smale's 9th problem.

Problem 10.3.2 (**The extended Smale's 9th problem**). *Given any of the problems in* (10.3.1) - (10.3.4)*, represented by the solution map Ξ mapping a class of inputs Ω into a metric space $(\mathcal{M}, d_{\mathcal{M}})$, is there an algorithm which decides the feasibility of the problem, and if so, produces an output that is correct up to K digits (where the error is measured via $\mathrm{dist}_{\mathcal{M}}$) and whose computational cost is bounded by a polynomial in K and the number of variables n?*

This question can be asked both in the Turing model, where the computational cost can be expressed either in terms of the number of steps performed by the Turing machine, or alternatively in terms of the total number of arithmetic operations and comparisons as well as the space complexity. In the BSS model, the computational cost is given by the total number of arithmetic operations and comparisons executed by the BSS machine. We will consider all these cases.

10.3.3 Example Theorem

Theorem 10.3.3 (The extended Smale's 9th problem - computing solutions)**.** *Let Ξ denote the solution map to any of the problems* (10.3.1) - (10.3.4) *with the regularisation parameters satisfying $\delta \in [0, 1]$, $\lambda \in (0, 1/3]$, and $\tau \in [1/2, 2]$ (and additionally being rational in the Turing case) and consider the $\|\cdot\|_p$-norm for measuring the error, for an arbitrary $p \in [1, \infty]$. Let $K > 2$ be an integer. There exists a class Ω of "well-conditioned" feasible inputs so that, simultaneously, we have the following.*

(i) No algorithm can produce K correct digits on each input in Ω. Moreover, for any $\mathrm{p} > \frac{1}{2}$*, no randomised algorithm can produce K correct digits with probability greater than or equal to* p *on each input in Ω.*

(ii) There does exist an algorithm (a Turing or a BSS machine) that produces $K-1$ correct digits for all inputs in Ω. However, any such algorithm will need an arbitrarily long time to achieve this. In particular, for any $T > 0$, and any algorithm Γ, there exists an input $\iota \in \Omega$ such that either Γ on input ι does not produce $K-1$ correct digits for $\Xi(\iota)$ or the runtime of Γ on ι exceeds T. Moreover, for any randomised algorithm Γ^{ran} and $\mathrm{p} < 1/2$ *there exists an input $\iota \in \Omega$ such that*

$$\begin{aligned}\mathbb{P}\big(\Gamma^{\mathrm{ran}}(\iota) \textit{ does not produce } K-1 \textit{ correct digits for } \Xi(\iota) \\ \textit{or the runtime of } \Gamma \textit{ on } \iota \textit{ exceeds } T\big) > \mathrm{p}.\end{aligned}$$

(iii) There exists a polynomial $\mathrm{pol} : \mathbb{R} \to \mathbb{R}$*, as well as a Turing machine and a BSS machine that both produce $K-2$ correct digits for all inputs in Ω, so that the number of arithmetic operations for both machines is bounded by* $\mathrm{pol}(n)$*, where $n = m + mN$ is the number of variables, and the number of digits required from the input oracle is bounded by* $\mathrm{pol}(\log(n))$*. Moreover, the space complexity of the Turing machine is bounded by* $\mathrm{pol}(n)$*.*

(iv) If one only considers (i) - (iii), Ω can be chosen with any fixed dimensions m and N provided that $m \geq 4$ and $N > m$. Moreover, if one only considers (i) then K can be chosen to be 1.

The problem of computing ϵ-approximations to the objective function of NP-hard optimisation problems often leads to phase transitions at the approximation threshold $\epsilon_{\mathrm{A}} > 0$. Indeed, assuming that P $\neq$ NP we often have the following:

Classical phase transition in hardness of approximation

$$\begin{tikzcd}
\boxed{\begin{array}{c}\text{Computing}\\ \epsilon\text{-approx} \in \text{P}\end{array}} \arrow[r, shift left, "\epsilon_A > \epsilon"] & \boxed{\begin{array}{c}\text{Computing } \epsilon\text{-approx}\\ \text{is NP-Hard (thus } \notin \text{P)}\end{array}} \arrow[l, shift left, "\epsilon_A < \epsilon"]
\end{tikzcd} \tag{10.3.9}$$

The fact that $\epsilon_{\mathrm{A}} > 0$ often follows from the PCP theorem [BGS98, ALM$^+$98, FGL$^+$96], for overviews see S. Arora and B. Barak [AB09] and references therein. The extended Smale's 9th problem leads to similar – yet more complex – phase transitions for the problem of computing ϵ-approximations to minimisers in the extended model for classical combinatorial optimisation problems such as LP and problems in continuous optimisation such as BP. This phenomenon is characterised by the *strong breakdown-epsilon* $\epsilon_{\mathrm{B}}^{\mathrm{s}}$ and the *weak breakdown-epsilon* $\epsilon_{\mathrm{B}}^{\mathrm{w}}$, yielding phase transitions in several directions for LP (the computational cost is measured as a function of the number of variables) independent of the P vs NP question:

New phase transitions due to the extended Smale's 9th

$$\begin{tikzcd}
 & \boxed{\begin{array}{c}\text{Computing } \epsilon\text{-approx} \notin \text{k-EXPTIME} \,\forall\, k \\ \text{but} \in \text{R (computable)}\end{array}} \arrow[dl, shift left, "\epsilon_{\mathrm{B}}^{\mathrm{w}}<\epsilon"] \arrow[dr, shift left, "\epsilon_{\mathrm{B}}^{\mathrm{s}}>\epsilon"] & \\
\boxed{\begin{array}{c}\text{Computing} \\ \epsilon\text{-approx} \in \text{P}\end{array}} \arrow[ur, shift left, "\epsilon_{\mathrm{B}}^{\mathrm{w}}>\epsilon>\epsilon_{\mathrm{B}}^{\mathrm{s}}"] \arrow[rr, shift left, "\epsilon_{\mathrm{B}}^{\mathrm{s}}=\epsilon_{\mathrm{B}}^{\mathrm{w}}>\epsilon"] & & \boxed{\begin{array}{c}\text{Computing} \\ \epsilon\text{-approx} \notin \text{R} \\ \text{(non-computable)}\end{array}} \arrow[ul, shift left, "\epsilon_{\mathrm{B}}^{\mathrm{s}}<\epsilon<\epsilon_{\mathrm{B}}^{\mathrm{w}}"] \arrow[ll, shift left, "\epsilon_{\mathrm{B}}^{\mathrm{s}}=\epsilon_{\mathrm{B}}^{\mathrm{w}}<\epsilon"]
\end{tikzcd} \tag{10.3.10}$$

The above integers (K, $K-1$, $K-2$) can be viewed as 'quantised' phase transition thresholds. In particular, we consider the integers $\lceil|\log(\epsilon_{\mathrm{B}}^{\mathrm{w}})|\rceil$ and $\lceil|\log(\epsilon_{\mathrm{B}}^{\mathrm{s}})|\rceil$, but one can easily state our main results with the actual breakdown-epsilons describing the 'unquantised' phase transition threshold as in (10.3.10).

10.3.4 Computing the exit flag - can correctness of algorithms be certified?

A crucial topic in computational mathematics is the reliability of algorithms and certification of their correctness. It is therefore natural to test whether the built-in algorithms in, for example MATLAB are reliable. We consider two concrete examples: the linear program

$$\min_{x\in\mathbb{R}^2} x_1 + x_2 \ \text{ subject to } \ x_1 + (1-\delta)x_2 = 1, \qquad x_1, x_2 \geq 0, \tag{10.3.11}$$

where $\delta > 0$ is a parameter, and the centred and standardised (so that the columns of the design matrix are normalised) Lasso problem

$$\min_{x\in^N} \frac{1}{m}\|A_\delta D_\delta x - y\|_2^2 + \lambda\|x\|_1, \tag{10.3.12}$$

where $m = 3, N = 2, \lambda \in (0, 1/\sqrt{3}]$,

$$A_\delta = \begin{pmatrix} \frac{1}{\sqrt{2}} - \delta & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} - \delta & -\frac{1}{\sqrt{2}} \\ 2\delta & 0 \end{pmatrix} \in^{3\times 2}, \quad y = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{pmatrix}^T \in^3, \tag{10.3.13}$$

and D_δ is the unique diagonal matrix such that each column of $A_\delta D_\delta$ has norm $\sqrt{m}$.

In order to compute a solution to (10.3.11), we consider MATLAB's `linprog` command; a well-established optimisation solver for linear programs. This is a general purpose solver, which offers three different algorithms: 'dual-simplex' (the default), 'interior-point', and 'interior-point-legacy'. Besides a minimiser, `linprog` also computes an additional output – `EXITFLAG` – which is an integer value corresponding to the reason for why the algorithm halted. Note that $+1$ indicates convergence to a minimiser, all other values indicate some form of failure. In Table 10.1 we apply the three `linprog` algorithms (with default settings) to the problem (10.3.11) with different values of δ. The results are fascinating. Not only does `linprog` completely fail to compute a minimiser accurately, it also fails to recognise that the computed minimiser is incorrect: in all cases, the `EXITFLAG` returns the value $+1$ indicating a successful termination.

To compute a solution to (10.3.12), we consider Matlab's `lasso` command. We test it with default settings as well as the tolerance parameter set to machine epsilon $\epsilon_{\mathrm{mach}} = 2^{-52}$ and also the maximum

	'dual-simplex'		'interior-point'		'interior-point-legacy'	
δ	Error	`EXITFLAG`	Error	`EXITFLAG`	Error	`EXITFLAG`
2^{-1}	0	1	0	1	$6.0 \cdot 10^{-12}$	1
2^{-15}	0	1	0	1	$3.0 \cdot 10^{-5}$	1
2^{-20}	0	1	0	1	$7.0 \cdot 10^{-7}$	1
2^{-24}	0	1	0	1	$7.1 \cdot 10^{-8}$	1
2^{-26}	1.4	1	1.4	1	$1.2 \cdot 10^{-1}$	1
2^{-28}	1.4	1	1.4	1	$4.6 \cdot 10^{-1}$	1
2^{-30}	1.4	1	1.4	1	$7.1 \cdot 10^{-1}$	1

Table 10.1: Testing the output of `linprog` applied to the problem in (10.3.11) for the algorithms 'dual-simplex', 'interior-point' and 'interior-point-legacy'. The table shows the error $\|\hat{x} - \tilde{x}\|_{\ell^2}$ and the value of `EXITFLAG` (1 means successful output), where $\hat{x}$ is the true minimiser of (10.3.11) and $\tilde{x}$ is the computed approximate minimiser. Note that machine epsilon is $\epsilon_{\text{mach}} = 2^{-52}$.

