The Foundations of Infinite-Dimensional Spectral Computations

Matthew J. Colbrook

St John's College
University of Cambridge

June 2020

This thesis is submitted for the degree of Doctor of Philosophy.

"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)

Declaration

This thesis is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text. Every lemma, proposition or theorem is original, except where explicitly indicated (and such results are given without proof and with a relevant citation). It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. Parts of this thesis are based on my work in a series of articles. In particular,

- Chapter 2 is based on my work in [Col19b],
- Chapter 3 is based on my work in [CRH19] and [CH19a],
- Chapter 4 is based on my work in [Col19a] and §4.5 is based on my work in [CHT20],
- Chapter 5 is based on my work in [Col19a],
- Chapter 6 is based on my work in [CH19a],
- Chapters 7 and 8 are based on my work in [Col19b],
- Chapter 9 is based on my work in [CH19b],
- Chapter 10 is based on my work in [Col19c].

Matthew J. Colbrook
Cambridge
June 2020

Abstract

Title: The Foundations of Infinite-Dimensional Spectral Computations

Author: Matthew J. Colbrook

Spectral computations in infinite dimensions are ubiquitous in the sciences. However, their many applications and theoretical studies depend on computations which are infamously difficult. This thesis, therefore, addresses the broad question,

"What is computationally possible within the field of spectral theory of separable Hilbert spaces?"

The boundaries of what computers can achieve in computational spectral theory and mathematical physics are unknown, leaving many open questions that have been unsolved for decades. This thesis provides solutions to several such long-standing problems.

To determine these boundaries, we use the Solvability Complexity Index (SCI) hierarchy, an idea which has its roots in Smale's comprehensive programme on the foundations of computational mathematics. The Smale programme led to a real-number counterpart of the Turing machine, yet left a substantial gap between theory and practice. The SCI hierarchy encompasses both these models and provides universal bounds on what is computationally possible. What makes spectral problems particularly delicate is that many of the problems can only be computed by using several limits, a phenomenon also shared in the foundations of polynomial root-finding as shown by McMullen. We develop and extend the SCI hierarchy to prove optimality of algorithms and construct a myriad of different methods for infinite-dimensional spectral problems, solving many computational spectral problems for the first time.

For arguably almost any operator of applicable interest, we solve the long-standing computational spectral problem and construct algorithms that compute spectra with error control. This is done for partial differential operators with coefficients of locally bounded total variation and also for discrete infinite matrix operators. We also show how to compute spectral measures of normal operators (when the spectrum is a subset of a regular enough Jordan curve), including spectral measures of classes of self-adjoint operators with error control and the construction of high-order rational kernel methods. We classify the problems of computing measures, measure decompositions, types of spectra (pure point, absolutely continuous, singular continuous), functional calculus, and Radon–Nikodym derivatives in the SCI hierarchy. We construct algorithms for and classify; fractal dimensions of spectra, Lebesgue measures of spectra, spectral gaps, discrete spectra, eigenvalue multiplicities, capacity, different spectral radii and the problem of detecting algorithmic failure of previous methods (finite section method). The infinite-dimensional QR algorithm is also analysed, recovering extremal parts of spectra, corresponding eigenvectors, and invariant subspaces, with convergence rates and error control. Finally, we analyse pseudospectra of pseudoergodic operators (a generalisation of random operators) on vector-valued l^p spaces.

All of the algorithms developed in this thesis are sharp in the sense of the SCI hierarchy. In other words, we prove that they are optimal, realising the boundaries of what digital computers can achieve. They are also implementable and practical, and the majority are parallelisable. Extensive numerical examples are given throughout, demonstrating efficiency and tackling difficult problems taken from mathematics and also physical applications.

In summary, this thesis allows scientists to rigorously and efficiently compute many spectral properties for the first time. The framework provided by this thesis also encompasses a vast number of areas in computational mathematics, including the classical problem of polynomial root-finding, as well as optimisation, neural networks, PDEs and computer-assisted proofs. This framework will be explored in the future work of the author within these settings.

Acknowledgements

First, I would like heartily to thank my supervisor Anders Hansen, who suggested the initial project that lead to this thesis. Ever since our initial meeting, Anders has been a constant encourager of my development in research. Anders is not only a brilliant mathematician who fearlessly straddles both pure and applied mathematics, but he is also extremely generous with his time, and I am very grateful for all of his advice. I have thoroughly enjoyed being his student whilst doing my PhD and look forward to many more years of friendship and collaboration.

Second, I have benefited greatly from the friendship and advice of other mathematicians. During my PhD, I have also had the opportunity to collaborate on projects disjoint from this thesis. Special thanks (in order of appearance) to Thanasis Fokas, Arieh Iserles, Sheehan Olver, Lorna Ayton, Alex Townsend, and Marcus Webb. To quote Arieh, "One of the great joys of doing mathematics is working with inspiring and brilliant people!" I must thank Thanasis for his encouragement, friendship and belief in my abilities as a mathematician ever since I took his Part III MMath course. He has been like a second supervisor and I have had the pleasure of working with him on projects disjoint from this thesis. Thanasis' mathematical enthusiasm and breadth are both extraordinary and infectious. I consider Arieh to be akin to a mathematical oracle, so vast is his mathematical knowledge, and I am very grateful for the many conversations over coffee and his kind patience with my questions. Thank you Sheehan for all our discussions, for sharing your extensive knowledge and for general career advice. I consider myself very lucky to be friends with such an exceptional numerical analyst. It has been a pleasure to interact with the Waves group at Cambridge and those in Lorna's group. Thank you Lorna for deepening my understanding of fluids/acoustic related topics, and for the opportunity to apply my ideas in acoustic scattering problems. I would like to especially thank Alex for his mentorship and friendship over the last year. As well as being a superb and knowledgeable mathematician, Alex is a dedicated and exceptional professional whom I admire greatly and from whom I have learnt a great deal. It has also been a joy getting to know his spectrally-accurate numerics group at Cornell. In particular, Andrew Horning with whom I have had the pleasure of developing a software package for computing spectral measures. Finally, I would like to thank Marcus for all our discussions, his advice, and his careful suggestions regarding the first chapter of this thesis.

Third, I would like to thank my family. Starting with my parents, whose sacrifice over my 26 years of existence is incalculable. Thank you for being the first to teach me about numbers (though this thesis may seem paradoxical in its lack thereof), all the help with homework when I was at school and encouraging me to pursue a degree at Cambridge. Thank you also to my siblings Stephen and Susanna. Thank you all for the love and support - I would not be in a position to write this thesis without it. Finally, special thanks are due to my wife Niamh. Not only did you kindly proof read this thesis (whilst doing a PhD of your own), but your constant love and support throughout my graduate studies have made it all possible. These last three years of marriage have been the happiest of my life. I fondly remember the first time we met and I tried to teach you quantum mechanics and spectral theory at the Bath House - I utterly failed in that regard but achieved something infinitely more valuable. If you still want to learn though, consult Chapter 3.

Contents

Chapter 1

Introduction

1.1 The Problem and Motivation

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, and compute solutions to PDEs. As such, 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 (see §7.1 for a detailed discussion of the finite section method, the most common approach which, while successful for many problems, can also fail catastrophically).

The ideas of using computational and algorithmic approaches to obtain spectral information date back to leading physicists and mathematicians such as Anderson [And58], Goldstine [GMvN59], Kato [Kat49], Murray [GMvN59], Schrödinger [Sch40], Schwinger [Sch60b, Sch60a] and von Neumann [GMvN59]. 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). 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 asymptotical behaviour of the ground state of a family of Schrödinger operators. The corresponding literature is vast, and we refer the reader to §1.3 for further comments. 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.

A reliable algorithm computing the spectrum should converge locally on compact subsets of $\mathbb{C}$ (con-

verging to the full spectrum and having no limiting points that are not in the spectrum), and guarantee that any point of the output is close to the spectrum, up to a chosen arbitrarily small error tolerance. A key question is whether 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 thesis addresses. The boundaries of what computers can achieve in computational spectral theory and mathematical physics are currently unknown, leaving many open questions that have been unsolved for decades. This thesis provides solutions to several such long-standing open problems. Mathematically determining these computational boundaries typically means the development of new algorithms that can handle problems previously out of reach, and providing mathematical proofs that the new algorithms are optimal.

Computational spectral problem

Questions concerning the foundations of computation and spectral computations have a rich history in mathematics and physics. The most well-known case is Hilbert's question regarding the existence of algorithms for decision problems [HA50] that led to Turing's seminal work [Tur36]. In spectral theory, a more recent example is the proof of the undecidability of the spectral gap [CPGW15]. Namely, one cannot construct an algorithm to determine whether a translationally invariant spin-lattice system is gapped or gapless in the thermodynamic limit. Another example is Smale's question regarding the existence of purely iterative (rational) generally convergent algorithms for polynomial root-finding [Sma85]. McMullen settled this problem as follows [McM87, McM88, Sma98]: yes, if the degree is three; no, if the degree is larger. However, in [DM89] Doyle and McMullen demonstrated a striking phenomenon: this problem can be solved in the case of the quartic and the quintic using several limits, a concept which we discuss below.

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 - see §1.3. 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 acts on $l^2(\mathbb{N})$ and is represented by an infinite matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots \\ a_{21} & a_{22} & a_{23} & \dots \\ a_{31} & a_{32} & a_{33} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{1.1.1}$$

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] with an infinite input tape corresponding to the matrix entries. 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 Chapter 9 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.

A key step in addressing the computational spectral problem was made 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

$$\lim_{n_3\to\infty} \lim_{n_2\to\infty} \lim_{n_1\to\infty} \Gamma_{n_3,n_2,n_1}(A) = \mathrm{Sp}(A) := \{z \in \mathbb{C} : (A - zI)^{-1} \text{ does not exist as a bounded operator}\}.$$

In other words, the process uses three successive limits. This result has given rise to the solvability complexity index (SCI). Informally, this can be described as the number of successive limits needed to solve a computational problem, a measure of its difficulty (see 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. The use of three limits in the algorithm of [Han11] is sharp if we consider the whole class of bounded operators, meaning 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 problems of descriptive set theory within the SCI hierarchy - see Chapter 2. A three limit algorithm is impossible to implement on a finite machine, and hence the result of [Han11] cannot be used for real-life numerical computation.

The fact that spectral problems are so high up in the SCI hierarchy poses a severe problem in applications: how can we guarantee that the outputs of numerical simulations converge and are sound? Fortunately,

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

there is another class in the SCI hierarchy (developed in §2.2.1): 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). This notion is explained further with a simple example below. One of the most important results of this thesis (Chapter 3) is that under very general assumptions, the spectral problem lies in Σ_1. We provide a set of algorithms that converge to the spectrum under mild assumptions which hold in the majority of applications. No previous algorithm converges in this generality, even for the case of general one-dimensional discrete Schrödinger operators. Furthermore, the algorithms converge with Σ_1 error control, and we show that this is sharp, realising the boundary of what digital computers can achieve. Finally, the algorithms are efficient and parallelisable.

For the simplest case of bounded operators $A \in \mathcal{B}(l^2(\mathbb{N}))$, this result can be understood as follows. 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

$$\mathrm{dist}(z, \mathrm{Sp}(A)) \leq E_n(A; z) \quad \forall z \in \Gamma_n(A) \quad \text{and} \quad \lim_{n\to\infty} \sup_{z\in\Gamma_n(A)} E_n(A; z) = 0. \tag{1.1.2}$$

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 §3.4). As an example, Figure 1.1 shows approximate states computed by the algorithm for the Penrose Laplacian, the canonical model of a 2D quasicrystal (see also §3.6.1). 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

$$d_{\mathrm{H}}(\Gamma_n(A), \mathrm{Sp}(A)) \leq E_n(A) \tag{1.1.3}$$

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

$$A = \begin{pmatrix} a_1 & & & \\ & a_2 & & \\ & & a_3 & \\ & & & \ddots \end{pmatrix}, \qquad a_j \in \mathbb{C}. \tag{1.1.4}$$

Since an algorithm can only deal with a finite amount of information at any one time (i.e. finitely many of the a_i - see §2.1), it is clear that the problem of computing the spectrum $\mathrm{Sp}(A)$ cannot be done with error control in the sense of (1.1.3). However, one can simply choose an algorithm Γ_n to collect $\{a_j\}_{j=1}^n$ and then one trivially has that $\Gamma_n(A) \to \mathrm{Sp}(A)$ as $n \to \infty$. We also clearly have the extra feature that

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

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

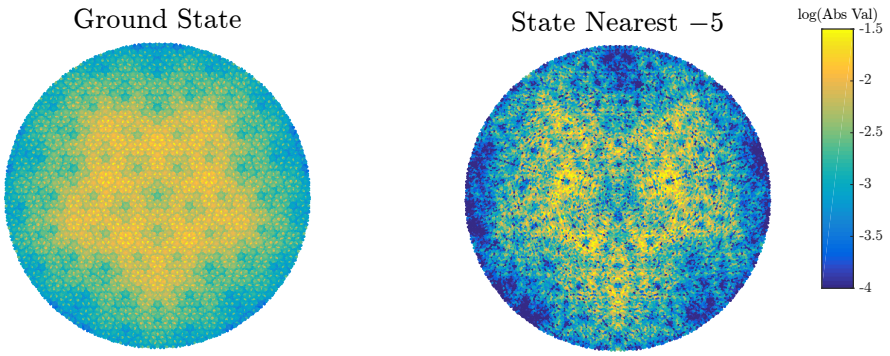


Figure 1.1: The ground 'state' for the Penrose Laplacian (from the cover of Physical Review Letters Volume 122, Issue 25 [CRH19]) and an approximate state corresponding to energy nearest -5. The algorithm allows us to choose which states to compute without direct diagonalisation. It should be emphasised that we are not necessarily approximating eigenvectors since the spectrum may not consist solely of eigenvalues.

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. The results of this thesis extend this type of convergence (up to an arbitrarily small user-chosen error tolerance given by the E_n in (1.1.2)) 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. There are special cases where the stricter form of error control in the sense of (1.1.3) is possible, such as finite rank perturbations of self-adjoint tridiagonal Toeplitz operators [WO17]. However, in general, such results require a large amount of structure.

Beyond spectra: a new computational paradigm

In order to classify and understand the difficulty of computational problems and develop techniques for their solution, one must go beyond the standard philosophy of numerical analysis. Many computational problems are solved as follows: a sequence of approximations is created by an algorithm, and the solution to the problem is the limit of this sequence. However, as discussed above, this is impossible for the general spectral problem and many other problems in computational mathematics. To deal with this, we use the SCI hierarchy. Current hierarchies in logic and computer science, such as the arithmetic hierarchy for sets of integers, are insufficient for such classifications. Hence, in order to establish the boundaries of what computers can achieve in the sciences, the SCI hierarchy is needed. Many existing foundational problems also become results in the SCI hierarchy.

The framework provided by this thesis encompasses a vast number of areas in computational mathematics. Establishing the boundaries of what computers can do in spectral theory is related to Smale's comprehensive programme concerning the foundations of computational mathematics initiated in the 1980s. This thesis closes the substantial current gap between the abstract theory and applications, and the framework is somewhat different from that of Smale's programme. Some of the other areas encompassed can be found in §2.5 and include the classical problem of polynomial root-finding (and its curious resolution by Doyle and McMullen), optimisation, neural networks, PDEs and computer-assisted proofs. This last point is 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]

Computer-assisted proofs are impossible to ignore, with 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. We provide some of the first results in the corresponding infinite classification theory.

Outline of chapter

The rest of this chapter is as follows. In §1.2 we summarise the contributions of the thesis. A discussion of relations to previous work is given in §1.3, and we finish the chapter with a summary of basic notation.

Finally, this thesis is written with both pure and applied mathematicians in mind, a reflection of the true cross-disciplinary flavour of the subject of infinite-dimensional spectral theory (which has its roots in the physical theory of quantum mechanics and Hilbert's work on integral equations, blossoming into one of the most beautiful and technical areas of mathematics). Throughout, standard graduate-level functional analysis and numerical analysis are assumed, though this thesis is mostly self-contained.

1.2 Summary of Thesis

For a lookup table of the computational spectral problems addressed in this thesis, with theorem and page numbers, we refer the reader to the concluding remarks on page 263. This thesis is split into three parts:

Part I solves the computational spectral problem, dating back to the work of Szegő [Sze20] and Schwinger [Sch60b, Sch60a], and summarised in the above quotation of Arveson. We show how to compute spectra (and pseudospectra) of a very large class of operators (both discrete operators and partial differential operators) with error control in the above Σ_1 sense. We then show how to 'diagonalise' normal operators (including unbounded) whose spectra are subsets of regular enough Jordan curves (such as self-adjoint and unitary operators) via algorithms that compute spectral measures and spectral decompositions. An example, demonstrating the efficiency of the new methods for magneto-graphene is shown in Figure 1.2.

Part II goes beyond the spectrum to algorithms that compute further spectral properties. As well as computing the spectrum, scientists may want to determine features of the spectrum such as its Lebesgue measure or fractal dimension, different types of spectral radii and numerical ranges, detect band gaps, or compute capacity, spectral gaps, discrete spectra etc. We use the resolvent norm (and generalisations) to develop the first algorithms that compute these quantities and many others, and prove that our methods are sharp in the SCI hierarchy. We also prove the curious result that detecting the failure of the finite

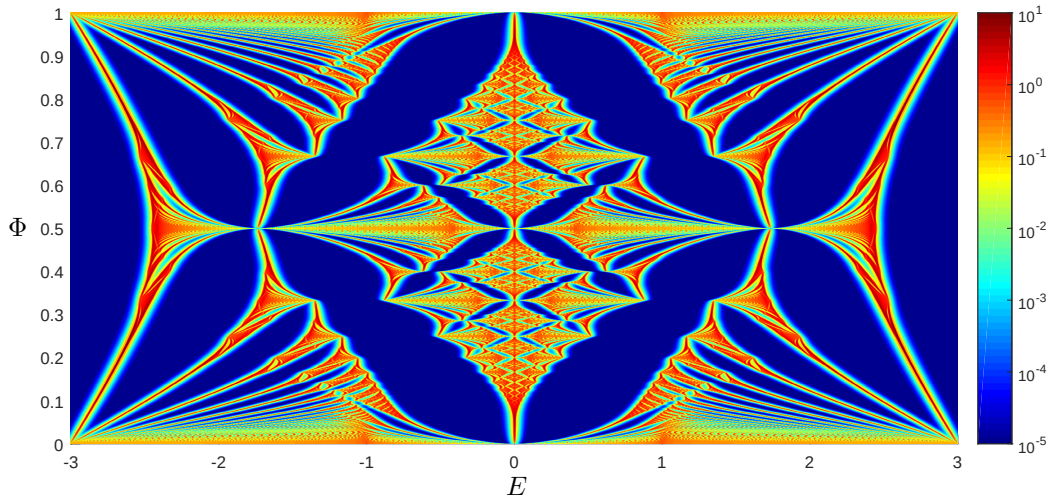


Figure 1.2: Radon–Nikodym derivative (`log10` scale) of the measure for various magnetic field strengths Φ. The axis label E (energy) stands for the spectral parameter. The Radon–Nikodym derivative is computed to high precision using a fourth-order kernel method developed in Chapter 4.

section method[6] (computing an error flag) is strictly harder than computing the spectrum. All of these problems have strong physical motivations and are themselves important open problems in the spectral theory community.

Part III provides examples where classical finite-dimensional algorithms can be extended to infinite dimensions in a useful manner. These provide further classifications in the SCI hierarchy, related to the classical finite section method. First, we develop results connected to the infinite-dimensional QR algorithm, and then we prove convergence of pseudospectra of periodic finite sections for pseudoergodic operators.

To classify the computational problems addressed in this thesis, we use the SCI hierarchy mentioned above and developed in Chapter 2. The computational spectral problem becomes an infinite classification theory, and there will, necessarily, have to be many different types of algorithms. Characterising the hierarchy will yield a myriad of different approaches, as different structures on the various classes of operators will require specific algorithms. We now summarise each chapter.

Chapter 2: The Solvability Complexity Index

In this chapter, we first summarise the basic SCI hierarchy as it already appears in the literature, and then extend the hierarchy to include notions of error control. This general framework goes beyond spectral theory, with applications in machine learning, optimisation, PDEs and computer-assisted proofs (see §2.5). We discuss how all of the algorithms in this thesis can be made to work using just arithmetic operations over the rationals $\mathbb{Q}$, with inexact input, and in a recursive manner.[7] This circumvents the current lack of a universally agreed definition of recursivity for algorithms over the fields $\mathbb{R}$ or $\mathbb{C}$. Furthermore, the proven lower bounds in this thesis hold in any model of computation. In a special case (which does not hold in general), we provide a link between the SCI hierarchy and the Baire hierarchy from descriptive set theory. This allows the construction of combinatorial problems arbitrarily high up in the SCI hierarchy, regardless of

[6] In its simplest form for operators given by (1.1.1), this corresponds to computing the spectrum of the upper-left $n \times n$ submatrix of A. Even in the case of tridiagonal self-adjoint operators, this does not converge due to 'spectral pollution', the appearance of persistent eigenvalues in gaps of the essential spectrum that have nothing to do with the spectrum of the full infinite-dimensional operator.

[7] This also allows their use for computer-assisted proofs and/or implementation using interval arithmetic [Tuc11].

the model of computation (arithmetical, radical, general etc.). By embedding these combinatorial problems into spectral problems, this provides the first technique for dealing with problems that have SCI greater than three, and also greatly simplifies the proofs of results lower down in the SCI hierarchy. We emphasise, however, that this thesis is not a thesis on logic or descriptive set theory - the contents of this chapter are self-contained.

Part I: Spectra, Spectral Measures and Spectral Decompositions

Chapter 3: Computing Spectra with Error Control

This chapter, based on [CRH19] and [CH19a], settles the long-standing problem of computing spectra (see Arveson's quotation in §1.1). We construct an algorithm computing spectra (and pseudospectra) of many operators, including non-normal operators, with rigorous error control in the Σ_1 sense. This is done both in the discrete infinite matrix setting (allowing unbounded operators defined on graphs or lattices), and also for partial differential operators. In the self-adjoint (or normal) case, the algorithm provides 'approximate states'. We also consider the decision problem of deciding if a non-empty compact set intersects the spectrum. The algorithms presented are optimal in the sense of the SCI hierarchy described in §2.2, and converge whilst also resolving the issue of spectral pollution discussed further in §7.1 and §7.3.2. We finish by showing that the new class of algorithms are efficient, as well as being completely parallelisable. Examples include a two-dimensional Penrose tile (a model of a quasicrystal), non-Hermitian Hamiltonians in superconductor theory and optics, and partial differential operators such as Schrödinger operators on unbounded domains.

Chapter 4: Computing Spectral Measures

In this chapter, we provide the first general[8] set of algorithms for the computation of spectral measures, as given by the classical spectral theorem, for a large class of self-adjoint and unitary operators (and discuss extensions to more general normal operators). This is an infinite-dimensional analogue of computing eigenvectors,[9] and 'diagonalises' the operator as an integral, thus resolving the diagonalisation problem discussed by Arveson in §1.1. We also consider the computation of the functional calculus, and the Radon–Nikodym derivative of the absolutely continuous part of the measure. We discuss how to accelerate convergence locally for smooth enough measures using different rational kernels with vanishing moments (which lend themselves to computations with infinite-dimensional operators). Under certain assumptions, this also allows computation with error control. The new algorithms are parallelisable, allowing large scale computations. Examples demonstrated include orthogonal polynomials on the real line (recovering the measure from their recurrence relations), a model of magneto-graphene that demonstrates high-resolution computation and the avoidance of spectral pollution, fractional diffusion on a quasicrystal and the solution of infinite-dimensional evolution equations with error control. Partial differential operators on the continuum are also studied, and the results of this chapter carry over by employing spectral methods to solve the relevant PDEs corresponding to the resolvent. As an example, we study a very efficient numerical method to compute highly oscillatory bound states of the Dirac operator whilst avoiding spectral pollution (this is important in computational chemistry).

[8] Although there is a rich literature on the theory of spectral measures, most of the efforts to develop computational tools have focused on specific examples where analytical formulas are available, or perturbations thereof.

[9] Of course eigenvectors exist in the infinite-dimensional case, but not all of the spectrum consists of eigenvalues. The projection-valued measure generalises the notion of projections onto eigenspaces.

Chapter 5: Computing Spectral Type

This chapter complements Chapter 4 and classifies the absolutely continuous, singular continuous and pure point parts of the spectrum in the SCI hierarchy. These different sets often characterise different physical properties in quantum mechanics, and we provide the first set of algorithms that can compute these quantities under general conditions. The impossibility results hold in general, even when restricted to tridiagonal operators, and even for structured operators such as bounded discrete Schrödinger operators on the lattices $\mathbb{N}$ or $\mathbb{Z}^d$.

Part II: Beyond Spectra

Chapter 6: Discrete Spectra and Spectral Gap

This chapter develops new algorithms for computing the discrete spectrum, multiplicities and eigenspaces of various classes of normal operators. Of course, a vast number of algorithms exist that compute eigenvalues of operators (even in infinite dimensions), but the algorithms of this chapter are the first that separate the discrete spectrum from the essential spectrum. We also provide SCI classifications of the decision problem of determining if the discrete spectrum is empty, and the spectral gap problem (related to the dichotomy between the discrete and essential spectrum and motivated from physical applications). For this last problem, we consider the infinite-dimensional version, as well as an extension to classifying the geometric/algebraic properties of the bottom of the spectrum. Finally, the effectiveness of the algorithm computing discrete spectra and eigenvectors is demonstrated.

Chapter 7: Geometric Features and Detecting Finite Section Failure

A highlight of this chapter is the proof that detecting the failure of finite section (computing an error flag) is harder than computing the spectrum itself (the problem solved in Chapter 3). This also settles the open problem on computing or detecting gaps in the essential spectrum of self-adjoint operators, a problem which has received considerable attention in the community. Furthermore, we classify various types of spectral radii, polynomial operator norms and capacity (useful for the analysis of Krylov numerical methods) in the SCI hierarchy for different classes of operators. 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. The results of this chapter (other than the spectral radius for normal operators) all present the first algorithms computing their corresponding spectral properties. Finally, we demonstrate the effectiveness of the algorithms with numerical examples for spectral radii, essential numerical ranges and capacity of spectra.

Chapter 8: Lebesgue Measure and Fractal Dimensions of Spectra

In this chapter, we consider the open problems 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. We provide the first algorithms solving these computational problems, with classifications in the SCI hierarchy. Numerical evidence is given that a portion of the spectrum of a two-dimensional model of a quasicrystal has fractal dimension approximately 0.8.

Part III: Extensions of Classical Finite-Dimensional Algorithms

Chapter 9: The Infinite-Dimensional QR Algorithm

In this chapter, we discuss how the most famous finite-dimensional algorithm, the QR algorithm, can be extended to infinite dimensions. The infinite-dimensional QR (IQR) algorithm is at least thirty years old (dating back to the work of Deift, Li and Tomei [DLT85], see also the work of Hansen [Han08b, Han08a]), but there is little existing analysis. We provide new convergence theorems for the IQR algorithm with convergence rates and error control. The results concern eigenvalues, eigenvectors and invariant subspaces (including non-normal operators). We prove that for infinite matrices with finitely many non-zero entries in each column, the IQR algorithm can be executed exactly, and that for general invertible operators, it can be executed with error control. We provide new classification results for the SCI hierarchy: Δ_1 classification for the extremal part of the spectrum and dominant invariant subspaces, and Σ_1 results for spectra of certain classes of compact operators (the general spectral problem for compact operators is not in Σ_1). We demonstrate the IQR algorithm and new convergence results on a variety of difficult problems. In some cases, the IQR algorithm performs much better than predicted by our theory, working on much larger classes of operators. Hence, we are left with many open problems on the theoretical understanding of the potential of this algorithm.

Chapter 10: Pseudoergodic Operators and Finite Section

In this chapter, we examine the so-called 'pseudoergodic' class of operators (a well-studied class encompassing generalisations of many random and non-normal operators in applications). We prove that pseudospectra of finite sections with periodic boundary conditions converge to the pseudospectrum of the full infinite-dimensional operator as the truncation parameter tends to infinity. This holds in any lattice dimension, and for any vector-valued l^p space with $p \in [1, \infty]$. Our results can be considered as a generalisation of the well-known classical result for banded Laurent operators and their circulant approximations. In terms of the SCI hierarchy, this gives a Σ_1 classification for the pseudospectral problem.

1.3 Relations to Previous Work

The results presented in this thesis follow in the long tradition of infinite-dimensional spectral computations. This field contains a vast literature that spans more than half a century, so we can only cite work that has had the most influence on the author. We split the comments into three categories: spectral computations, numerical approaches, and foundations of computational mathematics and computer-assisted proofs.

Spectral computations: We have already mentioned the work of Anderson, Digernes, Goldstine, Kato, Murray, Schrödinger, Schwinger, Varadarajan, Varadhan, von Neumann and Weyl [And58, GMvN59, Kat49, Sch40, Sch60b, Sch60a, Wey50, DVV94]. The results of [DVV94] yield an algorithm that converges in one limit without any form of error control. However, Chapter 3 extends these works considerably by providing a Σ_1 classification, and for a much broader class of operators. Not only is the Σ_1 classification sharp in the SCI hierarchy, but it also provides the useful, practical result of error control.