	Default settings			'RelTol' $= \epsilon_{\text{mach}}$			'RelTol' $= \epsilon_{\text{mach}}$ 'MaxIter' $= \epsilon_{\text{mach}}^{-1}$		
δ	Error	Runtime	`Warn`	Error	Runtime	`Warn`	Error	Runtime	`Warn`
2^{-1}	$1 \cdot 10^{-16}$	$< 0.01s$	0	$1 \cdot 10^{-16}$	$< 0.01s$	0	$1 \cdot 10^{-16}$	$< 0.01s$	0
2^{-7}	0.68	$< 0.01s$	0	$2 \cdot 10^{-16}$	$0.02s$	0	$2 \cdot 10^{-16}$	$0.02s$	0
2^{-15}	1.17	$< 0.01s$	0	1.17	$0.33s$	1	$1 \cdot 10^{-11}$	$1381.5s$	0
2^{-20}	1.17	$< 0.01s$	0	1.17	$0.33s$	1	no output	$> 12h$	0
2^{-24}	1.17	$< 0.01s$	0	1.17	$0.34s$	1	no output	$> 12h$	0
2^{-26}	1.17	$< 0.01s$	0	1.17	$0.34s$	1	no output	$> 12h$	0
2^{-28}	1.17	$< 0.01s$	0	1.17	$< 0.01s$	0	1.17	$< 0.01s$	0
2^{-30}	1.17	$< 0.01s$	0	1.17	$< 0.01s$	0	1.17	$< 0.01s$	0

Table 10.2: The output of `lasso` applied to (10.3.12) with inputs as in (10.3.13) and $\lambda = 0.1$. The table shows the error $\|\hat{x} - \tilde{x}\|_{\ell^2}$ (where $\hat{x}$ is the true minimiser and $\tilde{x}$ is the computed minimiser), the CPU runtime, and a boolean value indicating whether a `Warning` was issued.

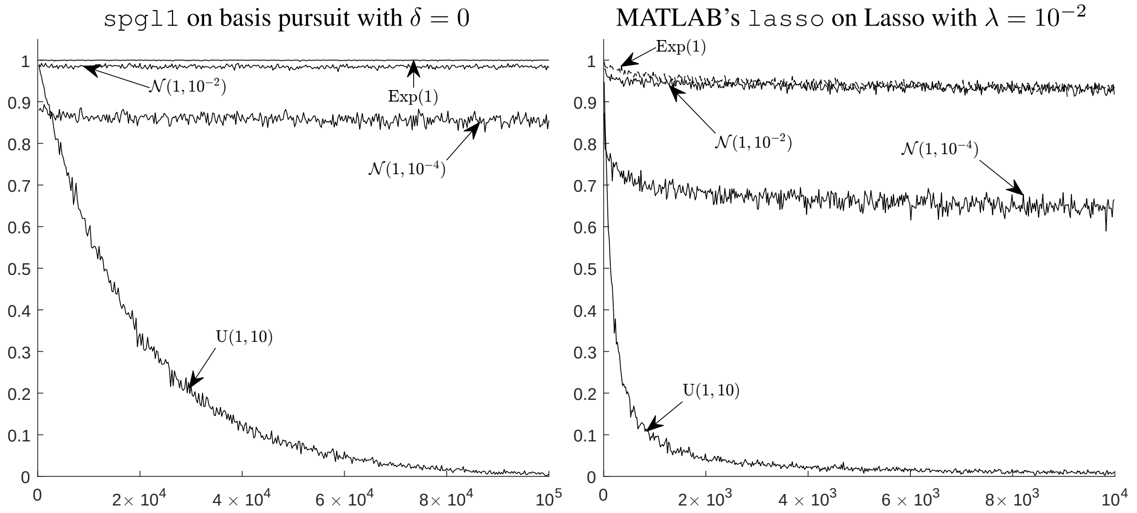


Figure 10.1: (**Random matrices – Non-computability is not rare**). The vertical axis represents the success rate $\frac{\#\text{ of successes}}{\#\text{ of trials}}$, where $\#$ of trials $= 1200$. Success $\Leftrightarrow$ computed solution is accurate to at least $K = 2$ digits ($\|\cdot\|_\infty$ norm). The horizontal axis shows the dimension N. In all cases, $A \in \mathbb{R}^{m\times N}$ in (10.3.2) and (10.3.3) is iid – as described in §10.3.5 – according to the distributions $\mathrm{U}(a, b)$, $\mathrm{Exp}(\nu)$ and $\mathcal{N}(\mu, \sigma^2)$, being the uniform distribution on $[a, b]$, the exponential distribution with parameter ν and the normal distribution with mean μ and variance σ.

number of iterations to $\epsilon_{\mathrm{mach}}^{-1}$. The `lasso` routine does not have an 'exit flag', however, it provides a `Warning` if it considers the output to be untrustworthy. The results of this experiment are summarised in Table 10.2, where we display 1 under the `Warn` column if a `Warning` was issued, or 0 if no warning was issued. As is evident, the failure of `lasso` is similar to the failure of `linprog`, however, an interesting observation is that the `Warning` parameter is occasionally able to verify the wrong solution, yet, most of the time, no warning is issued despite completely inaccurate outputs.

10.3.5 Non-computability is not rare

Theorem 10.3.3 demonstrates that for any integer K, there are – for all problems (10.3.1) - (10.3.5) – classes of inputs for which no algorithm can compute a correct K digit approximate solution. This statement, as is typical for a result regarding non-computability, describes a worst case scenario. However, the proof techniques of our theorems reveal much more. Indeed, for random matrix ensembles, one can characterise the probability of failure of algorithms. This is because our proof of Theorem 10.3.3 is constructive.

To be more precise, Figure 10.1 displays experiments with well-established algorithms such as `spgl1` [vdBF08] and MATLAB's `lasso`. We have tested these algorithms on BP (10.3.2) with $\delta = 0$ and Lasso (10.3.3) with $\lambda = 10^{-2}$. In both cases, all accuracy parameters in the algorithms were set to machine precision ϵ_{mach} in MATLAB, and the number of iterations in `spgl1` and `lasso` were set to 1000 and the default parameter respectively. We executed these algorithms on inputs $A \in \mathbb{R}^{m\times N}$ and $y \in \mathbb{R}^m$, where the entries of A are iid according to a distribution $\mathcal{D}$ and $y = Ae_i$ where $i \in \{1, \ldots, N\}$ is chosen uniformly at random. In particular, we examine the cases where $\mathcal{D}$ is a normal distribution $\mathcal{N}(\mu, \sigma^2)$ with mean μ and variance σ, $\mathcal{D}$ is a uniform distribution $\mathrm{U}(a, b)$ on the interval $[a, b]$, or $\mathcal{D}$ is an exponential distribution

$\mathrm{Exp}(\nu)$ with parameter ν. Figure 10.1 displays the results for $m = 1$ and varying Ns, where we plot the 'success rate' given by

$$\text{Success rate } = \frac{\#\text{ of successes}}{\#\text{ of trials}} \in [0,1],$$

as a function of the dimension N. In all cases $\#$ of trials $= 1200$. Given a distribution $\mathcal{D}$, let $\iota_N = (y, A)$ be such that $y \in^1$ and $A \in^{1\times N}$ are randomly chosen as described above. For any choice of algorithm Γ that solve Basis Pursuit or Lasso implemented in floating point arithmetic and any $K \in \mathbb{N}$ define (when the limit exists)

$$\lim_{N\to\infty} \mathbb{P}\big(\Gamma(\iota_N) \text{ provides } K \text{ correct digits}\big) =: P_{\infty}^{\Gamma}(\mathcal{D}).$$

One can show that, for example, when Γ represents the `spgl1` algorithm we have

$$P_{\infty}^{\Gamma}(\mathcal{D}) = 0 \text{ for } \mathcal{D} = \mathrm{U}(a, b).$$

Note that this asymptotic behaviour is already visible in Figure 10.1. For the Gaussian and Exponential distributions, the issue is more complicated and there may be transient behaviour. Indeed, let Ξ denote the solution map to the Lasso problem in Figure 10.1 (so that Ξ outputs a solution to (10.3.3)) and assume that $\mathcal{D}$ is $\mathcal{N}(\mu, \sigma^2)$ or that $\mathcal{D}$ is $\mathrm{Exp}(\nu)$. Then $\mathbb{P}(\|\Xi(\iota_N)\|_\infty < 10^{-k}) \to 1$, as $N \to \infty$ for all $k \in \mathbb{N}$. Thus, an algorithm that always outputs zero will eventually become correct with high probability. However, in Figure 10.1, there is behaviour of the following form: there exists an M so that

$$\mathbb{P}\big(\Gamma(\iota_N) \text{ provides } K \text{ correct digits}\big) > \mathbb{P}\big(\Gamma(\iota_{N+1}) \text{ provides } K \text{ correct digits}\big) \quad N \leq M,$$

for some large $M \in \mathbb{N}$, yet for any algorithm Γ – such that the objective function applied to $\Gamma(\iota_N)$ is ε (in this example $\varepsilon = 10^{-5}$ suffices) away from the true minimum – we have $P_{\infty}^{\Gamma}(\mathcal{D}) = 1$. The latter is typically true for both `spgl1` and `lasso`, thus Figure 10.1 demonstrates a transient behaviour for both the normal and exponential distributions. These phenomena can be mathematically analysed and proven by using the specific techniques used in the proof of Theorem 10.3.3.

10.4 Computer-assisted Proofs

Computer-assisted proofs are rapidly becoming an important part of modern pure mathematics:

> *"During the next century computers will become sufficiently good at proving theorems that the practice of pure mathematical research will be completely revolutionized."*
>
> — Sir W.T. Gowers (Fields medal 1998), Cambridge [Gow00]

Recent examples given in Hales' proof of Kepler's conjecture (Hilbert's 18th problem) [Hal05, HAB$^+$17] and Fefferman (Fields medal 1978) and Seco's proof of the Dirac–Schwinger conjecture [FS90, FS92, FS93, FS94b, FS94c, FS95, FS96b, FS96a, FS94a], see also the discussion of Fefferman's 2017 Wolf Prize [CST$^+$17]. A potentially surprising result is that both of these examples are computer-assisted proofs that use non-computable problems. This can be understood via the precise notions of error control in §2.2. The theory of computer-assisted proofs has not yet been developed, since, in general, it is not known which computational problems can be used in computer-assisted proofs.