Arveson [Arv94a, Arv93b, Arv94b, Arv93a] helped pioneer the combination of spectral computations and C^*-algebra techniques (which dates back to the work of Böttcher and Silbermann [BS83]). Part of

his work considered spectral densities, with weak* convergence to measures whose support is the essential spectrum, also related to Szegő's work [Sze20] on finite section approximations. Similar results are also obtained by Laptev and Safarov [LS96]. These results motivated the work in Chapter 4 and are related to the density of states studied in mathematical physics (discussed in detail at the end of §4.1.2). Note that we compute the spectral measures, which contain more spectral information than the density of states (which ignores, for example, discrete spectra below the essential spectrum). For instance, [Arv94a] considers a method to locate essential spectra via *"weight[ing] the count of eigenvalues in a way which eliminates spurious ones"*. However, such an approach causes the discrete spectrum to be ignored. Our results extend the previous work above by showing that it is indeed possible to recover the full spectral measure, which is supported on the spectrum, and study other aspects such as Radon–Nikodym derivatives, spectral/measure decompositions and the functional calculus. The results in Chapter 6 also show how to recover the discrete spectrum.

There is a large physics literature on the spectral gap problem, a problem we address in Chapter 6. The spectral gap problem is related to the Haldane conjecture [Hal83], which remains unsolved despite numerical evidence [GJL94]. Another important related problem is the Yang–Mills mass gap problem [BCD$^+$06]. The seminal paper by Cubitt, Perez–Garcia and Wolf [CPGW15] shows that the spectral gap problem is undecidable (not computable in the sense of Turing) when considering the thermodynamic limit of finite-dimensional Hamiltonians. In their conclusion, the authors note, *"Thus, any method of extrapolating the asymptotic behaviour from finite system sizes must fail in general."* This comment also serves as a warning for other spectral problems, where, in the literature, it is often wrongly assumed that a large system size captures the infinite-dimensional operator. We note that there is a subtle difference between the thermodynamic limit studied in [CPGW15] and the viewpoint of infinite-dimensional operators in this thesis. We study the infinite-dimensional version of the problem, determining the existence of a gap for Hamiltonians on a separable Hilbert space. We prove that the problem generically requires two limits in the SCI hierarchy, and hence our results can be considered as an extension of [CPGW15].

The finite-section method, intensely studied for spectral computation and often viewed in connection with Toeplitz theory, is very similar to Schwinger's idea of approximating in a finite-dimensional subspace. Typically, when applied to appropriate subclasses of operators, finite section approaches yield algorithms with no form of error control. The reader may wish to consult the pioneering work by Böttcher [Böt96, Böt94], Böttcher and Silberman [BS99], Böttcher, Brunner, Iserles and Nørsett [BBIN10], Brunner, Iserles and Nørsett [BIN11]. Some of these papers also discuss the failure of the finite section approach for certain classes of operators, see also the work of Hansen [Han10, Han08b]. An important result is that of Shargorodsky [Sha13] demonstrating that second order spectra methods [Dav98] (a variant of the finite section method) do not in general recover the whole spectrum. All of these have motivated the work on the problem of spectral pollution in Chapter 7, where we show that it is in general very difficult to detect the failure of the finite section method. Chapter 7 helps explain the richness of results for specific subclasses of operators regarding the finite section method.

Chapter 8 is motivated in part by recent progress in the field of Schrödinger operators with random or almost periodic potentials. For example, relevant work includes that of Avila et. al. [Avi09, Avi08, AJ09, AK06, AV07], Puig [Pui04] and Sütő [Süt89] (see [EDML06, EDML08] for numerical work for higher dimensional versions of the Fibonacci Hamiltonian) on specific examples of operators, including Cantor-like spectra. Numerical studies of fractal dimensions of spectra include the work of Han, Thouless, Hiramoto

and Kohmoto on Harper's equation [HTHK94] and Ketzmerick, Kruse, Kraut and Geisel on wavepacket spreading [KKKG97] (for many more references connected to this paper, see [KKL03]). Another well-studied area where fractal spectral properties appear is optics. For example, following the analytical and numerical work of Berry and coauthors [Ber01, BSVS01, Ber04], the fractal structure of modes of non-Hermitian operators are studied in laser theory [RGSE18, NYWM01]. Probably the most famous example of the Lebesgue measure of spectra is the formula in (8.4.1) for the almost Mathieu operator (the case of $\lambda = 1$ was one of Simon's problems [Sim00]), which was conjectured based on numerical evidence in the work of Aubry and André [AA80]. Following this paper, there have been many further numerical studies, for example, the work of Thouless [Tho83, Tho90] and Thouless and Tan [TT91]. Numerical studies of such operators typically look at periodic approximates, and computing the Lebesgue measure of periodic approximates of tridiagonal operators lies in Δ_1. In contrast, the tools we develop are much more general and do not assume such structure. A verification of our algorithms for the almost Mathieu operator is presented in §8.4.1. The almost Mathieu operator is only one of many operators with numerical studies of the Lebesgue measure of their spectra. For others, see, for example, the references in [AJM17, BS91, Sir89]. Whilst results are known for specific examples such as the almost Mathieu operator or the Fibonacci Hamiltonian, the problems of computing the Lebesgue measure and fractal dimensions of spectra remain open in the general case (see remarks in [DGS15] and references therein). Our results show the boundaries of what can be achieved numerically for different classes of operators.

The IQR algorithm provides another approach to spectral computations, which can be seen as a generalisation of the finite section method. The IQR algorithm was first studied in connection with Toda flows by Deift, Li and Tomei [DLT85] (covering self-adjoint infinite matrices with real entries). Despite being purely functional analytic and ignoring implementation issues, these results form some of the basic fundamentals of the IQR algorithm and provide a beautiful geometric interpretation. A convergence result for eigenvectors corresponding to eigenvalues outside the essential numerical range for normal operators was given in [Han08b]. However, this paper did not consider convergence rates, actual numerical calculation nor any classification results (implementation for banded operators was, however, given in [Han08a]). The results of Chapter 9, therefore, provide a significant step in the analysis of the IQR algorithm by showing that it can be implemented for invertible operators, and giving convergence results (with convergence rates and error control) for eigenvalues, eigenvectors and invariant subspaces (including non-normal operators).

Chapter 10 is motivated, in particular, by [Böt94] which shows that pseudospectra converge for truncated Wiener–Hopf operators and Toeplitz operators with piecewise continuous symbols, and further results concerning Toeplitz operators [BG05, BS99, Böt96]. In some sense, Chapter 10 is complementary by studying a generalisation of Toeplitz operators, but now requiring the operators to be banded (or banded in each lattice dimension). The result we prove was conjectured (for a one-dimensional tridiagonal case) in [DNS99], but has been an open problem since.

Finally, the work of Zworski [Zwo13, Zwo99] on computing resonances can be viewed in terms of the SCI hierarchy. In [Zwo13], the computational approach is based on expressing resonances as limits of non-self-adjoint spectral problems. This gives a two limit process, and hence fits directly into the SCI hierarchy. Resonances provide a way of studying the time evolution of quantum systems. Another approach, based on the new algorithms of Chapter 4, is discussed in §4.6.2. The recent work of Ben–Artzi, Marletta and Rösler [BAMR20a, BAMR20b] on computing resonances is also formulated in terms of the SCI hierarchy, though the results of [BAMR20a, BAMR20b] do not allow error control at the time of writing.

Numerical methods: Numerically, the point of view in this thesis is closest to the work of Olver, Townsend and Webb on practical infinite-dimensional linear algebra [OT14, OT13, Olv18, OW18, WO17]. This work includes efficient codes, such as the infinite-dimensional QL (IQL) algorithm [Web17], as well as theoretical results (see also the infinite-dimensional version of the FEAST algorithm in [HT20]). The key point to note is that all of these methods, including our own, deal with infinite-dimensional operators directly, rather than the discretise-then-solve paradigm that pervades previous numerical approaches. The set of algorithms this thesis provides can be considered as new members within the growing family of infinite-dimensional techniques.

The IQL algorithm is rather different from the IQR algorithm studied in Chapter 9. The IQL algorithm requires an analytical QL factorisation for the 'tail' of the operator (for instance Toeplitz-plus-finite-rank Jacobi operators). Such a QL factorisation does not always exist for bounded operators. The results of [Web17] complement Chapter 9 in the following sense. For a bounded Jacobi operator J such that there is an eigenvalue λ_0 with $0 < |\lambda_0| < \eta := \min_{\lambda \in \mathrm{Sp}(J)\setminus\lambda_0} |\lambda|$, the IQL algorithm converges in the top-left entry to λ_0 at rate $\mathcal{O}(|\lambda_0/\eta|^n)$. In other words, the IQL algorithm gives information on the part of the spectrum nearest the origin, whereas the IQR algorithm in Chapter 9 gives information on the extremal parts of the spectrum. There are advantages and disadvantages to both approaches. For example, our analysis of the IQR algorithm gives little information inside the essential numerical range, except in special cases. However, the IQL algorithm can deal with discrete spectra near the origin, even if the eigenvalues are surrounded by essential spectra. On the other hand, for operators which do not have a large amount of structure, computing the QL decomposition (if it exists) for the IQL algorithm is extremely difficult, whereas the IQR algorithm does not suffer from this setback.

The work of [WO17] is of particular relevance to Chapters 4 and 5. In [WO17], the authors studied Jacobi operators that are compact perturbations of Toeplitz operators through connection coefficients. Their results can be stated in terms of the SCI hierarchy:

- If the perturbation is finite rank (and known), the pure point spectrum can be computed in one limit with Δ_1 error control, and the absolutely continuous part of the spectral measure can be computed in finite time (the absolutely continuous spectrum is known analytically).
- If the perturbation is compact, with a known rate of decay at infinity, then the full spectrum can be computed in one limit with Δ_1 error control.

Chapters 4 and 5 extend the work of [WO17] by considering operators more general than tridiagonal compact perturbations of Toeplitz operators, allowing operators to be unbounded, and building algorithms that are arithmetic and can cope with inexact input. At the price of this greater generality, some of the objects we study are not computable with error control. However, they are still computationally useful as we shall demonstrate (many of them can be computed with one limit). Moreover, with certain regularity assumptions (see §4.5.2), we can compute spectral measures with error control. Our methods are also entirely different and rely on estimating the resolvent operator with error control. We also leverage this to construct methods with arbitrarily high orders of convergence.

Foundations and computer-assisted proofs: Smale's seminal work [Sma81, Sma97] and his programme on the foundations of computational mathematics and scientific computing initiated the pioneering work by McMullen [McM87, McM88, Sma98], and Doyle and McMullen [DM89] on polynomial root-finding. These are classification results in the SCI hierarchy.

Regarding the SCI hierarchy itself, this first appeared in [Han11] where it was shown that the $\mathrm{SCI} \leq 3$ for the computation of spectra of general operators (without any structural assumptions). This algorithm first computed pseudospectra with two limits, and then shrunk the pseudospectrum to the spectrum to produce a three limit algorithm. In Chapter 3, we show how this can be bypassed (for a very large class of operators) with an algorithm that converges in one limit with Σ_1 error control.

The number of examples of computer-assisted proofs in the literature is substantial and growing fast, so we can only mention a few examples (see also §2.5.1). In most cases, in order to prove that the computational proof is 100% accurate, one implicitly has to prove a classification in the SCI hierarchy. The work by Fefferman and Seco [FS90, FS92, FS93, FS94b, FS94c, FS95, FS96b, FS96a, FS94a] involves a Σ_1^A classification (where the superscript refers to restricting to purely arithmetical operations). Similarly, Hales' Flyspeck programme [Hal05, HAB+17], which provided a computer-assisted proof of Kepler's conjecture, relies on a Σ_1^A classification. Both of these examples are computer-assisted proofs done via non-computable problems. There are also computer-assisted proofs based on Δ_1^A classifications. For instance, the work of Gabai, Meyerhoff, and Milley [GMM09] on hyperbolic three-manifolds. Moreover, recent results using computer-assisted proofs in spectral theory include the work of Brown, Langer, Marletta, Tretter, and Wagenhofer [MBLM+10] and Bögli, Brown, Marletta, Tretter and Wagenhofer [BBM+14].

1.4 Notation

We end this chapter by listing the basic standard notation used in this thesis. 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
$\sigma_1(C)$	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 $\{z \in \mathbb{C} : R(z,A) \text{ does not exist as a bounded operator}\}$
$\mathrm{Sp}_\epsilon(A)$	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 $\{z \in \mathbb{C} : A - zI \text{ is not Fredholm}\}$
$r_{\mathrm{ess}}(A)$	essential numerical radius of operator A defined as $\sup\{\|z\| : z \in \mathrm{Sp}_{\mathrm{ess}}(A)\}$
$W(A)$	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 $\bigcap_{K \text{ compact}} \mathrm{cl}(W(A+K))$

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

$$\mathrm{Sp}_\epsilon(A) = \{z \in \mathbb{C} : \|R(z, A)\|^{-1} \leq \epsilon\}, \tag{1.4.1}$$

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

$$d_{\mathrm{H}}(\mathcal{S}, \mathcal{T}) = \max\left\{\sup_{\lambda\in\mathcal{S}} \mathrm{dist}(\lambda, \mathcal{T}), \sup_{\lambda\in\mathcal{T}} \mathrm{dist}(\lambda, \mathcal{S})\right\}, \tag{1.4.2}$$

where $\mathrm{dist}(\lambda, \mathcal{T}) = \inf_{\rho\in\mathcal{T}} |\rho - \lambda|$. 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 2

The Solvability Complexity Index

This chapter discusses the Solvability Complexity Index (SCI) hierarchy, which is needed to show that the algorithms in later chapters realise the boundary of what digital computers can achieve. The SCI was first introduced in [Han11] where it was shown that the $\mathrm{SCI} \leq 3$ for the computation of spectra of general operators (see §1.1 and §1.3). However, for operators with more structure, the spectral problem is fortunately much lower in the SCI hierarchy, with the first SCI-sharp algorithms appearing in [CRH19], which are the topic of Chapter 3. We have sought to place all of the results concerning the hierarchy itself in one chapter, for ease of reference. It should be mentioned that this is not a thesis on logic or descriptive set theory, and the contents of this chapter are self-contained.

This chapter begins with the basic set-up of the SCI in §2.1. Notions of error control are discussed in §2.2, as well as properties of the refined structure. A note on Turing towers and realisable computation is given in §2.3. This essentially says that all of the algorithms constructed in this thesis can be made recursive (in the classical Turing sense) with restrictions to arithmetic operations over $\mathbb{Q}$ and inexact input. However, the proven lower bounds in this thesis hold in any model of computation. In other words, it does not matter which model of computation one uses for a definition of 'algorithm', from a classification point of view they are equivalent for these infinite-dimensional spectral problems. This result is satisfying since it avoids the current lack of a universally agreed definition of recursivity for algorithms over the fields $\mathbb{R}$ or $\mathbb{C}$. In §2.4, we link the SCI hierarchy to the Baire hierarchy (in a special case), in order to provide combinatorial array problems arbitrarily high up in the SCI hierarchy. This is the most technical part of the chapter and we use some tools from descriptive set theory. These results will be used for proving lower bounds, where we typically embed such a problem within the spectral problem of interest. Any result in this thesis with $\mathrm{SCI} \geq 3$ will be proven using this technique and the results of this chapter. Moreover, these are the first problems in the SCI hierarchy requiring general towers of arbitrarily large height and the tools provided in this chapter may be used in other areas of computational mathematics. We also state the precise differences and similarities between the SCI hierarchy and the Baire hierarchy. Finally, we give some examples of the broader role of the SCI hierarchy in mathematics.

2.1 The Basic SCI Hierarchy

We begin with the basic set-up of the SCI, as it already appears in the literature [Han11, BACH$^+$19]. First, we define a computational problem. The basic objects of a computational problem are: Ω is some set, called

the primary set, Λ is a set of complex-valued functions on Ω (we allow any such function to map to some finite-dimensional space $\mathbb{C}^N$), called the evaluation set, $\mathcal{M}$ is a metric space, and $\Xi : \Omega \to \mathcal{M}$ is called the problem function. The set Ω is the class of objects that give rise to our computational problems. The problem function $\Xi : \Omega \to \mathcal{M}$ is the map we are interested in computing. Finally, the set Λ is the collection of functions that provide us with the information we are allowed to read. The collection $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ is referred to as a computational problem.

For example, we could consider the case that $\Omega = \mathcal{B}(l^2(\mathbb{N}))$ (the set of bounded linear operators acting on $l^2(\mathbb{N})$) 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 given by (1.4.2). In this case, Λ could correspond to the evaluation of matrix entries (with respect to the canonical basis) of a given $A \in \Omega$.

We can now define in the broadest sense, what we mean by an algorithm.

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] or a Blum–Shub–Smale (BSS) machine [BCSS98]. One can establish that the SCI hierarchy does not collapse (in particular for the spectral problem) regardless of the model of computation. A general algorithm has no restrictions on the operations allowed. Whilst complete generality in this sense may seem to be at odds with practical computation (and the theory of recursion), 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. Neither is this an issue for practical computation since the algorithms in this thesis can be made to work using only arithmetic operations over the rationals (see §2.3). Hence throughout this thesis, 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 non-computability results (proven lower bounds) of this thesis are due to the problem at hand being inherently non-computable. In other words, 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. This captures the notion of successive limits discussed in §1.1. However, before we do so, we will discuss the cases for which we may have a set-valued function. 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 a general algorithm produces a single-valued output i.e. $\Gamma(A) \in \mathcal{M}$ for $A \in \Omega$. 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 2.1.2 (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*

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

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

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

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

Throughout this thesis, 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 $\Gamma_{n_k, \ldots, n_1}$ in the hierarchy, in particular, what kind of operations may be allowed. A tower of algorithms for a computational problem is the toolbox allowed. A radical tower, as defined below, first appeared in [Han11] where it was referred to as a *"set of estimating functions"* for computing spectra. The definition here is substantially more general and allows for the use of these types of towers for a wide range of problems.

Definition 2.1.3 (Arithmetic and radical towers)**.** *Given a computational problem $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$:*

(i) An arithmetic tower of algorithms of height k for $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ is a tower of algorithms where the lowest functions $\Gamma = \Gamma_{n_k, \ldots, n_1} : \Omega \to \mathcal{M}$ 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 $A_f = f(A)$.

(ii) A radical tower of algorithms of height k for $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ is a tower of algorithms where the lowest functions $\Gamma = \Gamma_{n_k, \ldots, n_1} : \Omega \to \mathcal{M}$ satisfy the following: For each $A \in \Omega$ the action of Γ on A consists of only performing finitely many arithmetic operations, comparisons and extracting radicals of $\{A_f\}_{f \in \Lambda_\Gamma(A)}$.

For arithmetic towers we let $\alpha = A$ and for radical towers we let $\alpha = R$.

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

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

Remark 2.1.6. *In other words, a Δ_{m+1} problem is one that be computed in m limits.*

Remark 2.1.7. *In this thesis, we will concern ourselves only with deterministic algorithms. It is possible to extend the SCI hierarchy to probabilistic algorithms, which is useful for settings such as optimisation, and this will be the topic of future work.*

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. The following definition is motivated by the arithmetical hierarchy in logic.

Definition 2.2.1. *Suppose that $\mathcal{M} = \{0, 1\}$ with the discrete topology. We define the following:*

(i) We say that $\Xi : \Omega \to \mathcal{M}$ permits a representation by an alternating quantifier form of length m if

$$\Xi = (Q_m n_m) \cdots (Q_1 n_1) \Gamma_{n_m,\ldots,n_1},$$

where (Q_i) is a list of alternating quantifiers $(\forall)$ and $(\exists)$, and all $\Gamma_{n_m,\ldots,n_1} : \Omega \to \mathcal{M}$ are general algorithms in the sense of Definition 2.1.1.

(ii) We say that $\{\Xi, \Omega\}$ is Σ_m^α if an alternating quantifier form of length m exists with Q_m being $(\exists)$ and $\Gamma_{n_m,\ldots,n_1}$ algorithms of type α, and that $\{\Xi, \Omega\}$ is Π_m^α if an alternating quantifier form of length m exists with Q_m being $(\forall)$ and $\Gamma_{n_m,\ldots,n_1}$ algorithms of type α.

(iii) We say that $\{\Xi, \Omega\}$ *is* Δ_m^α *if* $\{\Xi, \Omega\}$ *is* Σ_m^α *and* Π_m^α.[1]

Definition 2.2.1, the following theorem and Proposition 2.2.4 are taken from [BACH$^+$19]. This section is based on work done in collaboration in [BACH$^+$19].

Theorem 2.2.2. *Following Definition 2.2.1, and supposing that* $\mathcal{M} = \{0, 1\}$*, the following is true.*

1. *If* $\mathrm{SCI}(\Xi, \Omega)_\alpha \leq m$ *then* Ξ *is* Δ_{m+1}^α.
2. *If* Ξ *is* Σ_m^α *or* Π_m^α *then* $\mathrm{SCI}(\Xi, \Omega)_\alpha \leq m$.
3. *For* $m \in \mathbb{N}$*, we have that* $\mathrm{SCI}(\Xi, \Omega)_\alpha = m$ *if and only if* m *is the smallest integer with* Ξ *being* Δ_{m+1}^α.

This motivates the following generalisation when $\mathcal{M}$ is a totally ordered set.

Definition 2.2.3 (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*

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

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

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

If the metric space $\mathcal{M} = \{0, 1\}$, it is clearly a totally ordered set and hence, from Definition 2.2.3, we obtain the SCI hierarchy for arbitrary decision problems. It is not immediately clear whether Definition 2.2.3 and Definition 2.2.1 agree when $\mathcal{M} = \{0, 1\}$. However, the next proposition provides the link.

Proposition 2.2.4 (Properties of the SCI hierarchy I)**.** *Given the above set-up we have the following.*

(i) The SCI hierarchy encompasses the arithmetical hierarchy.

(ii) If $\mathcal{M} = \{0, 1\}$*, then Definition 2.2.3 and Definition 2.2.1 are equivalent and hence the SCI encompasses generalisations of the arithmetical hierarchy. In particular, this holds for arithmetic towers which extends the arithmetical hierarchy to arbitrary domains.*

(iii) If $\mathcal{M} = \{0, 1\}$*, then* $\Delta_k^\alpha = \Sigma_k^\alpha \cap \Pi_k^\alpha$ *for all* k *and* α.

2.2.1 Extending the hierarchy for spectral 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

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

[1] This implies that there exist two alternating quantifier forms with distinct 'heads'.

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

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

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 (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 2.2.5 (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 $m \in \mathbb{N}$*

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

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

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

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:

$$\begin{array}{ccccccccccccccccc}
\Pi_0^\alpha & & & & \Pi_1^\alpha & & & & \Pi_2^\alpha & & & & & & \\
\| & & & \curlyvee & & \subsetneq & & \curlyvee & & \subsetneq & & \curlyvee & & & \\
\Delta_0^\alpha & \subsetneq & \Delta_1^\alpha & \subsetneq & \Sigma_1^\alpha \cup \Pi_1^\alpha & \subsetneq & \Delta_2^\alpha & \subsetneq & \Sigma_2^\alpha \cup \Pi_2^\alpha & \subsetneq & \Delta_3^\alpha & \subsetneq & \cdots & & \\
\| & & & \subsetneq & & \curlyvee & & \subsetneq & & \curlyvee & & & \subsetneq & & \\
\Sigma_0^\alpha & & & & \Sigma_1^\alpha & & & & \Sigma_2^\alpha & & & & & &
\end{array}$$

Note, it is precisely the classes Σ_1^α and Π_1^α that are crucial in computer-assisted proofs (see §2.5.1).

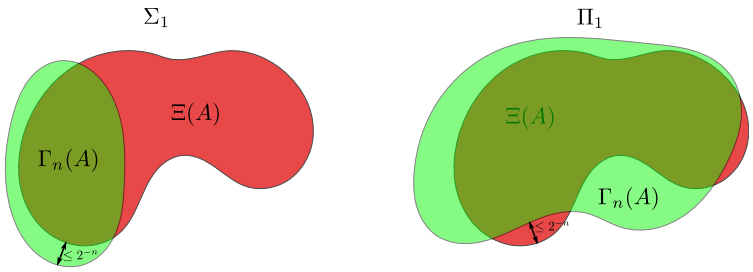


Figure 2.1: Meaning of Σ_1 and Π_1 convergence for problem function Ξ computed in the Hausdorff metric. The red area represents $\Xi(A)$, whereas the green areas represent the output of the algorithm $\Gamma_n(A)$. Σ_1 convergence means convergence as $n \to \infty$ but each output point in $\Gamma_n(A)$ 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).

Remark 2.2.6 (Warning!)**.** *We use the same Δ_k^α, Σ_k^α, Π_k^α notation from, for example, the arithmetical hierarchy. This similarity is deliberate, since classical hierarchies become special cases of the SCI hierarchy (Proposition 2.2.8). However, there is a substantial difference. In classical hierarchies, each Δ_k class is defined via $\Delta_k = \Sigma_k \cap \Pi_k$, which is not always the case in the SCI hierarchy. The Δ_k classes form the core of the SCI hierarchy, and it is only when there is extra structure on the metric space that the Σ_k and the Π_k classes can be defined. Furthermore, there may be cases in the SCI hierarchy where*

$$\Delta_k \neq \Sigma_k \cap \Pi_k.$$

In addition, classical hierarchies also have that $\Sigma_k \setminus \Delta_{k-1} \neq \emptyset$ and $\Pi_k \setminus \Delta_{k-1} \neq \emptyset$, which may not hold in general SCI hierarchies.

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 2.2.7. *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 (see Remark 2.2.9).

Proposition 2.2.8 (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 $\{\Xi, \Omega\} \in \Delta_k^A$.

Remark 2.2.9. *Throughout this thesis, 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).*

2.2.2 Proof of Proposition 2.2.8

In this subsection we prove Proposition 2.2.8, which we have placed in a separate section, allowing the reader to skip it if desired. Let $(\mathcal{M}, d)$ be a metric space with the Attouch–Wets or Hausdorff topology induced by another metric space $(\mathcal{M}', d_{\mathcal{M}'})$. For the Attouch–Wets topology and any fixed $x_0 \in \mathcal{M}'$ we set

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

for $C_1, C_2 \in \mathrm{Cl}(\mathcal{M}')$, where $\mathrm{Cl}(\mathcal{M}')$ denotes the set of non-empty closed subsets of $\mathcal{M}'$. In the case that $\mathcal{M}' = \mathbb{C}$ with the usual metric we take $x_0 = 0$. We have the following 'sandwich' lemma.

Lemma 2.2.10. *Suppose that $(\mathcal{M}, d)$ is the Hausdorff or Attouch–Wets topology induced by a metric space $(\mathcal{M}', d_{\mathcal{M}'})$. Let $\epsilon > 0$. Suppose also that $A, A', B, B', C \in \mathcal{M}$ with $A \subset_{\mathcal{M}'} A'$, $C \subset_{\mathcal{M}'} B'$, $d(C, A') \leq \epsilon$ and $d(B, B') \leq \epsilon$. Then*

$$d(A, C) \leq d(A, B) + 2\epsilon.$$

Proof. Suppose first that $(\mathcal{M}, d)$ is the Hausdorff topology. If $x \in C$ then $x \in B'$ and $\mathrm{dist}(x, A) \leq d(B', A) \leq d(A, B) + \epsilon$. On the other hand, if $x \in A$ then $x \in A'$ and $\mathrm{dist}(x, C) \leq d(A', C) \leq \epsilon$. The result now follows.

Suppose now that $(\mathcal{M}, d)$ is the Attouch–Wets topology and let $x \in \mathcal{M}'$. Since $C \subset_{\mathcal{M}'} B'$ we must have

$$\mathrm{dist}(x, A) - \mathrm{dist}(x, C) \leq \mathrm{dist}(x, A) - \mathrm{dist}(x, B') \leq |\mathrm{dist}(x, A) - \mathrm{dist}(x, B)| + |\mathrm{dist}(x, B) - \mathrm{dist}(x, B')|.$$

Similarly, since $A \subset_{\mathcal{M}'} A'$ we must have

$$\mathrm{dist}(x, C) - \mathrm{dist}(x, A) \leq \mathrm{dist}(x, C) - \mathrm{dist}(x, A') \leq |\mathrm{dist}(x, C) - \mathrm{dist}(x, A')|.$$

It follows that

$$|\mathrm{dist}(x, A) - \mathrm{dist}(x, C)| \leq |\mathrm{dist}(x, A) - \mathrm{dist}(x, B)| + |\mathrm{dist}(x, B) - \mathrm{dist}(x, B')| + |\mathrm{dist}(x, C) - \mathrm{dist}(x, A')|$$

and this finishes the proof of the Lemma. □

Proposition 2.2.11. *Let $(\mathcal{M}, d)$ be either a metric space with the Attouch–Wets or Hausdorff topology induced by another metric space $(\mathcal{M}', d_{\mathcal{M}'})$ or a totally ordered metric space with order respecting metric. Suppose we have a computational problem*

$$\Xi : \Omega \to \mathcal{M},$$

with a corresponding convergent Σ_k^α tower $\Gamma^1_{n_k,\ldots,n_1}$ and a corresponding convergent Π_k^α tower $\Gamma^2_{n_k,\ldots,n_1}$ (either both arithmetic or both general). Suppose also that $1 \leq k \leq 3$ and that, in the case of arithmetic towers, 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 $\{\Xi, \Omega\} \in \Delta_k^\alpha$.