Any computation that arises in a proof must be performed reliably with 100% verification. At first, one might expect that this can only be achieved with Δ_1^A computational problems, i.e., problems that are computable in the classical Turing sense. However, this is not the case and bears a resemblance to the notion of recursively enumerable sets in classical computation theory. For example, the computer-assisted proof of Kepler's conjecture is based on problems that are in Σ_1^A but not Δ_1^G. There are several such examples:

- ***Kepler's Conjecture (Hilbert's 18th problem) - SCI classification:*** $\in \Sigma_1^A$**,** $\notin \Delta_1^G$: Kepler conjectured that no packing of congruent balls in Euclidean three space has density greater than that of the face-centred cubic packing. The Flyspeck programme, led by Hales [Hal05, HAB$^+$17], provides a fully computer-assisted verification. The key computational part relies on deciding about 50000 linear programs with irrational inputs. More specifically, to decide whether there exists an $x \in \mathbb{R}^N$ such that

$$\langle x, c\rangle_K \leq M \text{ subject to } Ax = y, \quad x \geq 0, \tag{10.4.1}$$

$$\langle x, c\rangle_K = \lfloor 10^K \langle x, c\rangle \rfloor 10^{-K}, \quad K \in \mathbb{N}, \quad M \in \mathbb{Q}.$$

  Since A and y can be irrational, one can think of this as a decision problem with inexact input (a Turing machine or a BSS machine that can access $A \in \mathbb{R}^{m\times N}$ in the form of an oracle $\mathcal{O}_A$ such that $|\mathcal{O}_A(i,j,k) - A_{i,j}| \leq 2^{-k}$). The following facts about the problem (10.4.1) and its classification hold:

  (i) For any integer $\tilde{K} > 1$ there exists a class of inputs Ω such that the problem (10.4.1) with $K = \tilde{K}$ is $\notin \Sigma_1^G$. However, with the same input class Ω, we have that the problem (10.4.1), with $K = \tilde{K} - 1$ is $\in \Delta_1^A$.

  (ii) The raises the question of how the computer-assisted proof of Kepler's conjecture was at all possible, given that (10.4.1) must be decided for $K = 6$. Given the class Ω in (i), if the inequality $\langle x, c\rangle_K \leq M$ in (10.4.1) is replaced by a strict inequality $\langle x, c\rangle_K < M$, then the problem is in Σ_1^A. A similar (though much more complicated) analysis occurs, and leads to a series of Σ_1^A problems which are solved in the Flyspeck programme.

- ***Dirac–Schwinger conjecture - SCI classification:*** $\in \Sigma_1^A$**,** $\notin \Delta_1^G$: The Dirac–Schwinger conjecture was proven in a series of papers by Fefferman and Seco [FS90, FS92, FS93, FS94b, FS94c, FS95, FS96b, FS96a, FS94a]. Consider the Hamiltonian

$$H_{dZ} = \sum_{k=1}^{d} (-\triangle_{x_k} - Z|x_k|^{-1}) + \sum_{1\leq j\leq k\leq d} |x_j - x_k|^{-1}$$

  acting on antisymmetric functions in $L^2(\mathbb{R}^{3d})$. The ground state energy $E(d,Z)$ for d electrons and a nucleus of charge Z is then defined by

$$E(d,Z) := \inf\{\lambda \in \mathrm{Sp}(H_{dZ})\}.$$

  The ground state energy of an atom is then defined as $E(Z) := \min_{d\geq 1} E(d,Z)$. The key result is asymptotic behaviour of $E(Z)$ for large Z:

$$E(Z) = -c_0 Z^{7/3} + \frac{1}{8}Z^2 - c_1 Z^{5/3} + \mathcal{O}(Z^{5/3-1/2835}),$$

  for some explicitly defined constants c_0 and c_1. In order to show this, the proof verified that $F''(\omega) \leq c < 0$ for some specific function F, for some c and for all $\omega \in (0, \omega_c)$ where ω_c is specifically defined. A full discussion of the details is beyond the scope of this thesis, but the intricate computer-assisted proof hinges on several problems that are $\notin \Delta_1^G$ but $\in \Sigma_1^A$ (see, for example, Algorithm 3.7 and Algorithm 3.8 in [FS96b]).

- ***Boolean Pythagorean triples problem - SCI classification:*** $\in \Pi_1^A$***,*** $\notin \Delta_1^G$: The Boolean Pythagorean triples problem asks if it is possible to colour each of the positive integers either red or blue, so that no Pythagorean triple of integers a, b, c, satisfying $a^2 + b^2 = c^2$ are all the same colour. This is true up to $n = 7824$, and the proof, performed by Heule, Kullmann, and Marek (2016) [HKM16], is based on computations showing that this is not true for $n = 7825$. Clearly, for any finite set of integers, the combinatorial problem lies $\in \Delta_0^A$, but it is not $\in \Delta_0^G$ for the whole set $\mathbb{N}$. However, by checking each successive integer, it is clear that the problem does lie $\in \Pi_1^A$. Such proofs for counterexamples are common for disproving conjectures within number theory.
- ***Group theory:*** $\mathrm{Aut}(\mathbb{F}_5)$ ***has property*** (T) ***- SCI classification :*** $\in \Sigma_1^A$***,*** $\notin \Delta_1^G$: The fact that the automorphism group of the free group on five generators has Kazhdan's property (T), was shown by Kaluba, Nowak and Ozawa [KNO19]. The key computational problem involves a (root of a) minimiser of a semi-definite program. This is computed using floating-point arithmetic, which, at best, is equivalent to solving the semi-definite program with inexact input. This problem is $\notin \Delta_1^G$ but is $\in \Delta_2^A$. There is no concept of Σ_1^A for minimisers of semi-definite programs, but the reasoning in the paper [KNO19] regarding the verification implies that the final decision problem is $\in \Sigma_1^A$.

A key part in all of the examples above is that one must prove either Σ_1^A or Π_1^A classifications in order to demonstrate that the verification is possible. This is trivial in the Boolean Pythagorean triples problem, but is very technical in the proof of the Dirac–Schwinger conjecture. Regarding spectral problems, many of the results in this course led to Σ_1^A or Π_1^A classifications. It follows that such computations could be used as part of a proof. Figuring out exactly which problems can similarly be used will be a key part of mathematics in the coming decades.

Bibliography

[A^+08] X. Antoine et al. A review of transparent and artificial boundary conditions techniques for linear and nonlinear Schrödinger equations. 2008.

[AA68] Vladimir Igorevich Arnold and André Avez. *Ergodic Problems of Classical Mechanics*, volume 9. New York, Benjamin, 1968.

[AB09] Sanjeev Arora and Boaz Barak. *Computational Complexity - A Modern Approach*. Princeton University Press, 2009.

[ABHN01] W. Arendt, C. J. K. Batty, M. Hieber, and F. Neubrander. Cauchy problems. In *Vector-valued Laplace Transforms and Cauchy Problems*, pages 109–240. Springer, 2001.

[ABP06] Paola F. Antonietti, Annalisa Buffa, and Ilaria Perugia. Discontinuous Galerkin approximation of the Laplace eigenproblem. *Comput. Methods Appl. Mech. Engrg.*, 195(25-28):3483–3503, 2006.

[AEG14] Andrea Agazzi, Jean-Pierre Eckmann, and Gian M. Graf. The colored Hofstadter butterfly for the honeycomb lattice. *Journal of statistical physics*, 156(3):417–426, 2014.

[AES03] A. Arnold, M. Ehrhardt, and I. Sofronov. Discrete transparent boundary conditions for the Schrödinger equation: Fast calculation, approximation, and stability. *Commun. Math. Sci.*, 1(3):501–556, 2003.

[AG74] W.O. Amrein and V. Georgescu. On the characterization of bound states and scattering states in quantum mechanics. *Helv. Phys. Acta*, 46:635–658, 1973/74.

[AH12] Ben Adcock and Anders C. Hansen. Stable reconstructions in Hilbert spaces and the resolution of the Gibbs phenomenon. *Appl. Comput. Harm. Anal.*, 32(3):357–388, 2012.

[AJ09] Artur Avila and Svetlana Jitomirskaya. The Ten Martini Problem. *Annals of Mathematics (2)*, 170(1):303–342, 2009.

[AJM17] Artur Avila, Svetlana Jitomirskaya, and C. A. Marx. Spectral theory of extended Harper's model and a question by Erdős and Szekeres. *Inventiones mathematicae*, 210(1):283–339, 2017.

[AK06] Artur Avila and Raphaël Krikorian. Reducibility or nonuniform hyperbolicity for quasiperiodic Schrödinger cocycles. *Annals of Mathematics (2)*, 164(3):911–940, 2006.

[Akh65] Naum I. Akhiezer. *The classical moment problem and some related questions in analysis*. Translated by N. Kemmer. Hafner Publishing Co., New York, 1965.

[ALM^+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, May 1998.

[AMH11] A. H. Al-Mohy and N. J. Higham. Computing the action of the matrix exponential, with an application to exponential integrators. *SIAM J. Sci. Comput.*, 33(2):488–511, 2011.

[And58] Philip W. Anderson. Absence of diffusion in certain random lattices. *Physical Review*, 109(5):1492, 1958.

[Aro51] Nachman Aronszajn. Approximation methods for eigenvalues of completely continuous symmetric operators. In *Proceedings of the Symposium on Spectral Theory and Differential Problems*, pages 179–202. Oklahoma Agricultural and Mechanical College, Stillwater, Okla., 1951.

[Arv93a] William Arveson. Improper filtrations for C^*-algebras: spectra of unilateral tridiagonal operators. *Acta Sci. Math. (Szeged)*, 57(1-4):11–24, 1993.

[Arv93b] William Arveson. Noncommutative spheres and numerical quantum mechanics. In *Operator algebras, mathematical physics, and low-dimensional topology*, volume 5 of *Res. Notes Math.*, pages 1–10. A K Peters, Wellesley, MA, 1993.

[Arv94a] William Arveson. C^*-algebras and numerical linear algebra. *Journal of Functional Analysis*, 122(2):333–360, 1994.

[Arv94b] William Arveson. The role of C^*-algebras in infinite-dimensional numerical linear algebra. In *C^*-algebras: 1943–1993 (San Antonio, TX, 1993)*, volume 167 of *Contemp. Math.*, pages 114–129. Amer. Math. Soc., Providence, RI, 1994.

[AV07] Artur Avila and Marcelo Viana. Simplicity of Lyapunov spectra: proof of the Zorich-Kontsevich conjecture. *Acta Mathematica*, 198(1):1–56, 2007.

[Avi08] Artur Avila. The absolutely continuous spectrum of the almost Mathieu operator. *arXiv:0810.2965*, 2008.

[Avi09] Artur Avila. On the spectrum and Lyapunov exponent of limit periodic Schrödinger operators. *Communications in Mathematical Physics*, 288(3):907–918, 2009.

[BACH$^+$20] J. Ben-Artzi, M. J. Colbrook, A. C. Hansen, O. Nevanlinna, and M. Seidel. Computing Spectra – On the Solvability Complexity Index hierarchy and towers of algorithms. *arXiv:1508.03280*, 2020.

[BB98] Carl M. Bender and Stefan Boettcher. Real spectra in non-Hermitian Hamiltonians having PT symmetry. *Physical Review Letters*, 80(24):5243, 1998.

[BBG00] Daniele Boffi, Franco Brezzi, and Lucia Gastaldi. On the problem of spurious eigenvalues in the approximation of linear elliptic problems in mixed form. *Mathematics of Computation*, 69(229):121–140, 2000.

[BBG13] Daniele Boffi, Annalisa Buffa, and Lucia Gastaldi. Convergence analysis for hyperbolic evolution problems in mixed form. *Numer. Linear Algebra Appl.*, 20(4):541–556, 2013.

[BBIN10] Albrecht Böttcher, Hermann Brunner, Arieh Iserles, and Syvert P. Nørsett. On the singular values and eigenvalues of the Fox-Li and related operators. *New York J. Math.*, 16:539–561, 2010.