Remark 2.2.12. *This proposition essentially says that we can combine the two notions of error control Π_k and Σ_k to reduce the number of limits needed by one.*

Proof. **Step 1:** For $k = 1$ and the case that $(\mathcal{M}, d)$ is either a metric space with the Attouch–Wets or Hausdorff topology, this is a trivial consequence of Lemma 2.2.10. Let δ_{n_1} be an approximation of

$$d(\Gamma^1_{n_1}(A), \Gamma^2_{n_1}(A)) + 2 \cdot 2^{-n_1}$$

from above to accuracy $1/n_1$. Note that suitable approximations can easily be generated using approximations of $d(\Gamma^1_{n_1}(A), \Gamma^2_{n_1}(A))$. Let $\epsilon > 0$, then simply choose $n_1 \in \mathbb{N}$ minimal such that $\delta_{n_1} \leq \epsilon$. In the case that $(\mathcal{M}, d)$ is totally ordered with order respecting metric

$$d(\Gamma^1_{n_1}(A), \Xi(A)) \leq d(\Gamma^1_{n_1}(A), \Gamma^2_{n_1}(A)),$$

and we can take n_1 large such that the right hand side is less than the given ϵ (recall we can compute the right hand side to arbitrary precision). Set $\Gamma(A) = \Gamma^1(A)$, then we have

$$d(\Gamma_(A), \Xi(A)) \leq \epsilon.$$

Step 2: For larger k, we use the same idea, but we must be careful to ensure the first $k-1$ limits exist. For the rest of the proof, $\tilde{d}$ will denote an approximation of d to accuracy $1/n_1$ (which by assumption can always be computed).

We first deal with the case $k = 2$. Let $\epsilon > 0$ and consider the intervals $J^1_\epsilon = [0, \epsilon]$ and $J^2_\epsilon = [2\epsilon, \infty)$. Let $\delta_{n_2,n_1}(A)$ be an approximation of

$$d(\Gamma^1_{n_2,n_1}(A), \Gamma^2_{n_2,n_1}(A)) + 2 \cdot 2^{-n_2}$$

from above to accuracy $1/n_1$. Again note that we can easily construct such approximations. It is clear that $\lim_{n_1\to\infty} \delta_{n_2,n_1}(A) = d(\Gamma^1_{n_2}(A), \Gamma^2_{n_2}(A)) + 2 \cdot 2^{-n_2} =: \delta_{n_2}(A)$ and that $d(\Gamma^1_{n_2}(A), \Xi(A)) \leq \delta_{n_2}(A)$ (again appealing to Lemma 2.2.10 if we are in the case of the Attouch–Wets or Hausdorff topologies). Given n_1, n_2, let $l(n_2, n_1) \leq n_1$ be maximal such that $\delta_{n_2,l}(A) \in J^1_\epsilon \cup J^2_\epsilon$. If no such l exists or $\delta_{n_2,l}(A) \in J^1_\epsilon$ then define $\mathrm{Osc}(\epsilon; n_1, n_2, A) = 1$ otherwise define $\mathrm{Osc}(\epsilon; n_1, n_2, A) = 0$. Since $\delta_{n_2,n_1}(A)$ cannot oscillate infinitely often between the two intervals J^1_ϵ and J^2_ϵ, it follows that

$$\mathrm{Osc}(\epsilon; n_2, A) := \lim_{n_1\to\infty} \mathrm{Osc}(\epsilon; n_1, n_2, A)$$

exists. Define $\Gamma^\epsilon_{n_1}(A)$ as follows. Choose $j \leq n_1$ minimal such that $\mathrm{Osc}(\epsilon; n_1, j, A) = 1$ if such a j exists, and define $\Gamma^\epsilon_{n_1}(A) = \Gamma_{j,n_1}(A)$. If no such j exists then define $\Gamma^\epsilon_{n_1}(A) = C_0$ where C_0 is any fixed member of $(\mathcal{M}, d)$. In particular, $\Gamma^\epsilon_{n_1}$ is a type α algorithm. Now for large n_2, we must have $\delta_{n_2}(A) < \epsilon$ and hence $\mathrm{Osc}(\epsilon; n_2, A) = 1$. It follows that $\Gamma^\epsilon(A) = \lim_{n_1\to\infty} \Gamma^\epsilon_{n_1}(A)$ exists and is equal to $\Gamma^1_N(A)$ where $N \in \mathbb{N}$ is minimal with $\mathrm{Osc}(\epsilon; N, A) = 1$. It follows that $d(\Gamma^\epsilon(A), \Xi(A)) \leq 2\epsilon$.

We will use the $\Gamma^\epsilon_{n_1}(A)$ to construct a height one tower. Observe first of all that by our assumptions we can compute $\tilde{d}(\Gamma^{\epsilon_1}_m(A), \Gamma^{\epsilon_2}_n(A))$ for $m, n \in \mathbb{N}$ and $\epsilon_1, \epsilon_2 > 0$. Given n_1, choose $j = j(n_1) \leq n_1$ maximal such that for all $1 \leq l \leq j$ we have

$$\tilde{d}(\Gamma^{2^{-j}}_{n_1}(A), \Gamma^{2^{-l}}_{n_1}(A)) \leq 4(2^{-j} + 2^{-l}). \tag{2.2.1}$$

If no such j exists then set $\Gamma_{n_1} = C_0$, otherwise set $\Gamma_{n_1}(A) = \Gamma_{n_1}^{2^{-j(n_1)}}(A)$. Again, this is easily seen to be a type α algorithm. Pick any $N \in \mathbb{N}$, then by the convergence of the $\Gamma^\epsilon_{n_1}(A)$ and $d(\Gamma^\epsilon(A), \Xi(A)) \leq 2\epsilon$, (2.2.1) must hold for $j = N$ and $1 \leq l \leq N$ if n_1 is large enough. Hence by definition of $j(n_1)$,

$$\limsup_{n_1\to\infty} d(\Gamma_{n_1}(A), \Xi(A)) \leq \limsup_{n_1\to\infty} d(\Gamma_{n_1}^{2^{-N}}(A), \Xi(A)) + 2^{3-N} \leq 2^{4-N}.$$

Since N was arbitrary, we must have convergence to $\Xi(A)$.

Step 3: We now deal with $k = 3$. The strategy will be similar to the $k = 2$ case but now we construct $\Gamma^\epsilon_{n_2,n_1}(A)$ such that $\Gamma^\epsilon_{n_2}(A) := \lim_{n_1\to\infty} \Gamma^\epsilon_{n_2,n_1}(A)$ exists and is 3ϵ close to $\Xi(A)$ for large n_2, but may not converge in $(\mathcal{M}, d)$. Using this, we will construct a height two type α tower.

As in step 2, let $\epsilon > 0$ and consider the intervals $J^1_\epsilon = [0, \epsilon]$ and $J^2_\epsilon = [2\epsilon, \infty)$. Let $\delta_{n_3,n_2,n_1}(A)$ be an approximation of

$$d(\Gamma^1_{n_3,n_2,n_1}(A), \Gamma^2_{n_3,n_2,n_1}(A)) + 2 \cdot 2^{-n_3},$$

from above to accuracy $1/n_1$. Again, we have

$$\lim_{n_2\to\infty} \lim_{n_1\to\infty} \delta_{n_3,n_2,n_1}(A) = d(\Gamma^1_{n_3}(A), \Gamma^2_{n_3}(A)) + 2 \cdot 2^{-n_3} =: \delta_{n_3}(A)$$

exists with $d(\Gamma^1_{n_3}(A), \Xi(A)) \leq \delta_{n_3}(A)$. Given n_1, n_2 and j, let $l(j, n_2, n_1) \leq n_1$ be maximal such that $\delta_{j,n_2,l}(A) \in J^1_\epsilon \cup J^2_\epsilon$. If no such l exists or $\delta_{j,n_2,l}(A) \in J^1_\epsilon$ then define $\mathrm{Osc}(\epsilon; n_1, n_2, j, A) = 1$ otherwise define $\mathrm{Osc}(\epsilon; n_1, n_2, j, A) = 0$. Arguing as in step 1 we have that

$$\mathrm{Osc}(\epsilon; n_2, j, A) := \lim_{n_1\to\infty} \mathrm{Osc}(\epsilon; n_1, n_2, j, A)$$

exists. Now consider $\mathrm{Osc}(\epsilon; n_1, n_2, j, A)$ for $j \leq n_2$. If such a j exists with $\mathrm{Osc}(\epsilon; n_1, n_2, j, A) = 1$ then let $j(n_1, n_2)$ be the minimal such j and set $\Gamma^\epsilon_{n_2,n_1}(A) = \Gamma^1_{j(n_1,n_2),n_2,n_1}(A)$. Otherwise set $\Gamma^\epsilon_{n_2,n_1}(A) = C_0$, where again C_0 is some fixed member of $(\mathcal{M}, d)$. Since we only deal with finitely many $j \leq n_2$, it is clear that $\Gamma^\epsilon_{n_2,n_1}$ is a type α algorithm. Furthermore, we must have that $\Gamma^\epsilon_{n_2}(A) := \lim_{n_1\to\infty} \Gamma^\epsilon_{n_2,n_1}(A)$ exists and is defined as follows. Let $j(n_2) \leq n_2$ be minimal with $\mathrm{Osc}(\epsilon; n_2, j, A) = 1$ (if such a j exists). If such a j exists then $\Gamma^\epsilon_{n_2}(A) = \Gamma^1_{j(n_2),n_2}(A)$, otherwise $\Gamma^\epsilon_{n_2}(A) = C_0$.

Now there exists $N \in \mathbb{N}$ such that $\delta_N(A) < \epsilon/2$ and hence $\delta_{N,n_2}(A) < \epsilon$ for large n_2. But this implies that $\mathrm{Osc}(\epsilon; n_2, N, A) = 1$. Hence for n_2 large we must have $j(n_2) \leq N$. If $\delta_l(A) > 2\epsilon$ then for large n_2 we must have $\delta_{l,n_2}(A) > 2\epsilon$ and hence $\mathrm{Osc}(\epsilon; n_2, l, A) = 0$. As n_2 increases, $j(n_2)$ may not converge. However, the above arguments show that for large n_2 it can take only finitely many values, say in the set $S = \{s_1, ..., s_m\}$, all of which must have $\delta_{s_i}(A) \leq 2\epsilon$. It follows that for large n_2 we must have

$$d(\Gamma^\epsilon_{n_2}(A), \Xi(A)) \leq 3\epsilon. \tag{2.2.2}$$

Now we get to work using these 'towers' (which do not necessarily converge in the last limit) and the trick to avoid oscillations. Define

$$F(n_1, n_2, j, l, A) := \tilde{d}(\Gamma^{2^{-j}}_{n_2,n_1}(A), \Gamma^{2^{-l}}_{n_2,n_1}(A)),$$
$$F(n_2, j, l, A) := \lim_{n_1\to\infty} F(n_1, n_2, j, l, A) = d(\Gamma^{2^{-j}}_{n_2}(A), \Gamma^{2^{-l}}_{n_2}(A))$$

and the intervals $J^1_{j,l} = [0, 4(2^{-j} + 2^{-l})], J^2_{j,l} = [8(2^{-j} + 2^{-l}), \infty)$. Given j, l, n_1 and n_2, we define $i(j, l, n_2, n_1) \leq n_1$ be maximal such that $F(i, n_2, j, l, A) \in J^1_{j,l} \cup J^2_{j,l}$. If no such i exists or if it does and $F(i, n_2, j, l, A) \in J^1_{j,l}$ then define $\widehat{\mathrm{Osc}}(n_1, n_2, j, l, A) = 1$ otherwise define $\widehat{\mathrm{Osc}}(n_1, n_2, j, l, A) = 0$.

Choose $j = j(n_1, n_2) \leq n_2$ maximal such that for all $1 \leq l \leq j$ we have $\widehat{\mathrm{Osc}}(n_1, n_2, j, l, A) = 1$. If no such j exists then set $\Gamma_{n_2,n_1} = C_0$, otherwise set $\Gamma_{n_2,n_1}(A) = \Gamma_{n_2,n_1}^{2^{-j(n_1,n_2)}}(A)$. Again, this is easily seen to be a type α algorithm.

Arguing as before, we have the existence of

$$\widehat{\mathrm{Osc}}(n_2, j, l, A) := \lim_{n_1 \to \infty} \widehat{\mathrm{Osc}}(n_1, n_2, j, l, A).$$

Now define $h = h(n_2) \leq n_2$ maximal such that for all $1 \leq l \leq h$ we have $\widehat{\mathrm{Osc}}(n_2, h, l, A) = 1$. If no such h exists then we must have

$$\Gamma_{n_2}(A) := \lim_{n_1 \to \infty} \Gamma_{n_2,n_1}(A) = C_0,$$

otherwise we must have

$$\Gamma_{n_2}(A) := \lim_{n_1 \to \infty} \Gamma_{n_2,n_1}(A) = \Gamma_{n_2}^{2^{-h(n_2)}}(A).$$

By (2.2.2), for any fixed j, l we have $\widehat{\mathrm{Osc}}(n_2, j, l, A) = 1$ for large n_2 and hence $h(n_2)$ exists for large n_2 and diverges to ∞. Now let $N \in \mathbb{N}$ then it follows that

$$\begin{aligned}\limsup_{n_2 \to \infty} d(\Gamma_{n_2}^{2^{-h(n_2)}}(A), \Xi(A)) &\leq \limsup_{n_2 \to \infty} d(\Gamma_{n_2}^{2^{-N}}(A), \Xi(A)) + d(\Gamma_{n_2}^{2^{-h(n_2)}}(A), \Gamma_{n_2}^{2^{-N}}(A)) \\ &\leq 3 \cdot 2^{-N} + \limsup_{n_2 \to \infty} 8(2^{-h(n_2)} + 2^{-N}) \leq 11 \cdot 2^{-N}.\end{aligned}$$

Since N was arbitrary we must have convergence to $\Xi(A)$. □

Proof of Proposition 2.2.8. The statement regarding intersections follows directly from Proposition 2.2.11 and the following remark - no assumptions regarding the ability to compute distances between outputs of algorithms is necessary when considering general towers. For the sharpness result in (i), we deal with $X = \Sigma$ and the $X = \Pi$ follows from an identical argument. Suppose that $\Delta_k^G \not\ni \{\Xi, \Omega\} \in \Sigma_k^\alpha$. If $\{\Xi, \Omega\} \in \Pi_k^\alpha$, we would have $\{\Xi, \Omega\} \in \Sigma_k^\alpha \cap \Pi_k^\alpha \subset \Sigma_k^G \cap \Pi_k^G = \Delta_k^G$, a contradiction. □

2.3 Turing Towers and Algorithmic Unification

So far, we have considered algorithms with exact input, which can also store and perform arithmetic on real numbers. Whilst such assumptions may be useful from a numerical analysis point of view, they are not how the mechanics of computation operate in the real world. In this section, we aim to cross the bridge between this point of view and classical computation theory. We conclude that for the problems in this thesis, the difference in points of view are irrelevant - both give the same classification of computational difficulty in the SCI hierarchy. This section also shows that the Σ_1 and Π_1 classifications proven in this thesis can be used for computer-assisted proofs.

Suppose we are given a computational problem $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$, and that the evaluation set $\Lambda = \{f_j : \Omega \to \mathbb{C}^{k_j}\}_{j \in \mathcal{I}}$, where $\mathcal{I}$ is some countable index set that can be finite or infinite. However, obtaining f_j may be a computational task on its own. For instance, $f_j(A)$ could be the number $e^{\frac{\pi}{j} i}$ for example or a matrix value from an inner product integral. Hence, we cannot access $f_j(A)$, but rather $f_{j,n}(A)$ where $f_{j,n}(A) \to f_j(A)$ as $n \to \infty$. Or, just as for problems that are high up in the SCI hierarchy, it could be that we need several limits, in particular one may need mappings $f_{j,n_m,\ldots,n_1} : \Omega \to [\mathbb{Q} + i\mathbb{Q}]^{k_j}$ such that

$$\lim_{n_m \to \infty} \ldots \lim_{n_1 \to \infty} f_{j,n_m,\ldots,n_1}(A) = f_j(A) \quad \forall A \in \Omega. \tag{2.3.1}$$

In particular, we may view the problem of obtaining $f_j(A)$ as a problem in the SCI hierarchy, where Δ_1 classification would correspond to the existence of mappings $f_{j,n} : \Omega \to [\mathbb{Q} + i\mathbb{Q}]^{k_j}$ such that

$$\|f_{j,n}(A) - f_j(A)\| \leq 2^{-n} \quad \forall A \in \Omega. \tag{2.3.2}$$

This idea is formalised in the following definition.

Definition 2.3.1 (Δ_m-information)**.** *Let $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ be a computational problem. For $m \in \mathbb{N}$ we say that Λ has Δ_{m+1}-information if each $f_j \in \Lambda$ is not available, however, there are mappings $f_{j,n_m,\ldots,n_1} : \Omega \to [\mathbb{Q} + i\mathbb{Q}]^{k_j}$ such that* (2.3.1) *holds. Similarly, for $m = 1$ there are mappings $f_{j,n} : \Omega \to [\mathbb{Q} + i\mathbb{Q}]^{k_j}$ such that* (2.3.2) *holds. Finally, if $k \in \mathbb{N}$ and $\hat{\Lambda}$ is a collection of such functions described above such that Λ has Δ_k-information, we say that $\hat{\Lambda}$ provides Δ_k-information for Λ. Moreover, we denote the family of all such $\hat{\Lambda}$ by $\mathcal{L}^k(\Lambda)$.*

With this definition, we can define a computational problem with Δ_m-information.

Definition 2.3.2 (Computational problem with Δ_m-information)**.** *Given $m \in \mathbb{N}$, a computational problem where Λ has Δ_m-information is denoted by $\{\Xi, \Omega, \mathcal{M}, \Lambda\}^{\Delta_m}$ and denotes the family of computational problems $\{\Xi, \Omega, \mathcal{M}, \hat{\Lambda}\}$ where $\hat{\Lambda} \in \mathcal{L}^m(\Lambda)$.*

Definition 2.3.3 (Tower with Δ_m-information)**.** *A tower of algorithms of height k with Δ_m-information is a tower of algorithms of height k for the computational problem $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$, where Λ has Δ_m-information such that the tower converges (all m-limits) for any evaluation set $\hat{\Lambda} \in \mathcal{L}^m(\Lambda)$.*

The above three definitions are due to discussions between the author and Alex Bastounis. The SCI hierarchy, given Δ_m-information, is then defined in the obvious way, where the convergence has to happen given any $\hat{\Lambda} \in \mathcal{L}^m(\Lambda)$. We will use the notation

$$\{\Xi, \Omega, \mathcal{M}, \Lambda\}^{\Delta_m} \in \Delta_k^\alpha$$

to denote that the computational problem is in Δ_k^α with respect to towers of algorithms with Δ_m-information. Since $\{\Xi, \Omega, \mathcal{M}, \Lambda\}^{\Delta_m}$ is the collection of all computational problems with Λ replaced by $\hat{\Lambda} \in \mathcal{L}^m(\Lambda)$, we note that the use of $\in$ is a slight abuse of notation. When $\mathcal{M}$ and Λ are obvious then we will write $\{\Xi, \Omega\}^{\Delta_m} \in \Delta_k^\alpha$ for short. In exactly the same way as above, we can define Π_k^α and Σ_k^α for $\{\Xi, \Omega, \mathcal{M}, \Lambda\}^{\Delta_m}$ if the metric space that Ξ maps to is totally ordered or a Attouch–Wets/Hausdorff metric space.

To make a connection with the classical theory of computation, consider the case where $\Lambda = \{f_j\}_{j \in \mathcal{I}}$ has some natural (countably infinite) ordering $\mathcal{I}$. For example, in the case of spectral computations for general $A \in \mathcal{B}(l^2(\mathbb{N}))$ we have the matrix evaluations $f_j(A) = \langle Ae_{\phi_2(j)}, e_{\phi_1(j)} \rangle$, where $\phi = (\phi_1, \phi_2)$ is an effective bijection from $\mathbb{N}$ to $\mathbb{N}^2$. Of course given $\hat{\Lambda} \in \mathcal{L}^1(\Lambda)$ we must replace $\mathcal{I}$ by $\mathcal{I} \times \mathbb{N}$. By a suitable effective enumeration of $\mathbb{Q} + i\mathbb{Q}$, we can assume each $f_{j,n}$ maps into $\mathbb{N}$. We can also view the evaluation functions as an oracle through the mapping defined by

$$\hat{\Lambda}(A) : \mathcal{I} \times \mathbb{N} \ni (j, n) \mapsto f_{j,n}(A) \in \mathbb{N}.$$

Now suppose that our metric space $(\mathcal{M}, d)$ is the Hausdorff metric on non-empty compact subsets of $\mathbb{C}$, the Attouch–Wets metric on non-empty closed subsets of $\mathbb{C}$, $\mathbb{R}$ with its usual topology or some (at most countable) discrete ordered space.

Definition 2.3.4 (Turing Tower)**.** *Given a computational problem $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ where $(\mathcal{M}, d)$ is one of the above metric spaces, a Turing Tower of Algorithms of height k for $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ is a tower of algorithms of height k with Δ_1-information where the lowest level algorithms*

$$\Gamma = \Gamma_{n_k,\ldots,n_1} : \Omega \to \mathcal{M}$$

satisfy the following. For each $A \in \Omega$ and $\hat{\Lambda} \in \mathcal{L}^1(\Lambda)$:

1. *We can view the output as lying in the space $\{0,1\}^*$ by a suitable effective enumeration. For example, if $(\mathcal{M}, d) = \mathbb{C}$ (or $\mathbb{R}$) with the usual metric, the output $\Gamma(A) \in \mathbb{Q} + i\mathbb{Q}$ (or $\mathbb{Q}$). If $(\mathcal{M}, d)$ is the Hausdorff metric on the non-empty compact subsets of $\mathbb{C}$ or the Attouch–Wets metric on non-empty closed subsets of $\mathbb{C}$, $\Gamma(A)$ is a finite collection of points in $\mathbb{Q} + i\mathbb{Q}$.*

2. *Γ is an oracle Turing machine such that given the input $(n_1, \ldots, n_k)$ and oracle $\hat{\Lambda}(A)$, it computes $\Gamma_{n_k,\ldots,n_1}(A)$.*

Such a tower will be denoted by the superscript T.

Remark 2.3.5. *Although the above may seem complicated, it can be summarised as follows. A Turing tower is a tower of algorithms for which the lowest levels can be implemented using a Turing machine. The Δ_1-information is needed to model the fact that we can never store an arbitrary real number to full precision on a finite computer.*

Remark 2.3.6. *Note that we still require the convergence of our towers in the original metric space $\{\mathcal{M}, d\}$, which of course may not be compatible with the metric induced by the coding of our range space.*

A remarkable consequence of our results is that for all of the problems considered in this thesis, the SCI classification does not change if we consider Turing towers instead of general towers or arithmetic towers. In other words, it does not matter which model of computation one uses for a definition of 'algorithm'; from a classification point of view, they are equivalent for these spectral problems. This is a straightforward application of Church's thesis, along with a careful analysis of the stability of our algorithms, which are in general based upon computing (generalisations of) the resolvent or its norm. Explicitly, for the algorithms based on `DistSpec` (see §3.5.1) it is possible to carry out an error analysis with Δ_1-information. If we know the errors and can also bound numerical errors (or use exact arithmetic on $\mathbb{Q}$), then we can incorporate this uncertainty for the estimation of $\|R(z, A)\|^{-1}$ and still gain the same classification of our problems. This also holds for other algorithms based on similar functions. This leads to rigorous Σ_k^α or Π_k^α type error control suitable for verifiable numerics. In particular, for Σ_1^α or Π_1^α towers of algorithms, this could be useful for computer-assisted proofs.

2.4 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 the only known method of providing 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 $\mathrm{SCI} >$

2, where we typically reduce the problems discussed in this section to the given spectral problem. The technique can often be quite fiddly and depends on the problem at hand.

It is beyond the scope of this thesis 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.[2] 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.

2.4.1 Recalling some results from descriptive set theory

We briefly recall 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

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

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 2.4.1 ([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 2.4.2 (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.*

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

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

Definition 2.4.3, Proposition 2.4.4 and the idea of using well-orderings in part of the proof of Theorem 2.4.5 below are due to discussions between the author and Arno Pauly. 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 2.4.3. *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 2.4.4. *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 2.4.5. *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 2.4.6. *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 2.4.4 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 $x \in \cap_{l\le n_i}\Gamma^{-1}_{i,l}(U_{i,l})$ and $\Lambda_{\Gamma_{i,l}}(x) \subset \{\lambda_j : j \le k\}$ for $l \le n_i$. 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 $\{\lambda_j : j \le I(a)\}$ 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 $j \le I(a)$ 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 2.4.4 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 2.4.2, 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

$$\Xi(a) = \lim_{n_m\to\infty} \dots \lim_{n_1\to\infty} f_{n_m,\dots,n_1}(a). \tag{2.4.1}$$

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

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

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, (2.4.1) holds with $f_{n_m,\dots,n_1}(a) = \Gamma_{n_m,\dots,n_1}$. □

2.4.3 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 $\{a_{m_1,...,m_k}\}_{m_1,...,m_k \in \mathbb{N}}$ with entries $a_{m_1,...,m_k} \in \{0,1\}$. As usual Λ_k is the set of component-wise evaluations/projections. Consider the formulas

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

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

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 $a \in \Omega_k$

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

that is, so that all quantifier types alternate.

Theorem 2.4.7. *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 $R = P, Q$,*

$$\Delta_{k+1}^G \not\ni \{\Xi_{k,R}, \Omega_k, \mathcal{M}, \Lambda_k\} \in \Delta_{k+2}^A.$$

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.

Remark 2.4.8. *Note that we allow both discrete and continuous spaces $\mathcal{M}$, which will be important for our reduction arguments when proving lower bounds for classifications of spectral problems for non-discrete $\mathcal{M}$. The lower bound is a strong result in the sense that it holds regardless of the model of computation. In other words, it is the intrinsic combinatorial complexity of the problems that makes the problems hard.*

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 $\{\Xi_{k,P}, \Omega_k, \mathcal{M}, \Lambda_k\} \in \Delta_{k+2}^A$. 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

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

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 $\{\Xi_{k,P}, \Omega_k, \mathcal{M}, \Lambda_k\} \in \Sigma^A_{k+1}$ if k is even and $\{\Xi_{k,P}, \Omega_k, \mathcal{M}, \Lambda_k\} \in \Pi^A_{k+1}$ 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 2.4.5. For the lower bound we note that P is Σ^0_3 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 $\Sigma^0_3(\Omega_k)$ but not $\Pi^0_3(\Omega_k)$. It also follows that $\Xi_{k,P}$ is $\Sigma^0_{k+1}(\Omega_k)$ complete if k is even and $\Pi^0_{k+1}(\Omega_k)$ complete otherwise. Now suppose for a contradiction that $\{\Xi_{k,P}, \Omega_k, \mathcal{M}, \Lambda_k\} \in \Delta^G_{k+1}$. But then Theorem 2.4.5 implies that $\Xi_{k,P} \in \mathcal{B}_k(\Omega_k, \mathcal{M})$ and hence by Theorem 2.4.2, $\Xi_{k,P}$ is $\Sigma^0_{k+1}(\Omega_k)$ measurable. $\Xi_{k,P}$ is the indicator function of set, also denoted by $\Xi_{k,P}$, which is either $\Sigma^0_{k+1}(\Omega_k)$ or $\Pi^0_{k+1}(\Omega_k)$ complete depending on the parity of k. But 0 and 1 are separated in $\mathcal{M}$ and hence since $\Xi_{k,P}$ is $\Sigma^0_{k+1}(\Omega_k)$ measurable, $\Xi_{k,P}$ and its complement both lie in $\Sigma^0_{k+1}(\Omega_k)$. It follows that $\Xi_{k,P} \in \Sigma^0_{k+1}(\Omega_k) \cap \Pi^0_{k+1}(\Omega_k)$, contradicting the stated completeness. □

For our applications to spectral problems, 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 see clearly from the proof of Theorem 2.4.7 that it holds for a much wider class of decision problems, but these four are the only ones that we shall use in the sequel.

Remark 2.4.9. *The results of this section point towards the extension of the SCI hierarchy to countable ordinals and beg the question of whether this could be useful. This will be explored in future work.*

2.4.4 Key similarities and differences between the SCI and Baire hierarchies

We end this section by discussing the key similarities and differences between the SCI and Baire hierarchies.

Similarities between the SCI and Baire hierarchies. The main similarity between the hierarchies is the concept of pointwise limits. In some special cases, we have equivalence (see Theorem 2.4.5), but, in general, this is not the case.

Differences between the SCI and Baire hierarchies. The hierarchies describe very different problems and have different motivations.

(i) (*Generality*). The SCI hierarchy is designed to be able to handle all types of computational problems such as Smale's problem on iterative polynomial root-finding, spectral problems, and solving PDEs. This is not within the scope, nor is it the intention of the Baire hierarchy.

(ii) (*Refinements*). When extra structure on $\mathcal{M}$ is available, the SCI hierarchy can be refined as in §2.2. In particular, we obtain the Σ^α_k and Π^α_k classes. This type of refinement is not captured by the Baire hierarchy.

(iii) (*Topology vs information*). The most striking difference is that the Baire hierarchy is based on (metrisable) topologies, whereas the SCI hierarchy is based on the information Λ available to the algorithm. This makes the SCI hierarchy a more natural fit for scientific computation - often the type of information presented to us is fixed and cannot be changed. To illustrate this, consider the computational

spectral problem. Let $\Xi : \Omega \ni A \mapsto \mathrm{Sp}(A) \in \mathcal{M}$ where Ω is the set of self-adjoint operators in $\mathcal{B}(l^2(\mathbb{N}))$ and $\mathcal{M}$ is the collection of non-empty compact subsets of $\mathbb{C}$ with the Hausdorff metric. The spectrum then depends continuously on the operator norm and hence, if we equip Ω with the operator norm topology, Ξ is Baire class 0. However, the SCI for this computational problem is two if Λ consists of matrix entry evaluations. Changing the metric on Ω, causes the Baire class to change, but does not alter the SCI. Instead, the SCI changes with Λ (becoming one if we have the bounded dispersion information in Chapter 3).