[BBJ02] Carl M. Bender, Dorje C. Brody, and Hugh F. Jones. Complex extension of quantum mechanics. *Physical Review Letters*, 89(27):270401, 2002.

[BBKK21] Steven L Brunton, Marko Budišić, Eurika Kaiser, and J Nathan Kutz. Modern Koopman theory for dynamical systems. *arXiv preprint arXiv:2102.12086*, 2021.

[BC71] Erik Balslev and Jean-Michel Combes. Spectral properties of many-body Schrödinger operators with dilatation-analytic interactions. *Communications in Mathematical Physics*, 22:280–294, 1971.

[BCD$^+$06] E. Bombieri, S. Cook, P. Deligne, C.L. Fefferman, J. Gray, A. Jaffe, J. Milnor, A. Wiles, and E. Witten. The Millennium Prize Problems. *CMI/AMS*, 2006.

[BCJ09] Annalisa Buffa, Patrick Ciarlet, Jr., and Erell Jamelot. Solving electromagnetic eigenvalue problems in polyhedral domains with nodal finite elements. *Numerische Mathematik*, 113(4):497–518, 2009.

[BCN01] Albrecht Böttcher, A.V. Chithra, and M.N.N. Namboodiri. Approximation of approximation numbers by truncation. *Integral Equations and Operator Theory*, 39(4):387–395, 2001.

[BCSS98] Lenore Blum, Felipe Cucker, Michael Shub, and Steve Smale. *Complexity and Real Computation*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1998.

[BDG99] Daniele Boffi, Ricardo G. Duran, and Lucia Gastaldi. A remark on spurious eigenvalues in a square. *Appl. Math. Lett.*, 12(3):107–114, 1999.

[Bee93] Gerald Beer. *Topologies on closed and closed convex sets*, volume 268 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1993.

[Ben07] Carl M. Bender. Making sense of non-Hermitian Hamiltonians. *Rep. Prog. Phys.*, 70(6):947, 2007.

[BFG$^+$] Oliver Beckstein, Shujie Fan, Richard J. Gowers, Micaela Matta, and Lily Wang. AdK equilibrium dataset. `https://www.mdanalysis.org/MDAnalysisData/adk_equilibrium.html`.

[BFKS09] Oliver Bendix, Ragnar Fleischmann, Tsampikos Kottos, and Boris Shapiro. Exponentially fragile PT symmetry in lattices with localized eigenmodes. *Physical Review Letters*, 103(3):030402, 2009.

[BGS98] Mihir Bellare, Oded Goldreich, and Madhu Sudan. Free bits, pcps, and nonapproximability – towards tight results. *SIAM J. Comput.*, 27(3):804–915, 1998.

[BGT03] Nicolas Burq, Pierre Gérard, and N Tzetkov. Two singular dynamics of the nonlinear Schrödinger equation on a plane domain. *Geometric And Functional Analysis*, 13(1):1–19, 2003.

[BH20] Simon Becker and Anders Hansen. Computing solutions of Schrödinger odinger equations on unbounded domains - on the brink of numerical algorithms. *arXiv preprint arXiv:2010.16347*, 2020.

[BHP07] Annalisa Buffa, Paul Houston, and Ilaria Perugia. Discontinuous Galerkin computation of the Maxwell eigenvalues on simplicial meshes. *Journal of Computational and Applied Mathematics*, 204(2):317–333, 2007.

[BHV21] Alexander Bastounis, Anders C Hansen, and Verner Vlačić. The extended Smale's 9th problem–On computational barriers and paradoxes in estimation, regularisation, computer-assisted proofs and learning. *arXiv preprint arXiv:2110.15734*, 2021.

[BIN11] Hermann Brunner, Arieh Iserles, and Syvert P. Nørsett. The computation of the spectra of highly oscillatory Fredholm integral operators. *Journal of Integral Equations and Applications*, 23(4):467–519, 2011.

[BK05] Jean Bourgain and Carlos E. Kenig. On localization in the continuous Anderson-Bernoulli model in higher dimension. *Invent. Math.*, 161(2):389–426, 2005.

[BK19] Steven L Brunton and J Nathan Kutz. *Data-driven Science and Engineering: Machine learning, Dynamical systems, and Control*. Cambridge University Press, 2019.

[BMM12] Marko Budišić, Ryan Mohr, and Igor Mezić. Applied Koopmanism. *Chaos*, 22(4):047510, 2012.

[BMT20] Sabine Bögli, Marco Marletta, and Christiane Tretter. The essential numerical range for unbounded linear operators. *Journal of Functional Analysis*, page 108509, 2020.

[BO13] Carl M. Bender and Steven A. Orszag. *Advanced mathematical methods for scientists and engineers I: Asymptotic methods and perturbation theory*. Springer Science & Business Media, 2013.

[Böt94] Albrecht Böttcher. Pseudospectra and singular values of large convolution operators. *Journal of Integral Equations and Applications*, 6(3):267–301, 1994.

[Böt96] Albrecht Böttcher. Infinite matrices and projection methods. In *Lectures on operator theory and its applications (Waterloo, ON, 1994)*, volume 3 of *Fields Inst. Monogr.*, pages 1–72. Amer. Math. Soc., Providence, RI, 1996.

[BP84] N. Beer and D. G. Pettifor. The recursion method and the estimation of local densities of states. In *The Electronic Structure of Complex Systems*, pages 769–777. Springer, 1984.

[BP06] Annalisa Buffa and Ilaria Perugia. Discontinuous Galerkin approximation of the Maxwell eigenproblem. *SIAM Journal on Numerical Analysis*, 44(5):2198–2226, 2006.

[BPW09] Annalisa Buffa, Ilaria Perugia, and Tim Warburton. The mortar-discontinuous Galerkin method for the 2D Maxwell eigenproblem. *J. Sci. Comput.*, 40(1-3):86–114, 2009.

[BRS16] Miguel A. Bandres, Mikael C. Rechtsman, and Mordechai Segev. Topological photonic quasicrystals: Fractal topological spectrum and protected transport. *Physical Review X*, 6(1):011016, 2016.

[BS83] Albrecht Böttcher and Bernd Silbermann. The finite section method for Toeplitz operators on the quarter-plane with piecewise continuous symbols. *Math. Nachr.*, 110:279–291, 1983.

[BS91] Vincenzo G. Benza and Clément Sire. Band spectrum of the octagonal quasicrystal: Finite measure, gaps, and chaos. *Physical Review B*, 44(18):10343, 1991.

[BS99] Albrecht Böttcher and Bernd Silbermann. *Introduction to large truncated Toeplitz matrices*. Universitext. Springer-Verlag, New York, 1999.

[BT79] P. Brenner and V. Thomée. On rational approximations of semigroups. *SIAM J. Numer. Anal.*, 16(4):683–694, 1979.

[BT89] Martin Blümlinger and Robert F. Tichy. Topological algebras of functions of bounded variation I. *Manuscripta Mathematica*, 65(2):245–255, 1989.

[BW73] Carl M. Bender and Tai T. Wu. Anharmonic oscillator. II. A study of perturbation theory in large order. *Physical Review D*, 7(6):1620, 1973.

[CA22] Matthew J Colbrook and Lorna J Ayton. A contour method for time-fractional PDEs and an application to fractional viscoelastic beam equations. *Journal of Computational Physics*, page 110995, 2022.

[Caf98] Russel E Caflisch. Monte Carlo and quasi-Monte Carlo methods. *Acta Numer.*, 7:1–49, 1998.

[CAH22a] Matthew J. Colbrook, Vegard Antun, and Anders C. Hansen. The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and smale's 18th problem. *Proceedings of the National Academy of Sciences*, 119(12):e2107151119, 2022.

[CAH22b] Matthew J Colbrook, Vegard Antun, and Anders C Hansen. The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem. *Proceedings of the National Academy of Sciences*, 119(12):e2107151119, 2022.

[CAS22] Matthew J Colbrook, Lorna Ayton, and Máté Szőke. Residual Dynamic Mode Decomposition: Robust and verified Koopmanism. *arXiv preprint arXiv:2111.14889*, 2022.

[CFKS87] Hans L. Cycon, Richard G. Froese, Werner Kirsch, and Barry Simon. *Schrödinger operators with application to quantum mechanics and global geometry*. Texts and Monographs in Physics. Springer-Verlag, Berlin, study edition, 1987.

[CGM80] E. Caliceti, S. Graffi, and M. Maioli. Perturbation theory of odd anharmonic oscillators. *Communications in Mathematical Physics*, 75(1):51–66, 1980.

[CH19] Matthew J. Colbrook and Anders C. Hansen. On the infinite-dimensional QR algorithm. *Numerische Mathematik*, 143(1):17–83, 2019.

[Cha19] K. Chang. A physics magic trick: Take 2 sheets of carbon and twist. *The New York Times*, Oct 2019.

[CHns] Matthew J. Colbrook and Anders C. Hansen. The foundations of spectral computations via the solvability complexity index hierarchy. *Journal of the European Mathematical Society*, under revisions.

[CHT21] Matthew J. Colbrook, Andrew Horning, and Alex Townsend. Computing spectral measures of self-adjoint operators. *SIAM Review*, 63(3):489–524, 2021.

[CHTW21] Matthew J Colbrook, Andrew Horning, Kyle Thicke, and Alexander B Watson. Computing spectral properties of topological insulators without artificial truncation or supercell approximation. *arXiv preprint arXiv:2112.03942*, 2021.

[CL90] René Carmona and Jean Lacroix. *Spectral theory of random Schrödinger operators*. Probability and its Applications. Birkhäuser Boston, Inc., Boston, MA, 1990.

[CLPT93] M. Crouzeix, S. Larsson, S. Piskarev, and V. Thomée. The stability of rational approximations of analytic semigroups. *BIT*, 33(1):74–84, 1993.

[Col20a] Matthew J. Colbrook. *The Foundations of Infinite-Dimensional Spectral Computations*. PhD thesis, University of Cambridge, 2020.

[Col20b] Matthew J. Colbrook. Pseudoergodic operators and periodic boundary conditions. *Mathematics of Computation*, 89(322):737–766, 2020.

[Col21] Matthew J. Colbrook. Computing spectral measures and spectral types. *Communications in Mathematical Physics*, 384(1):433–501, 2021.

[Col22] Matthew J Colbrook. Computing semigroups with error control. *SIAM Journal on Numerical Analysis*, 60(1):396–422, 2022.

[Colns] Matthew J. Colbrook. On the computation of geometric features of spectra of linear operators on hilbert spaces. *Foundations of Computational Mathematics*, under revisions.

[Com93] Jean-Michel Combes. Connections between quantum dynamics and spectral properties of time-evolution operators. In *Differential equations with applications to mathematical physics*, volume 192 of *Math. Sci. Engrg.*, pages 59–68. Academic Press, Boston, 1993.

[CPGW15] Toby S. Cubitt, David Perez-Garcia, and Michael M. Wolf. Undecidability of the spectral gap. *Nature*, 528(7581):207, 2015.