2.5 The Role of the SCI Hierarchy in Mathematics

The SCI hierarchy encompasses many key computational problems in the history of mathematics with many applications in the mathematical sciences. To end this chapter, we discuss a non-exhaustive list below.

2.5.1 The SCI hierarchy and computer-assisted proofs

Computer-assisted proofs are quickly becoming a central part of mathematics (see, for example, the quotation of Gowers in §1.1). 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^T 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 examples of this kind:

- ***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

$$\langle x, c\rangle_K \leq M \text{ subject to } Ax = y, \quad x \geq 0, \tag{2.5.1}$$
$$\langle x, c\rangle_K = \lfloor 10^K \langle x, c\rangle \rfloor 10^{-K}, \quad K \in \mathbb{N}, \quad M \in \mathbb{Q}.$$

  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 (2.5.1) and its classification hold:

  (i) For any integer $\tilde{K} > 1$ there exists a class of inputs Ω such that the problem (2.5.1) with $K = \tilde{K}$ is $\notin \Sigma_1^G$. However, with the same input class Ω, we have that the problem (2.5.1), 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 (2.5.1) must be decided for $K = 6$. Given the class Ω in (i), if the inequality $\langle x, c\rangle_K \leq M$ in (2.5.1) 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 $\in \Sigma_1^A$.

Remark 2.5.1 (Proving Σ_1^A or Π_1^A results)**.** *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.*

2.5.2 Smale's problem on iterative generally convergent algorithms and the SCI

In the 1980s, Smale initiated a comprehensive programme concerning the foundations of computational mathematics [Sma81, BCSS98], focusing on problems in scientific computing rather than classical computer science (the goal being to establish a rigorous complexity theory for real-number calculations). One

of the key problems considered was polynomial root-finding. Newton's method may not converge for this problem, even for a cubic polynomial. A natural question was formulated in terms of the existence of iterative generally convergent algorithms [Sma85], *"Is there any purely iterative generally convergent algorithm for polynomial zero finding?"* McMullen [McM87, McM88, Sma98] answered this problem as follows: yes, if the degree is three; no, if the degree is higher. Doyle and McMullen later demonstrated a striking phenomenon [DM89]: this problem can be solved in the case of the quartic and the quintic using several limits. They introduced a 'tower of algorithms' in order to make this precise and showed that one could not handle the problem for degree six or larger, regardless of the height of the tower (number of limits used). In particular, Smale's problem on the existence of iterative generally convergent algorithms and the theory of McMullen and Doyle become classification problems in the SCI (with a certain restriction on the type of algorithm allowed).

2.5.3 Further examples

(i) *Insolvability of the quintic:* The insolvability of the quintic becomes a classification problem in the SCI hierarchy. The classic Abel–Ruffini theorem (insolvability of the quintic) shows that the SCI of the problem of computing the zeros of a polynomial, when one can only use arithmetic operations and radicals, is greater than zero for polynomials of degree five. Note that this (along with a construction of a convergent algorithm) shows the general finite-dimensional computational spectral problem lies in Δ_1^A and not in Δ_0^R.

(ii) *Optimisation:* As discussed in §2.5.1, deciding feasibility of linear programs given irrational inputs is not only undecidable ($\notin \Delta_1^G$) but $\notin \Sigma_1^G$. This also holds for many other key problems in optimisation such as finding minimisers of Basis pursuit and Lasso. These form the basis of many areas of information theory, such as compressed sensing, statistical estimation, areas of machine learning etc.

(iii) *Spectral problems:* As discussed in §1.1, in the nineties Arveson noted, regarding the lack of algorithms that could handle general spectral problems, that [Arv94b], *"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."* Due to the example of the diagonal matrices in (1.1.4), most infinite-dimensional computational spectral problems of interest are not in Δ_1^G. Many are also not Δ_2^G. Hence, none of the existing methods at the time could handle them. This explains the problem in Arveson's quotation - the standard methods were based on one limit approaches, and would therefore never capture the depth of the computational spectral problem. However, an important exception is given by the algorithms and computational problems in Chapter 3. It is also true that devising towers of algorithms can often inform us which information is needed to reduce the SCI of a problem.

Most of the classical literature on spectral computation is devoted to establishing algorithms that, in view of the SCI hierarchy, would provide Δ_2^A classification for specific subclasses of operators. Note that according to Turing's definition of computability, problems that are not in Δ_1^T are non-computable. Hence, the field of computational spectral theory has, even from the beginning, been concerned with non-computable problems.

Part I

Spectra, Spectral Measures and Spectral Decompositions

Chapter 3

Computing Spectra with Error Control

We begin the study of infinite-dimensional spectral computations with the problem of computing the spectrum. This chapter is based on the article [CRH19] and the generalisations to unbounded operators in [CH19a]. These algorithms 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.
- They can be efficiently implemented.
- They are local (one can compute the spectrum in any desired region of the complex plane) 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.[1] Previous methods aimed at tackling the *general* problem either suffer from spectral pollution (discussed further in §7.1 and §7.3.2) or do not converge to the full spectrum. Even in the cases where it converges, the finite section method 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.

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. In the analytic case, we do this via the power series representation 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. This idea will reappear in Chapters 6–8, since it allows us to grasp geometric properties of the spectrum. Similar ideas used to compute this approximation can be used to compute 'approximate states'.

[1] There are examples where such methods exist in certain cases - see §1.3.

Computing spectra of operators is a fundamental problem in the sciences, and it is hard to overestimate its importance, with wide-ranging applications (outlined in §1.1). This is highlighted by the current interest in the spectral properties of systems with complicated spectra. The study of aperiodic systems, such as quasicrystals [SBGC84, Sta12], often leads to complicated, even fractal-like spectra [HSYY$^+$13, DWM$^+$13], which can make current methods of computation difficult. Another example is given by recent experimental breakthroughs in open systems in optics, which typically yield non-Hermitian Hamiltonians, as there is no guaranteed energy preservation [RBM$^+$12, GSD$^+$09, RMEG$^+$10]. We shall demonstrate how the algorithms of this chapter can be implemented in a computationally efficient manner, allowing us to tackle problems that before, regardless of computing power, seemed unreachable. Examples provided include a two-dimensional Penrose tile (a model of a quasicrystal), non-Hermitian Hamiltonians in superconductor theory and optics, and partial differential operators such as Schrödinger operators.

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 to have 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 pseudospectrum when approximating pseudospectra, so 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 (1.4.2). Indeed, both can be viewed in terms of metrics on spaces of continuous functions [Bee93]. In other words, we seek local uniform convergence. We must also 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 given in (1.4.1).

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 3.1.10 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, as discussed in [CRH19]. Some of the results of this chapter also build upon and extend work done by the author in collaboration in [BACH+19] and classification results higher up in the SCI hierarchy can be found in [BACH+19].

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. We consider the problems of computing

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

To define the computational problem we have to define the domain Ω as well as Λ, the set of evaluation functions. 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^*, where $\{e_j\}_{j\in\mathbb{N}}$ is the canonical basis for $l^2(\mathbb{N})$.

(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 $\{e_1, \ldots, e_n\}$ of the canonical basis. 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).

The concept of bounded dispersion in (3.1.1) generalises the notion of a banded or sparse matrix to knowledge of off-diagonal decay of our operator viewed as a matrix. Moreover, 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.

In order 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. 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 $\mathrm{dist}(z, \mathrm{Sp}(A)) = \|R(z, A)\|^{-1}$. 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 $m \in \mathbb{N}$.*

We can now state our main result in this section:

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*

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

and in the case of Ξ_2, the output of the constructed algorithm is guaranteed to be inside 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. 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.*

The algorithm used to compute the pseudospectrum can be applied to cases where the spectrum or pseudospectrum are empty and we provide a numerical example of this below. Finally, we consider two discrete problems which also include the case when the spectrum may be empty. Let K be a non-empty and compact set in $\mathbb{C}$ and denote the collection of such subsets by $\mathcal{K}(\mathbb{C})$. Consider

$$\Xi_3 : (A, K) \to \text{Is } \mathrm{Sp}(A) \cap K = \emptyset?$$
$$\Xi_4 : (A, K) \to \text{Is } \mathrm{Sp}_\epsilon(A) \cap K = \emptyset?$$

More precisely, 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 $d_{\mathrm{H}}(K_n, K) \leq 2^{-(n+1)}$.[2] 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. We also see 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 $\mathbb{R}$.

Remark 3.1.7. *By considering singletons $K = \{z\}$, we can test whether a point lies in the spectrum or pseudospectrum. Even when restricting to such K, the proof makes clear that the classification remains the same.*

3.1.2 Spectra of partial differential operators

In this section, we provide classification results for general classes of differential operators. What may be surprising is that with very general assumptions, we obtain Σ_1^A classifications for the spectrum. This means that despite these operators being hard to analyse for spectral theoretical purposes, the problem of computing their spectra is not harder than computing the spectra of diagonal matrices (see §1.1). 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 throughout 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. We consider two cases. First, the algorithm can access point samples of the functions, and second, the algorithm can access coefficients in the series expansion of the functions (in the case that the a_k are analytic on the whole of $\mathbb{R}^d$). Note that these are very different computational problems.

The set-up

To make our problems well-defined, we let Ω consist of all such T such that the following assumptions hold:

[2] This is an example where functions in Λ take values in $\mathbb{C}^2$ - for a given n the first coordinate tells us how many points are in K_n, then we can use a bijection $\mathbb{C}^{|K_n|} \leftrightarrow \mathbb{C}$ to encode the set K_n in the second coordinate

(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 of the functions $a_k(x)$ 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 the case of §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 3.1.8 (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, our algorithm is the first that computes the spectrum of such operators with error control in the sense of Σ_1^A. As described, this has a wide number of applications and the problem has received a lot of attention [BO13, Wen96, BW73, FMT89].

Remark 3.1.9. *Throughout this section, the functions $\{g_m\}$ are not needed to compute the pseudospectrum.*

General case with function evaluations

In this section 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

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

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^2_{TV} to denote the class of problems where we have this weaker requirement. We can now define the mappings

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

and state the first theorem.

Theorem 3.1.10. *Let $\Xi^1_j, \Xi^2_j, \Omega^1_{\mathrm{TV}}$ and Ω^2_{TV} be as above. Then for $j = 1, 2$*

$$\Delta^G_1 \not\ni \{\Xi^1_j, \Omega^1_{\mathrm{TV}}\} \in \Sigma^A_1,$$
$$\Sigma^G_1 \cup \Pi^G_1 \not\ni \{\Xi^2_j, \Omega^2_{\mathrm{TV}}\} \in \Delta^A_2.$$

The proof also shows the stronger result that even if we had included the information $\{A_k, B_k\}_{|k|\leq N}$ for operators in Ω^2_{TV}, we would still have $\{\Xi^2_j, \Omega^2_{\mathrm{TV}}\} \notin \Sigma^G_1 \cup \Pi^G_1$.

Remark 3.1.11. *This result is of interest since it gives a computational problem where no Σ or Π error control is available in its corresponding Δ (SCI) class.*

Analytic coefficients

In this section, we assume that the functions a_k and $\widetilde{a}_k$ are analytic on the whole of $\mathbb{R}^d$. In particular, we assume we can evaluate $\{c_j\}_{j\in\mathbb{N}}$, an enumeration (where we know the ordering) of the coefficients of the power series of each of the $a_k(x)$. In this special case, we can compute the corresponding coefficients of the $\widetilde{a}_k(x)$ using finitely many arithmetic operations on $\{c_j\}$. We will assume that as well as the information $\{g_m\}$, $\{c_j\}$ and $\{A_k, B_k\}$, our algorithms can read the following information. Given

$$a_k(x) = \sum_{m\in(\mathbb{Z}_{\geq 0})^d} a^m_k x^m, \quad \widetilde{a}_k(x) = \sum_{m\in(\mathbb{Z}_{\geq 0})^d} \widetilde{a}^m_k x^m,$$

for each $n \in \mathbb{N}$ we know a constant d_n such that

$$|a_k^m|, |\widetilde{a}_k^m| \leq d_n(n+1)^{-|m|}, \qquad \forall m \in (\mathbb{Z}_{\geq 0})^d, |k| \leq N. \tag{3.1.6}$$

It is straightforward to show that such a d_n must exist using the fact that the power series converges absolutely on the whole of $\mathbb{R}^d$. Let

$$\Omega_{\mathrm{AN}}^1 = \{T \in \Omega \mid \text{ such that (1) – (5), the functions } a_k \text{ are analytic and (3.1.6) hold}\}.$$

Moreover, in this case we let Λ^1 contain functions that allow us to access sample of the functions $\{g_m\}_{m\in\mathbb{N}}$, the constants $\{A_k, B_k\}_{|k|\leq N}, \{c_n\}_{n\in\mathbb{N}}$, and $\{d_n\}_{n\in\mathbb{N}}$. As the proof makes clear, the information d_n can be replaced by any suitable information that allows us to control the remainder term in the truncated Taylor series uniformly on compact subsets of $\mathbb{R}^d$. For example, we could use Cauchy's formula, together with bounds on the functions a_k on compact subsets of $\mathbb{C}^d$. One could consider a weaker requirement on Λ^1 by replacing knowledge of A_k, B_k and d_n by some sequence of positive numbers b_n with

$$\sup_{n\in\mathbb{N}} \sup_{m\in(\mathbb{Z}_{\geq 0})^d} \frac{\max\{|a_k^m|\,(n+1)^{|m|}, |\widetilde{a}_k^m|\,(n+1)^{|m|} : |k| \leq N\}}{b_n} < \infty.$$

With a slight abuse of notation, we use Ω_{AN}^2 to denote the class of problems where we have this weaker requirement. Moreover, let Ω_p denote the class of operators in Ω_{AN}^2 such that each a_k is a polynomial (where we can let b_n be $n!$ say). We can now define the mappings

$$\Xi_j^3, \Xi_j^4 : \Omega_{\mathrm{AN}}^1, \Omega_{\mathrm{AN}}^2, \Omega_p \ni T \mapsto \begin{cases} \mathrm{Sp}(T) \in \mathcal{M}_{\mathrm{AW}}, & j = 1 \\ \mathrm{Sp}_\epsilon(T) \in \mathcal{M}_{\mathrm{AW}}, & j = 2, \end{cases}$$

and state the second theorem.

Theorem 3.1.12. *Let $\Xi_j^3, \Xi_j^4, \Omega_{\mathrm{AN}}^1$, Ω_{AN}^2 and Ω_p be as above. Then for $j = 1, 2$*

$$\Delta_1^G \not\ni \{\Xi_j^3, \Omega_{\mathrm{AN}}^1\} \in \Sigma_1^A, \qquad \Sigma_1^G \cup \Pi_1^G \not\ni \{\Xi_j^4, \Omega_{\mathrm{AN}}^2\} \in \Delta_2^A, \qquad \Sigma_1^G \cup \Pi_1^G \not\ni \{\Xi_j^4, \Omega_p\} \in \Delta_2^A.$$

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 $P_{f(n)}(A - zI)P_n$. 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

$$E_n(z) \approx \sigma_1(P_{f(n)}(A - zI)|_{P_n(l^2(\mathbb{N}))}).$$

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 $n \to \infty$. 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 (and do not converge in general - see §7.1).