[CRH19] Matthew J. Colbrook, Bogdan Roman, and Anders C. Hansen. How to compute spectra with error control. *Physical Review Letters*, 122(25):250201, 2019.

[CST$^+$17] Antonio Córdoba, Elias Stein, Terence Tao, Louis Nirenberg, Joseph J. Kohn, Sun-Yung Alice Chang, C. Robin Graham, Diego Córdoba, Bo'az Klartag, Jürg Fröhlich, Luis Seco, and Michael Weinstein. Ad honorem Charles Fefferman. *Notices Amer. Math. Soc.*, 64(11):1254–1273, 2017.

[CT21] Matthew J Colbrook and Alex Townsend. Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems. *arXiv preprint arXiv:2111.14889*, 2021.

[CW13] Snorre H. Christiansen and Ragnar Winther. On variational eigenvalue approximation of semidefinite operators. *IMA J. Numer. Anal.*, 33(1):164–189, 2013.

[Dav98] E. Brian Davies. Spectral enclosures and complex resonances for general self-adjoint operators. *LMS J. Comput. Math.*, 1:42–74, 1998.

[DDG$^+$12] Ron O Dror, Robert M Dirks, JP Grossman, Huafeng Xu, and David E Shaw. Biomolecular simulation: a computational microscope for molecular biology. *Ann. Rev. Biophys.*, 41:429–452, 2012.

[DDT01] Patrick Dorey, Clare Dunning, and Roberto Tateo. Spectral equivalences, Bethe ansatz equations, and reality properties in PT-symmetric quantum mechanics. *Journal of Physics A: Mathematical and General*, 34(28):5679, 2001.

[Dei99] Percy Deift. *Orthogonal polynomials and random matrices: a Riemann-Hilbert approach*, volume 3 of *Courant Lecture Notes in Mathematics*. New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 1999.

[DFJ90] K. G. Dyall and K. Fægri Jr. Kinetic balance and variational bounds failure in the solution of the Dirac equation in a finite Gaussian basis set. *Chem. Phys. Let.*, 174(1):25–32, 1990.

[DG81] Gordon W. F. Drake and S. P. Goldman. Application of discrete-basis-set methods to the Dirac equation. *Physical Review A*, 23(5):2093, 1981.

[DGS15] David Damanik, Anton Gorodetski, and Boris Solomyak. Absolutely continuous convolutions of singular measures and an application to the square Fibonacci Hamiltonian. *Duke Math. J.*, 164(8):1603–1640, 2015.

[DLT85] Percy Deift, Luenchau C. Li, and Carlos Tomei. Toda flows with infinitely many variables. *Journal of Functional Analysis*, 64(3):358–402, 1985.

[DN86] Joanne Dombrowski and Paul Nevai. Orthogonal polynomials, measures and recurrence relations. *SIAM Journal on Mathematical Analysis*, 17(3):752–759, 1986.

[Dod15] Benjamin Dodson. Global well-posedness and scattering for the mass critical nonlinear Schrödinger equation with mass below the mass of the ground state. *Advances in mathematics*, 285:1589–1618, 2015.

[Don63] William F. Donoghue, Jr. On a problem of Nieminen. *Inst. Hautes Études Sci. Publ. Math.*, pages 31–33, 1963.

[DS06a] D. Damanik and B. Simon. Jost functions and Jost solutions for Jacobi matrices, I. A necessary and sufficient condition for Szegő asymptotics. *Invent. Math.*, 165(1):1–50, 2006.

[DS06b] Laurent Demanet and Wilhelm Schlag. Numerical verification of a gap condition for a linearized nonlinear Schrödinger equation. *Nonlinearity*, 19(4):829–852, 2006.

[DS11] Roland Donninger and Wilhelm Schlag. Numerical study of the blowup/global existence dichotomy for the focusing cubic nonlinear Klein–Gordon equation. *Nonlinearity*, 24(9):2547, 2011.

[DVV94] Trond Digernes, Veeravalli S. Varadarajan, and S. R. Srinivasa Varadhan. Finite approximations to quantum systems. *Rev. Math. Phys.*, 6(4):621–648, 1994.

[DWM$^+$13] Cory R. Dean, L. Wang, P. Maher, C. Forsythe, F. Ghahari, Y. Gao, J. Katoch, M. Ishigami, P. Moon, M. Koshino, et al. Hofstadter's butterfly and the fractal quantum Hall effect in moire superlattices. *Nature*, 497(7451):598–602, 2013.

[ELO94] V. D. Efros, W. Leidemann, and G. Orlandini. Response functions from integral transforms with a Lorentz kernel. *Phys. Lett. B*, 338(2-3):130–133, 1994.

[ELOB07] V. D. Efros, W. Leidemann, G. Orlandini, and N. Barnea. The Lorentz integral transform (LIT) method and its applications to perturbation-induced reactions. *J. Phys. G*, 34(12):R459, 2007.

[ELS08] Maria J. Esteban, Mathieu Lewin, and Eric Séré. Variational methods in relativistic quantum mechanics. *Bull. Amer. Math. Soc. (N.S.)*, 45(4):535–593, 2008.

[ELS19] V. D. Efros, W. Leidemann, and V. Y. Shalamova. On calculating response functions via their Lorentz integral transforms. *Few-Body Sys.*, 60(2):35, 2019.

[EM77] B. Engquist and A. Majda. Absorbing boundary conditions for numerical simulation of waves. *Proc. Natl. Acad. Sci.*, 74(5):1765–1766, 1977.

[EM79] B. Engquist and A. Majda. Radiation boundary conditions for acoustic and elastic wave calculations. *Comm. Pure Appl. Math.*, 32:313–357, 1979.

[Ens78] Volker Enss. Asymptotic completeness for quantum mechanical potential scattering. I. Short range potentials. *Communications in Mathematical Physics*, 61(3):285–291, 1978.

[Eps69] Edward S Epstein. Stochastic dynamic prediction. *Tellus*, 21(6):739–759, 1969.

[Eva10] L. C. Evans. *Partial Differential Equations*, volume 19. Amer. Math. Soc., second edition, 2010.

[Fal03] Kenneth Falconer. *Fractal geometry*. John Wiley & Sons, Inc., Hoboken, NJ, second edition, 2003.

[FGL$^+$96] Uriel Feige, Shafi Goldwasser, Laszlo Lovász, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *J. ACM*, 43(2):268–292, 1996.

[FH10] Søren Fournais and Bernard Helffer. *Spectral methods in surface superconductivity*, volume 77 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser Boston, Inc., Boston, MA, 2010.

[Fis78] Michael E. Fisher. Yang-Lee edge singularity and ϕ^3 field theory. *Physical Review Letters*, 40(25):1610, 1978.

[FMSG14] Manuel Fernández-Martínez and Miguel A. Sánchez-Granero. Fractal dimension for fractal structures: a Hausdorff approach revisited. *Journal of Mathematical Analysis and Applications*, 409(1):321–330, 2014.

[FMSG15] Manuel Fernández-Martínez and Miguel A. Sánchez-Granero. How to calculate the Hausdorff dimension using fractal structures. *Applied Mathematics and Computation*, 264:116–131, 2015.

[FMT89] Francisco M. Fernández, Q. Ma, and R. H. Tipping. Tight upper and lower bounds for energy eigenvalues of the Schrödinger equation. *Physical Review A*, 39(4):1605, 1989.

[FS90] Charles Fefferman and Luis Seco. On the energy of a large atom. *Bull. Amer. Math. Soc. (N.S.)*, 23(2):525–530, 1990.

[FS92] Charles Fefferman and Luis Seco. Eigenvalues and eigenfunctions of ordinary differential operators. *Adv. Math.*, 95(2):145–305, 1992.

[FS93] Charles Fefferman and Luis Seco. Aperiodicity of the Hamiltonian flow in the Thomas-Fermi potential. *Rev. Mat. Iberoamericana*, 9(3):409–551, 1993.

[FS94a] Charles Fefferman and Luis Seco. The density in a one-dimensional potential. *Adv. Math.*, 107(2):187–364, 1994.

[FS94b] Charles Fefferman and Luis Seco. The eigenvalue sum for a one-dimensional potential. *Adv. Math.*, 108(2):263–335, 1994.

[FS94c] Charles Fefferman and Luis Seco. On the Dirac and Schwinger corrections to the ground-state energy of an atom. *Adv. Math.*, 107(1):1–185, 1994.

[FS95] Charles Fefferman and Luis Seco. The density in a three-dimensional radial potential. *Adv. Math.*, 111(1):88–161, 1995.

[FS96a] Charles Fefferman and Luis Seco. The eigenvalue sum for a three-dimensional radial potential. *Adv. Math.*, 119(1):26–116, 1996.

[FS96b] Charles Fefferman and Luis Seco. Interval arithmetic in quantum mechanics. In *Applications of interval computations (El Paso, TX, 1995)*, volume 3 of *Appl. Optim.*, pages 145–167. Kluwer Acad. Publ., Dordrecht, 1996.

[GAE$^+$00] William K George, Hans Abrahamsson, Jan Eriksson, Rolf I Karlsson, Lennart Löfdahl, and Martin Wosnik. A similarity theory for the turbulent plane wall jet without external stream. *J. Fluid Mech.*, 425:367–411, 2000.

[Ger15] Klaus Gersten. The asymptotic downstream flow of plane turbulent wall jets without external stream. *J. Fluid Mech.*, 779:351, 2015.

[GG13a] Andre K. Geim and Irina V. Grigorieva. Van der Waals heterostructures. *Nature*, 499(7459):419–425, 2013.

[GG13b] T. Göckler and V. Grimm. Convergence analysis of an extended Krylov subspace method for the approximation of operator functions in exponential integrators. *SIAM J. Numer. Anal.*, 51(4):2189–2213, 2013.

[Gil03] Michael I. Gil. *Operator functions and localization of spectra*. Springer, 2003.

[GJL94] O. Golinelli, T. Jolicoeur, and R. Lacaze. Finite-lattice extrapolations for a Haldane-gap antiferromagnet. *Physical Review B*, 50(5):3037, 1994.

[GK98] Ilya Ya Goldsheid and Boris A. Khoruzhenko. Distribution of eigenvalues in non-Hermitian Anderson models. *Physical Review Letters*, 80(13):2897, 1998.

[GKKS18] Dimitrios Giannakis, Anastasiya Kolchinskaya, Dmitry Krasnov, and Jörg Schumacher. Koopman analysis of the long-term evolution in a turbulent convection cell. *Journal of Fluid Mechanics*, 847:735–767, 2018.

[GKP91] T. Geisel, R. Ketzmerick, and G. Petschel. New class of level statistics in quantum systems with unbounded diffusion. *Physical Review Letters*, 66(13):1651, 1991.

[GL81] Peter Gács and Laszlo Lovász. Khachiyan's algorithm for linear programming. In *Mathematical Programming at Oberwolfach*, pages 61–68. Springer, 1981.

[GLS88] Martin Grötschel, Lászlo Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*. Springer, 1988.

[GMvN59] H. H. Goldstine, F. J. Murray, and J. von Neumann. The Jacobi method for real symmetric matrices. *J. ACM*, 6(1):59–96, January 1959.

[Gow00] W. T. Gowers. Rough structure and classification. *Geom. Funct. Anal.*, pages 79–117, 2000.

[Gra94] Gian M. Graf. Anderson localization and the space-time characteristic of continuum states. *Journal of Statistical Physics*, 75(1-2):337–346, 1994.

[Gri12] V. Grimm. Resolvent Krylov subspace approximation to operator functions. *BIT*, 52(3):639–659, 2012.

[GS97a] Fritz Gesztesy and Barry Simon. m-functions and inverse spectral analysis for finite and semi-infinite Jacobi matrices. *J. Anal. Math.*, 73:267–297, 1997.

[GS97b] David Gottlieb and Chi-Wang Shu. On the Gibbs phenomenon and its resolution. *SIAM Rev.*, 39(4):644–668, 1997.

[GS03] V. Girardin and R. Senoussi. Semigroup stationary processes and spectral representation. *Bernoulli*, 9(5):857–876, 2003.

[GVL13] Gene H. Golub and Charles F. Van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, fourth edition, 2013.

[HAB$^+$17] Thomas Hales, Mark Adams, Gertrud Bauer, Tat Dat Dang, John Harrison, Le Truong Hoang, Cezary Kaliszyk, Victor Magron, Sean McLaughlin, Tat Thang Nguyen, Quang Truong Nguyen, Tobias Nipkow, Steven Obua, Joseph Pleso, Jason Rute, Alexey Solovyev, Thi Hoai An Ta, Nam Trung Tran, Thi Diep Trieu, Josef Urban, Ky Vu, and Roland Zumkeller. A formal proof of the Kepler conjecture. *Forum Math. Pi*, 5:e2, 29, 2017.

[Hal50] Paul R. Halmos. *Measure Theory*. D. Van Nostrand Company, Inc., New York, N. Y., 1950.

[Hal60] John H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multidimensional integrals. *Numerische Mathematik*, 2:84–90, 1960.

[Hal71] Paul R. Halmos. Capacity in Banach algebras. *Indiana Univ. Math. J.*, 20:855–863, 1970/1971.

[Hal83] F. Duncan Haldane. Nonlinear field theory of large-spin Heisenberg antiferromagnets: semiclassically quantized solitons of the one-dimensional easy-axis Néel state. *Physical Review Letters*, 50(15):1153, 1983.

[Hal05] Thomas C. Hales. A proof of the Kepler conjecture. *Annals of Mathematics (2)*, 162(3):1065–1185, 2005.

[Hal17] Paul R Halmos. *Lectures on Ergodic Theory*. Courier Dover Publications, 2017.

[Han11] Anders C. Hansen. On the solvability complexity index, the n-pseudospectrum and approximations of spectra of operators. *Journal of the American Mathematical Society*, 24(1):81–124, 2011.

[Hel13] Bernard Helffer. *Spectral theory and its applications*, volume 139 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2013.

[HHK72] R. Haydock, Volker Heine, and M. J. Kelly. Electronic structure based on the local atomic environment for tight-binding bands. *Journal of Physics C: Solid State Physics*, 5(20):2845, 1972.

[HHT08] Nicholas Hale, Nicholas J. Higham, and Lloyd N. Trefethen. Computing $\mathbf{A}^{\alpha}$, $\log(\mathbf{A})$, and related matrix functions by contour integrals. *SIAM Journal on Numerical Analysis*, 46(5):2505–2523, 2008.

[Hig05] N. J. Higham. The scaling and squaring method for the matrix exponential revisited. *SIAM J. Matrix Anal. Appl.*, 26(4):1179–1193, 2005.

[HKM16] Marijn J. H. Heule, Oliver Kullmann, and Victor W. Marek. Solving and verifying the boolean pythagorean triples problem via cube-and-conquer. In Nadia Creignou and Daniel Le Berre, editors, *Theory and Applications of Satisfiability Testing – SAT 2016*, pages 228–245, 2016.

[HN96] Naomichi Hatano and David R. Nelson. Localization transitions in non-Hermitian quantum mechanics. *Physical Review Letters*, 77(3):570, 1996.

[HO10] Marlis Hochbruck and Alexander Ostermann. Exponential integrators. *Acta Numerica*, 19:209–286, 2010.

[Hof76] Douglas R. Hofstadter. Energy levels and wave functions of Bloch electrons in rational and irrational magnetic fields. *Physical Review B*, 14(6):2239, 1976.

[HPR09] Justin Holmer, Rodrigo Platte, and Svetlana Roudenko. Blow-up criteria for the 3d cubic nonlinear Schrödinger equation. *arXiv preprint arXiv:0911.3955*, 2009.

[HR08] Justin Holmer and Svetlana Roudenko. A sharp condition for scattering of the radial 3d cubic nonlinear Schrödinger equation. *Communications in mathematical physics*, 282(2):435–467, 2008.

[HS02] Dirk Hundertmark and Barry Simon. Lieb-Thirring inequalities for Jacobi matrices. *Journal of Approximation Theory*, 118(1):106–130, 2002.

[HSYY^{+}13] B. Hunt, J. D. Sanchez-Yamagishi, A. F. Young, M. Yankowitz, Brian J. LeRoy, K. Watanabe, T. Taniguchi, P. Moon, M. Koshino, P. Jarillo-Herrero, et al. Massive Dirac fermions and Hofstadter butterfly in a van der Waals heterostructure. *Science*, 340(6139):1427–1430, 2013.

[HTF09] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer-Verlag, 2009.

[HTHK94] J. H. Han, D. J. Thouless, H. Hiramoto, and M. Kohmoto. Critical and bicritical properties of Harper's equation with next-nearest-neighbor coupling. *Physical Review B*, 50(16):11365, 1994.

[HTT^{+}09] Anthony JG Hey, Stewart Tansley, Kristin Michele Tolle, et al. *The fourth paradigm: data-intensive scientific discovery*, volume 1. Microsoft Research Redmond, WA, 2009.

[IKS18] A. Iserles, K. Kropielnicka, and P. Singh. Magnus–Lanczos methods with simplified commutators for the Schrödinger equation with a time-dependent potential. *SIAM J. Numer. Anal.*, 56(3):1547–1569, 2018.

[JCN^{+}21] Dean Johnstone, Matthew J Colbrook, Anne EB Nielsen, Patrik Öhberg, and Callum W Duncan. Bulk localised transport states in infinite and finite quasicrystals via magnetic aperiodicity. *arXiv preprint arXiv:2107.05635*, 2021.

[Joh78] Charles R. Johnson. Numerical determination of the field of values of a general complex matrix. *SIAM Journal on Numerical Analysis*, 15(3):595–602, 1978.

[JWP96] Bo-Nan Jiang, Jie Wu, and Louis A. Povinelli. The origin of spurious solutions in computational electromagnetics. *Journal of Computational physics*, 125(1):104–123, 1996.

[Kac96] Alexander Grigoryevich Kachurovskii. The rate of convergence in ergodic theorems. *Russian Math. Sur.*, 51(4):653–703, 1996.

[Kal63] Rudolf Emil Kalman. Mathematical description of linear dynamical systems. *J. Soc. Ind. Appl. Math.*, 1(2):152–192, 1963.

[Kat49] Tosio Kato. On the upper and lower bounds of eigenvalues. *Journal of the Physical Society of Japan*, 4:334–339, 1949.

[KGM08] Shachar Klaiman, Uwe Günther, and Nimrod Moiseyev. Visualization of branch points in PT-symmetric waveguides. *Physical Review Letters*, 101(8):080402, 2008.

[Kha80] Leonid G Khachiyan. Polynomial algorithms in linear programming. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 20(1):51–68, 1980.

[Kir07] Werner Kirsch. An invitation to random Schrödinger operators. *arXiv:0709.3707*, 25:1–119, 2007.

[KKKG97] R. Ketzmerick, K. Kruse, S. Kraut, and T. Geisel. What determines the spreading of a wave packet? *Physical Review Letters*, 79(11):1959, 1997.

[KKS16] Stefan Klus, Peter Koltai, and Christof Schütte. On the numerical approximation of the Perron-Frobenius and Koopman operator. *J. Comput. Dyn.*, 3(1):51–79, 2016.

[KL87] Alexander S. Kechris and Alain Louveau. *Descriptive set theory and the structure of sets of uniqueness*, volume 128 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1987.

[Kla80] M. Klaus. On the point spectrum of Dirac operators. *Helv. Phys. Acta*, 53(3):453–462 (1981), 1980.

[KLY19] K. Kormann, C. Lasser, and A. Yurova. Stable interpolation with isotropic and anisotropic Gaussians using Hermite generating function. *SIAM J. Sci. Comput.*, 41(6):3839–59, 2019.

[KM71] G. Kallianpur and V. Mandrekar. Spectral theory of stationary H-valued processes. *J. Multivar. Anal.*, 1(1):1–16, 1971.

[KM82] Werner Kirsch and Fabio Martinelli. On the spectrum of Schrödinger operators with a random potential. *Communications in Mathematical Physics*, 85(3):329–350, 1982.

[KM18] Milan Korda and Igor Mezić. On convergence of extended dynamic mode decomposition to the Koopman operator. *J. Nonlin. Sci.*, 28(2):687–710, 2018.

[KNK$^+$18] Stefan Klus, Feliks Nüske, Péter Koltai, Hao Wu, Ioannis Kevrekidis, Christof Schütte, and Frank Noé. Data-driven model reduction and transfer operator approximation. *J. Nonlin. Sci.*, 28(3):985–1010, 2018.

[KNO19] Marek Kaluba, Piotr W. Nowak, and Narutaka Ozawa. $\mathrm{Aut}(\mathbb{F}_5)$ has property (T). *Mathematische annalen*, 375(3):1169, 2019.

[Koo31] Bernard O Koopman. Hamiltonian systems and transformation in Hilbert space. *Proc. Nat. Acad. Sci. USA*, 17(5):315, 1931.

[KPG92] R. Ketzmerick, G. Petschel, and T. Geisel. Slow decay of temporal correlations in quantum systems with Cantor spectra. *Physical Review Letters*, 69(5):695, 1992.

[KR97a] Richard V. Kadison and John R. Ringrose. *Fundamentals of the theory of operator algebras. Vol. I*, volume 15 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1997.

[KR97b] Richard V. Kadison and John R. Ringrose. *Fundamentals of the theory of operator algebras. Vol. II*, volume 16 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1997.

[KR01] Denis Krutikov and Christian Remling. Schrödinger operators with sparse potentials: asymptotics of the Fourier transform of the spectral measure. *Communications in Mathematical Physics*, 223(3):509–532, 2001.

[KRH$^+$19] Austin W Kleinfelter, Russell Repasky, Nandita Hari, Stefan Letica, Vidya Vishwanathan, Lee Organski, Jon Schwaner, William N Alexander, and William J Devenport. Development and calibration of a new anechoic wall jet wind tunnel. In *AIAA Scitech 2019 Forum*, page 1936, 2019.

[KS03] R. Killip and B. Simon. Sum rules for Jacobi matrices and their applications to spectral theory. *Ann. Math.*, pages 253–321, 2003.

[KSM20] Stefan Klus, Ingmar Schuster, and Krikamol Muandet. Eigendecompositions of transfer operators in reproducing kernel Hilbert spaces. *J. Nonlin. Sci.*, 30(1):283–315, 2020.

[KST87] Mahito Kohmoto, Bill Sutherland, and Chao Tang. Critical wave functions and a Cantor-set spectrum of a one-dimensional quasicrystal model. *Physical Review B*, 35(3):1020, 1987.

[KSTV15] D. Krejčiřík, P. Siegl, M. Tater, and J. Viola. Pseudospectra in non-Hermitian quantum mechanics. *J. Math. Phys.*, 56(10):103513, 32, 2015.

[Kut84] Werner Kutzelnigg. Basis set expansion of the Dirac operator without variational collapse. *International Journal of Quantum Chemistry*, 25(1):107–129, 1984.

[Kut97] W. Kutzelnigg. Relativistic one-electron Hamiltonians for electrons only and the variational treatment of the Dirac equation. *Chem. Phys.*, 225(1-3):203–222, 1997.

[KvN32] Bernard O Koopman and John von Neumann. Dynamical systems of continuous spectra. *Proc. Nat. Acad. Sci. USA*, 18(3):255, 1932.

[Las96] Yoram Last. Quantum dynamics and decompositions of singular continuous spectra. *Journal of Functional Analysis*, 142(2):406–445, 1996.

[Law80] Eugene L Lawler. The great mathematical sputnik of 1979. *The Mathematical Intelligencer*, 2(4):191–198, 1980.

[LDBK17] Qianxiao Li, Felix Dietrich, Erik M Bollt, and Ioannis G Kevrekidis. Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the Koopman operator. *Chaos*, 27(10):103111, 2017.

[LGDV15] Stefano Longhi, Davide Gatti, and Giuseppe Della Valle. Robust light transport in non-Hermitian photonic lattices. *Sci. Rep.*, 5, 2015.

[Lie05] Elliott H. Lieb. *The stability of matter: from atoms to stars*. Springer, Berlin, fourth edition, 2005.

[LL20] C. Lasser and C. Lubich. Computing quantum dynamics in the semiclassical regime. *arXiv preprint arXiv:2002.00624*, 2020.

[Lor63] Edward N Lorenz. Deterministic nonperiodic flow. *J. Atmo. Sci.*, 20(2):130–141, 1963.

[LRF$^+$11] Liad Levi, Mikael Rechtsman, Barak Freedman, Tal Schwartz, Ofer Manela, and Mordechai Segev. Disorder-enhanced transport in photonic quasicrystals. *Science*, 332(6037):1541–1544, 2011.

[LS96] Ari Laptev and Yu Safarov. Szegő type limit theorems. *Journal of Functional Analysis*, 138(2):544–559, 1996.

[LS09] Mathieu Lewin and Éric Séré. Spectral pollution and how to avoid it. *Proceedings of the London mathematical society*, 100(3):864–900, 2009.

[LS13] J. Liesen and Z. Strakos. *Krylov subspace methods: principles and analysis*. OUP, 2013.

[LS14] Mathieu Lewin and Éric Séré. Spurious modes in Dirac calculations and how to avoid them. In *Many-Electron Approaches in Physics, Chemistry and Mathematics*, pages 31–52. Springer, 2014.

[LSN17] Jörg Liesen, Olivier Sète, and Mohamed M. S. Nasser. Fast and accurate computation of the logarithmic capacity of compact sets. *Computational Methods and Function Theory*, 17(4):689–713, 2017.

[LSY16] Lin Lin, Yousef Saad, and Chao Yang. Approximating spectral densities of large matrices. *SIAM Review*, 58(1):34–65, 2016.

[LSY$^+$19] X. Lu, P. Stepanov, Yang, et al. Superconductors, orbital magnets and correlated states in magic-angle bilayer graphene. *Nature*, 574(7780):653–657, 2019.

[Lub08a] C. Lubich. On splitting methods for Schrödinger-Poisson and cubic nonlinear Schrödinger equations. *Math. Comp.*, 77(264):2141–2153, 2008.

[Lub08b] Christian Lubich. *From quantum to classical molecular dynamics: reduced models and numerical analysis*. Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS), Zürich, 2008.

[M9̈2] V. Müller. Local behaviour of the polynomial calculus of operators. *J. Reine Angew. Math.*, 430:61–68, 1992.

[Mar10] Marco Marletta. Neumann-Dirichlet maps and analysis of spectral pollution for non-self-adjoint elliptic PDEs with real essential spectrum. *IMA J. Numer. Anal.*, 30(4):917–939, 2010.

[Mat95] Pertti Mattila. *Geometry of sets and measures in Euclidean spaces*, volume 44 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995.

[Mez05] Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlin. Dyn.*, 41(1):309–325, 2005.

[Mez13] Igor Mezić. Analysis of fluid flows via spectral properties of the Koopman operator. *Ann. Rev. Fluid Mech.*, 45:357–378, 2013.

[Mez20] Igor Mezić. Spectrum of the Koopman operator, spectral expansions in functional spaces, and state-space geometry. *J. Nonlin. Sci.*, 30(5):2091–2145, 2020.

[Mez21] I. Mezić. Koopman operator, geometry, and learning of dynamical systems. *Not. Amer. Math. Soc.*, 2021.

[MFF20] Takaaki Murata, Kai Fukami, and Koji Fukagata. Nonlinear mode decomposition with convolutional neural networks for fluid dynamics. *J. Fluid Mech.*, 882, 2020.

[Mos09] Yiannis N. Moschovakis. *Descriptive set theory*, volume 155 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, second edition, 2009.

[MQ02] Robert I. McLachlan and G. Reinout W. Quispel. Splitting methods. *Acta Numerica*, 11:341–434, 2002.

[MRT18] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. MIT press, 2018.

[NBLOLT17] Gerardo G. Naumis, Salvador Barraza-Lopez, Maurice Oliva-Leyva, and Humberto Terrones. Electronic and optical properties of strained graphene and other strained 2D materials: a review. *Reports on Progress in Physics*, 80(9):096501, 2017.

[Nev93] Olavi Nevanlinna. *Convergence of iterations for linear equations*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 1993.

[Nev95] Olavi Nevanlinna. Hessenberg matrices in Krylov subspaces and the computation of the spectrum. *Numer. Funct. Anal. Optim.*, 16(3-4):443–473, 1995.

[NFBK10] Béla Sz Nagy, Ciprian Foias, Hari Bercovici, and László Kérchy. *Harmonic Analysis of Operators on Hilbert Space*. Springer Science & Business Media, 2010.

[NGP$^+$09] C. A. H. Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim. The electronic properties of graphene. *Rev. Modern Phys.*, 81(1):109, 2009.

[Nie62] Toivo Nieminen. *A condition for the selfadjointness of a linear operator*. Suomalainen tiedeakatemia, 1962.

[Nie92] Harald Niederreiter. *Random number generation and quasi-Monte Carlo methods*, volume 63 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[NKPH$^+$14] Feliks Nuske, Bettina G Keller, Guillermo Pérez-Hernández, Antonia SJS Mey, and Frank Noé. Variational approach to molecular kinetics. *J. Chem. Theory Comput.*, 10(4):1739–1752, 2014.

[Nov11] K. S. Novoselov. Nobel lecture: Graphene: Materials in the flatland. *Reviews of Modern Physics*, 83(3):837, 2011.

[Orl64] George H. Orland. On a class of operators. *Proc. Amer. Math. Soc.*, 15:75–79, 1964.

[OT13] Sheehan Olver and Alex Townsend. A fast and well-conditioned spectral method. *SIAM Review*, 55(3):462–489, 2013.

[Pal93] C. Palencia. A stability result for sectorial operators in Banach spaces. *SIAM J. Numer. Anal.*, 30(5):1373–1384, 1993.

[Par98] Beresford N. Parlett. *The symmetric eigenvalue problem*, volume 20 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998.

[Paz12] A. Pazy. *Semigroups of linear operators and applications to partial differential equations*, volume 44. Springer Science & Business Media, 2012.

[PGY$^+$13] L. A. Ponomarenko, R. V. Gorbachev, G. L. Yu, D. C. Elias, R. Jalil, A. A. Patel, A. Mishchenko, A. S. Mayorov, C. R. Woods, J. R. Wallbank, et al. Cloning of Dirac fermions in graphene superlattices. *Nature*, 497(7451):594–597, 2013.

[Poi99] Henri Poincaré. *Les Méthodes Nouvelles de la Mécanique Céleste*, volume 3. Gauthier-Villars, 1899.

[Pok79] Andrzej Pokrzywa. Method of orthogonal projections and approximation of the spectrum of a bounded operator. *Studia Mathematica*, 65(1):21–29, 1979.

[Pol96] Alexei G. Poltoratski. On the distributions of boundary values of Cauchy integrals. *Proc. Amer. Math. Soc.*, 124(8):2455–2463, 1996.

[PSZ10] Alexei G. Poltoratski, Barry Simon, and Maxim Zinchenko. The Hilbert transform of a measure. *J. Anal. Math.*, 111:247–265, 2010.

[Pui04] Joaquim Puig. Cantor spectrum for the almost Mathieu operator. *Communications in Mathematical Physics*, 244(2):297–309, 2004.

[Put79] Calvin R. Putnam. Operators satisfying a G_1 condition. *Pacific Journal of Mathematics*, 84(2):413–426, 1979.

[Rap77] Jacques Rappaz. Approximation of the spectrum of a non-compact operator given by the magnetohydrodynamic stability of a plasma. *Numerische Mathematik*, 28(1):15–24, 1977.

[Rem98] Christian Remling. The absolutely continuous spectrum of one-dimensional Schrödinger operators with decaying potentials. *Communications in Mathematical Physics*, 193(1):151–170, 1998.

[Ros91] M. Rosenblatt. Stochastic curve estimation. In *NSF-CBMS Regional Conference Series in Probability and Statistics*. IMS, 1991.

[RS75] Michael Reed and Barry Simon. *Methods of modern mathematical physics. II. Fourier analysis, self-adjointness*. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1975.

[RS78] Michael Reed and Barry Simon. *Methods of modern mathematical physics. IV. Analysis of operators*. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1978.

[RS80] Michael Reed and Barry Simon. *Methods of modern mathematical physics. I.* Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, second edition, 1980.

[RSHSPV97] Jacques. Rappaz, J. Sanchez Hubert, E. Sanchez Palencia, and D. Vassiliev. On spectral pollution in the finite element approximation of thin elastic "membrane" shells. *Numerische Mathematik*, 75(4):473–500, 1997.

[RTN14] Pedro Roman-Taboada and Gerardo G. Naumis. Spectral butterfly, mixed Dirac-Schrödinger fermion behavior, and topological states in armchair uniaxial strained graphene. *Physical Review B*, 90(19):195435, 2014.

[Rue69] David Ruelle. A remark on bound states in potential-scattering theory. *Nuovo Cimento A (10)*, 61:655–662, 1969.

[Sal72] Norberto Salinas. Operators with essentially disconnected spectrum. *Acta Sci. Math. (Szeged)*, 33:193–205, 1972.

[Sch40] Erwin Schrödinger. A method of determining quantum-mechanical eigenvalues and eigenfunctions. *Proc. Roy. Irish Acad. Sect. A.*, 46:9–16, 1940.

[Sch60a] Julian Schwinger. The special canonical group. *Proc. Nat. Acad. Sci. U.S.A.*, 46:1401–1415, 1960.

[Sch60b] Julian Schwinger. Unitary operator bases. *Proc. Nat. Acad. Sci. U.S.A.*, 46:570–579, 1960.

[Sch01] Bernhard Scholkopf. The kernel trick for distances. *Adv. Neur. Info. Proc. Syst.*, pages 301–307, 2001.

[SDB$^+$22] Máté Szőke, William J Devenport, Aurélien Borgoltz, W Nathan Alexander, Nandita Hari, Stewart AL Glegg, Ang Li, Rahul Vallabh, and Abdel-Fattah M Seyam. Investigating the aeroacoustic properties of porous fabrics. *AIAA Journal*, pages 1–10, 2022.

[SDS$^+$09] David E Shaw, Ron O Dror, John K Salmon, JP Grossman, Kenneth M Mackenzie, Joseph A Bank, Cliff Young, Martin M Deneroff, Brannon Batson, Kevin J Bowers, et al. Millisecond-scale molecular dynamics simulations on Anton. In *Proc. Conf. High Perfor. Comput. Net., Stor. and Anal.*, pages 1–11, 2009.

[SH84] Richard E. Stanton and Stephen Havriliak. Kinetic balance: A partial solution to the problem of variational safety in Dirac calculations. *The Journal of chemical physics*, 81(4):1910–1918, 1984.

[SH98] Andrew Stuart and Anthony R Humphries. *Dynamical Systems and Numerical Analysis*, volume 2. Cambridge University Press, 1998.

[Sha08] Eugene Shargorodsky. On the level sets of the resolvent norm of a linear operator. *Bull. Lond. Math. Soc.*, 40(3):493–504, 2008.

[Sil18] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Routledge, 2018.

[Sim83] Barry Simon. Some quantum operators with discrete spectrum but classically continuous spectrum. *Annals of Physics*, 146(1):209–220, 1983.

[Sim90] Barry Simon. Absence of ballistic motion. *Communications in Mathematical Physics*, 134(1):209–212, 1990.

[Sir89] Clément Sire. Electronic spectrum of a 2D quasi-crystal related to the octagonal quasi-periodic tiling. *EPL (Europhysics Letters)*, 10(5):483, 1989.

[SK12] Petr Siegl and David Krejčiřík. On the metric operator for the imaginary cubic oscillator. *Physical Review D*, 86(12):121702, 2012.

[SL09] Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *Science*, 324(5923):81–85, 2009.

[Sma81] Steve Smale. The fundamental theorem of algebra and complexity theory. *Bull. Amer. Math. Soc. (N.S.)*, 4(1):1–36, 1981.

[Sma97] Steve Smale. Complexity theory and numerical analysis. In *Acta numerica, 1997*, volume 6 of *Acta Numer.*, pages 523–551. Cambridge Univ. Press, Cambridge, 1997.

[SP13] Christian R Schwantes and Vijay S Pande. Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *J. Chem. Theory Comput.*, 9(4):2000–2009, 2013.

[SP15] Christian R Schwantes and Vijay S Pande. Modeling molecular kinetics with tICA and the kernel trick. *J. Chem. Theory Comput.*, 11(2):600–608, 2015.

[SST03] D. Sheen, I. H. Sloan, and V. Thomée. A parallel method for time discretization of parabolic equations based on Laplace transformation and quadrature. *IMA J. Numer. Anal.*, 23(2):269–299, 2003.

[Sta65] Joseph G. Stampfli. Hyponormal operators and spectral density. *Trans. Amer. Math. Soc.*, 117:469–476, 1965.

[Sti94] Thomas J. Stieltjes. Recherches sur les fractions continues. *Ann. Fac. Sci. Toulouse Sci. Math. Sci. Phys.*, 8(4):J1–J122, 1894.

[Sto90] Marshall H. Stone. *Linear transformations in Hilbert space*, volume 15 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 1990.

[STY$^+$04] V. M. Shabaev, I. I. Tupitsyn, V. A. Yerokhin, G. Plunien, and G. Soff. Dual kinetic balance approach to basis-set expansions for the Dirac equation. *Physical Review Letters*, 93(13):130405, 2004.

[Süt89] András Sütő. Singular continuous spectrum on a Cantor set of zero Lebesgue measure for the Fibonacci Hamiltonian. *Journal of Statistical Physics*, 56(3-4):525–531, 1989.

[Sze20] Gabor Szegő. Beiträge zur Theorie der Toeplitzschen Formen. *Mathematische Zeitschrift*, 6(3-4):167–202, 1920.

[Sze04] J. Szeftel. Design of absorbing boundary conditions for Schrödinger equations in $\mathbb{R}^d$. *SIAM J. Numer. Anal.*, 42(4):1527–1551, 2004.

[Tai06] Trinh D. Tai. On the simpleness of zeros of Stokes multipliers. *Journal of Differential Equations*, 223(2):351–366, 2006.

[Tal86] J. D. Talman. Minimax principle for the Dirac equation. *Physical Review Letters*, 57(9):1091, 1986.

[TBI97] Lloyd N. Trefethen and David Bau III. *Numerical linear algebra*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

[TE05] Lloyd N. Trefethen and Mark Embree. *Spectra and pseudospectra*. Princeton University Press, Princeton, NJ, 2005.

[Tes00] Gerald Teschl. *Jacobi operators and completely integrable nonlinear lattices*, volume 72 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2000.

[TFUT91] Hirokazu Tsunetsugu, Takeo Fujiwara, Kazuo Ueda, and Tetsuji Tokihiro. Electronic properties of the Penrose lattice. I. energy spectrum and wave functions. *Physical Review B*, 43(11):8879, 1991.

[TGB$^+$14] Dimitrii Tanese, Evgeni Gurevich, Florent Baboux, Thibaut Jacqmin, Aristide Lemaître, Elisabeth Galopin, Isabelle Sagnes, Alberto Amo, Jacqueline Bloch, and Eric Akkermans. Fractal energy spectrum of a polariton gas in a Fibonacci quasiperiodic potential. *Physical Review Letters*, 112(14):146404, 2014.

[Tha92] Bernd Thaller. *The Dirac equation*. Texts and Monographs in Physics. Springer-Verlag, Berlin, 1992.

[TOD12] T. Trogdon, S. Olver, and B. Deconinck. Numerical inverse scattering for the Korteweg–de Vries and modified Korteweg–de Vries equations. *Phys. D: Nonlinear Pheno.*, 241(11):1003–1025, 2012.

[Tre19] Lloyd N. Trefethen. *Approximation Theory and Approximation Practice*, volume 164. SIAM, 2019.

[Tsy98] S. V. Tsynkov. Numerical solution of problems on unbounded domains. A review. *Appl. Numer. Math.*, 27(4):465–532, 1998.

[TT13] Alex Townsend and Lloyd N. Trefethen. An extension of Chebfun to two dimensions. *SIAM J. Sci. Comput.*, 35(6):C495–C518, 2013.

[Tur36] Alan M. Turing. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proc. London Math. Soc. (2)*, 42(3):230–265, 1936.

[TW14] L. N. Trefethen and J. A. C. Weideman. The exponentially convergent trapezoidal rule. *SIAM Rev.*, 56(3):385–458, 2014.

[Val13] Leslie Valiant. *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World*. Basic Books, Inc., New York, NY, USA, 2013.

[vdBF08] E. van den Berg and M. P. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912, 2008.

[Wal48] Hubert S. Wall. *Analytic Theory of Continued Fractions*. D. Van Nostrand Company, Inc., New York, N. Y., 1948.

[WC15] J. Wilkening and A. Cerfon. A spectral transform method for singular Sturm–Liouville problems with applications to energy diffusion in plasma physics. *SIAM J. Appl. Math.*, 75(2):350–392, 2015.

[Wei82] Michael I Weinstein. Nonlinear Schrödinger equations and sharp interpolation estimates. *Communications in Mathematical Physics*, 87(4):567–576, 1982.

[Wen96] Ernst J. Weniger. A convergent renormalized strong coupling perturbation expansion for the ground state energy of the quartic, sextic, and octic anharmonic oscillator. *Annals of Physics*, 246(1):133–165, 1996.

[Wey50] Hermann Weyl. *The theory of groups and quantum mechanics*. Dover Publications, Inc., New York, 1950.

[Wil65] James H. Wilkinson. *The algebraic eigenvalue problem*. Clarendon Press, Oxford, 1965.

[WKR15] Matthew O. Williams, Ioannis G. Kevrekidis, and Clarence W. Rowley. A data–driven approximation of the Koopman operator: Extending dynamic mode decomposition. *J. Nonlin. Sci.*, 25(6):1307–1346, 2015.

[WRK15] Matthew O Williams, Clarence W Rowley, and Ioannis G Kevrekidis. A kernel-based method for data-driven Koopman spectral analysis. *J. Comput. Dyn.*, 2(2):247, 2015.

[WT88] J. A. C. Weideman and Lloyd N. Trefethen. The eigenvalues of second-order spectral differentiation matrices. *SIAM Journal on Numerical Analysis*, 25(6):1279–1298, 1988.

[WT07] J. A. C. Weideman and L. N. Trefethen. Parabolic and hyperbolic contours for computing the Bromwich integral. *Math. Comp.*, 76(259):1341–1356, 2007.

[Zas02] George M Zaslavsky. Chaos, fractional kinetics, and anomalous transport. *Phys. Reports*, 371(6):461–580, 2002.

[Zha07] Shan Zhao. On the spurious solutions in the high-order finite difference methods for eigenvalue problems. *Computer methods in applied mechanics and engineering*, 196(49-52):5031–5046, 2007.

[Zha15] Zhimin Zhang. How many numerical eigenvalues can we trust? *Journal of Scientific Computing*, 65(2):455–466, 2015.

[ZJ00] E. S. Zijlstra and T. Janssen. Density of states and localization of electrons in a tight-binding model on the Penrose tiling. *Physical Review B*, 61(5):3377, 2000.

[Zwo99] Maciej Zworski. Resonances in physics and geometry. *Notices Amer. Math. Soc.*, 46(3):319–328, 1999.

[Zwo13] Maciej Zworski. Scattering resonances as viscosity limits. In *Algebraic and Analytic Microlocal Analysis*, pages 635–654. Springer, 2013.