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 $d_{\mathrm{AW}}(C_n, C) \to 0$ holds if and only if $d_K(C_n, C) \to 0$ 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.

Proof. For the proof of (1), we may take $a_n^m = \sup\{E_n(z) : z \in \Gamma_n(A) \cap B_m(0)\}$ and the result follows. Note that we need $\Gamma_n(A)$ to be finite to be able to compute this number with finitely many arithmetic operations and comparisons. We next show (2) by defining

$$A_n^m = \big((\Xi(A) + B_{a_n^m}(0)) \cap B_m(0)\big) \cup (\Gamma_n(A) \cap \{z : |z| \geq m\}).$$

It is clear that $\Gamma_n(A) \subset A_n^m$ and given $\Gamma_n(A)$ we can easily compute a lower bound m_1 such that $\Xi(A) \cap B_{m_1}(0) \neq \emptyset$. Compute this from $\Gamma_1(A)$ and then fix it. Suppose that $m \geq 4m_1$, and suppose that $|z| < \lfloor m/4 \rfloor$. Then the points in A_n^m and $\Xi(A)$ nearest to z must lie in $B_m(0)$ and hence

$$\mathrm{dist}(z, A_n^m) \leq \mathrm{dist}(z, \Xi(A)), \quad \mathrm{dist}(z, \Xi(A)) \leq \mathrm{dist}(z, A_n^m) + a_n^m.$$

It follows that

$$d_{\mathrm{AW}}(A_n^m, \Xi(A)) \leq a_n^m + 2^{-\lfloor m/4 \rfloor}.$$

We now choose a sequence $m(n)$ such that setting $A_n = A_n^{m(n)}$ and $b_n = a_n^{m(n)} + 2^{-\lfloor m(n)/4 \rfloor}$ proves the result. Clearly it is enough to ensure that b_n converges to zero. If $n < 4m_1$ then set $m(n) = 4m_1$, otherwise consider $4m_1 \leq k \leq n$. If such a k exists with $a_n^k \leq 2^{-k}$ then let $m(n)$ be the maximal such k and finally if no such k exists then set $m(n) = 4m_1$. For a fixed m, $a_n^m \to 0$ as $n \to \infty$. It follows that for large n, $a_n^{m(n)} \leq 2^{-m(n)}$ and that $m(n) \to \infty$. □

Remark 3.2.3. *We will only consider algorithms where the output of $\Gamma_n(A)$ is at most finite for each n. Hence the above restriction does not matter in what follows.*

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

$$\sigma_1(A) = \inf\{\|Ax\| : x \in \mathcal{D}(A), \|x\| = 1\}, \tag{3.2.1}$$

and define the function

$$\gamma(z, A) = \min\{\sigma_1(A - zI), \sigma_1(A^* - \bar{z}I)\}.$$

Lemma 3.2.4. *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 $z \in \mathrm{Sp}(A)$.*

Proof. We deal with the case $z \notin \mathrm{Sp}(A)$ first, where we prove $\sigma(A - zI) = \sigma(A^* - \bar{z}I) = 1/\|R(z, A)\|$. We show this for $\sigma_1(A - zI)$ and the other case is similar using the fact that $R(z, A)^* = R(\overline{z}, A^*)$ and $\|R(z, A)\| = \|R(z, A)^*\|$. Let $x \in \mathcal{D}(A)$ with $\|x\| = 1$ then

$$1 = \|R(z, A)(A - zI)x\| \leq \|R(z, A)\| \, \|(A - zI)x\|$$

and hence upon taking infinum, $\sigma_1(A - zI) \geq 1/\|R(z, A)\|$. Conversely, let $x_n \in l^2(\mathbb{N})$ such that $\|x_n\| = 1$ and $\|R(z, A)x_n\| \to \|R(z, A)\|$. It follows that

$$1 = \|(A - zI)R(z, A)x_n\| \geq \sigma_1(A - zI) \, \|R(z, A)x_n\| .$$

Letting $n \to \infty$ we get $\sigma_1(A - zI) \leq 1/\|R(z, A)\|$.

Now suppose that $z \in \mathrm{Sp}(A)$. If at least one of $A - zI$ or $A^* - \bar{z}I$ is not injective on their respective domain then we are done, so assume both are one to one. Suppose also that $\sigma_1(A - zI), \sigma_1(A^* - \bar{z}I) > 0$ otherwise we are done. It follows that $\mathcal{R}(A - zI)$ is dense in $l^2(\mathbb{N})$ by injectivity of $A^* - \bar{z}I$ since $\mathcal{R}(A - zI)^\perp = N(A^* - \bar{z}I)$. It follows that we can define $(A - zI)^{-1}$, bounded on the dense set $\mathcal{R}(A - zI)$. We can extend this inverse to a bounded operator on the whole of $l^2(\mathbb{N})$. Closedness of A now implies that $(A - zI)(A - zI)^{-1} = I$. Clearly $(A - zI)^{-1}(A - zI)x = x$ for all $x \in \mathcal{D}(A)$. Hence, $(A - zI)^{-1} = R(z, A) \in \mathcal{B}(l^2(\mathbb{N}))$ so that $z \notin \mathrm{Sp}(A)$, a contradiction. □

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 3.2.5. *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

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

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

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

There exists N_1 such that if $n \geq N_1$ then (3.2.4) holds and $2/n + g^{-1}_{m+1}(\epsilon_n + 1/n) \leq \delta$ 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

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

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 3.2.6. *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*

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

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 $y_{n_j} \in \mathtt{Grid}(n_j)$ with $\left|y_{n_j} - w\right| \leq 1/n_j$. 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 $\left|y_{n_j} - z\right| \leq |z-w| + \left|y_{n_j} - w\right| \leq \delta/2 + 1/n_j$, 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 3.2.7. *Let $A \in \hat{\Omega}$ and define the function*

$$\tilde{\gamma}_n(z,A) = \min\{\sigma_1(P_{f(n)}(A - zI)|_{P_n(l^2(\mathbb{N}))}), \sigma_1(P_{f(n)}(A^* - \bar{z}I)|_{P_n(l^2(\mathbb{N}))})\}.$$

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 $\sigma_1((A - zI)|_{P_n(l^2(\mathbb{N}))}) \downarrow \sigma_1(A - zI)$ as $n \to \infty$. It is trivial that $\sigma_1((A - zI)|_{P_n(l^2(\mathbb{N}))}) \geq \sigma_1(A - zI)$ and that $\sigma_1((A - zI)|_{P_n(l^2(\mathbb{N}))})$ is non-increasing in n. Using Lemma 3.2.4, let $\epsilon > 0$ and $x \in \mathcal{D}(A)$ such that $\|x\| = 1$ and $\|(A - zI)x\| \leq \sigma_1(A - zI) + \epsilon$. 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

$$\sigma_1((A - zI)|_{P_{n_j}(l^2(\mathbb{N}))}) \leq \frac{\left\|(A - zI)P_{n_j}x_{n_j}\right\|}{\left\|P_{n_j}x_{n_j}\right\|} \to \|(A - zI)x\| \leq \sigma_1(A - zI) + \epsilon.$$

Since $\epsilon > 0$ was arbitrary, this shows the convergence of $\sigma_1((A - zI)|_{P_n(l^2(\mathbb{N}))})$. The fact that $\mathrm{span}\{e_n : n \in \mathbb{N}\}$ forms a core of A^* can also be used to show that $\sigma_1((A - zI)^*|_{P_n(l^2(\mathbb{N}))}) \downarrow \sigma_1(A^* - \bar{z}I)$.

Next we will use the assumption of bounded dispersion. For any bounded operators B, C, it holds that $|\sigma_1(A) - \sigma_1(B)| \leq \|A - B\|$. The definition of bounded dispersion now implies that

$$\left|\tilde{\gamma}_n(z,A) - \min\{\sigma_1((A - zI)|_{P_n(l^2(\mathbb{N}))}), \sigma_1((A - zI)^*|_{P_n(l^2(\mathbb{N}))})\}\right| \leq c_n.$$

The monotone convergence of $\min\{\sigma_1((A - zI)|_{P_n(l^2(\mathbb{N}))}), \sigma_1((A - zI)^*|_{P_n(l^2(\mathbb{N}))})\}$, 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 and Lemma 3.2.8 below, 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. □

Recall that every finite Hermitian matrix B (not necessarily positive definite) has a decomposition

$$PBP^T = LDL^*,$$

where L is lower triangular with 1's along its diagonal, D is block diagonal with block sizes 1 or 2 and P is a permutation matrix. Furthermore, this decomposition can be computed with finitely many arithmetic operations and comparisons. Throughout, we will assume without loss of generality that P is the identity matrix.

Lemma 3.2.8. *Let $B \in \mathbb{C}^n$ be self-adjoint (Hermitian), then we can determine the number of negative eigenvalues of B in finitely many arithmetic operations and comparisons (assuming no round-off errors) on the matrix entries of B.*

Proof. We can compute the decomposition $B = LDL^*$ in finitely many arithmetical operations and comparisons. By Sylvester's law of inertia (the Hermitian version), D has the same number of negative eigenvalues as B. It is then clear that we only need to deal with 2×2 matrices corresponding to the maximum block size of D. Let λ_1, λ_2 be the two eigenvalues of such a matrix, then we can determine their sign pattern from the trace and determinant of the matrix. □

This lemma has a corollary that will be useful in §6.3.

Corollary 3.2.9. *Let $B \in \mathbb{C}^n$ be self-adjoint (Hermitian) and list its eigenvalues in increasing order, including multiplicity, as $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$. In exact arithmetic, given $\epsilon > 0$, we can compute $\lambda_1, \lambda_2, ...\lambda_n$ to precision ϵ using only finitely many arithmetic operations and comparisons.*

Proof. Consider $A(\lambda) = B - \lambda I$. We will apply Lemma 3.2.8 to $A(\lambda)$ for various λ. First by considering the sequences $-1, -2, ...$ and $1, 2, ...$ we can find $m_1 \in \mathbb{N}$ such that $\mathrm{Sp}(B) \subset (-m_1, m_1)$. Now let $m_2 \in \mathbb{N}$ such that $1/m_2 < \epsilon$ and let a_j be the output of Lemma 3.2.8 applied to $A(j/m_2)$ for $-m_1 m_2 \leq j \leq m_1 m_2$. Set

$$\tilde{\lambda}_k = \min\{j : -m_1 m_2 \leq j \leq m_1 m_2, a_j \geq k\}, \quad k = 1, ..., n.$$

If $\lambda_k \in [j/m_2, (j+1)/m_2)$ then $\tilde{\lambda}_k = (j+1)/m_2$ and hence $\left|\tilde{\lambda}_k - \lambda_k\right| \leq 1/m_2 < \epsilon$. □

Remark 3.2.10. *Of course, in practice, there are much more computationally efficient ways to numerically compute eigenvalues or singular values - the above is purely used to show this can be done to any precision with finitely many arithmetic operations.*

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. We can now prove our main result.

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 3.2.7 together with Proposition 3.2.5 and its described algorithm. For the pseudospectrum, we use the same function γ_n described in Theorem 3.2.7 and convergence follows from using the algorithm in Proposition 3.2.6.

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 3.2.6 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 3.2.5 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

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

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 $\Gamma_{n_j}(A)$, $z \in \operatorname{Sp}(A)$ 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 3.2.11. *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 $A_m = \operatorname{diag}\{1, 1/2, ..., 1/m\} \in \mathbb{C}^{m \times m}$, $B_m = \operatorname{diag}\{1, 1, ..., 1\} \in \mathbb{C}^{m \times m}$ and $C = \operatorname{diag}\{1, 1, ...\}$. 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. $\operatorname{Sp}(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C) = \{1, 1/2, ..., 1/m\}$ 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 Theorems 3.1.10 and 3.1.12. 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. In order to do so, we must first select a suitable basis and then compute matrix values. Recall that our aim is to compute the spectrum and pseudospectrum from the information given to us regarding the functions a_k and $\widetilde{a}_k$, with the information we can evaluate made precise by the mappings Ξ_j^1, Ξ_j^2, Ξ_j^3 and Ξ_j^4. We will start by constructing the algorithms used for the positive results in Theorems 3.1.10 and 3.1.12 and then prove the lower bounds.

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 $C_H(\mathbb{R}) = \text{span}\{\psi_m : m \in \mathbb{Z}_{\geq 0}\}$. 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^*.*

Proof. Let $f \in C_H(\mathbb{R})$ and choose $\phi \in C_0^\infty(\mathbb{R})$ (the space of compactly supported smooth functions) bounded by 1 such that $\phi(x) = 1$ for all $|x| \leq 1$. It is straightforward using the fact that the a_k's are polynomially bounded to show that

$$\lim_{n\to\infty} \phi(x/n)f(x) = f(x), \quad \lim_{n\to\infty} T\phi(x/n)f(x) = (Tf)(x)$$

in $L^2(\mathbb{R})$, where Tf is the formal differential operator applied to f. The fact that T is closed implies that $f \in \mathcal{D}(T)$. Let $\tilde{T}$ denote the closure of the formal operator T, acting on $C_H(\mathbb{R})$, then we have shown that $\tilde{T}$ exists with $\tilde{T} \subset T$. Hence to show that $C_H(\mathbb{R})$ forms a core of T, we must show that $C_0^\infty(\mathbb{R}) \subset \mathcal{D}(\tilde{T})$. Let $g \in C_0^\infty(\mathbb{R})$ then in the L^2 sense write

$$g = \sum_{m\geq 0} b_m \psi_m.$$

Define $g_n = \sum_{m=0}^n b_m\psi_m$ then, since $\tilde{T}$ is closed, it is enough to show that $\tilde{T}g_n$ converges as $n \to \infty$. Let H denote the closure of the operator $-d^2/dx^2 + x^2$ with initial domain $C_0^\infty(\mathbb{R})$ then $H\psi_m = (2m+1)\psi_m$ and H is self-adjoint. Note also that $g \in \mathcal{D}(H^n)$ for any $n \in \mathbb{N}$. But $\langle Hg, \psi_m\rangle = (2m+1)\langle g, \psi_m\rangle = (2m+1)b_m$, so $\{(2m+1)\,|b_m|\}$ is square summable. We can repeat this argument any number of times to get that the coefficients b_m decay faster than any inverse polynomial. To prove the required convergence, it is enough to consider one of the terms $a_k(x)\partial^k$ that defines $\tilde{T}$ acting on $C_H(\mathbb{R})$. The coefficient $a_k(x)$ is polynomially bounded almost everywhere, and for some A_k and B_k

$$\langle a_k\partial^k\psi_m, a_k\partial^k\psi_m\rangle \leq A_k^2 \int_{\mathbb{R}} (1+|x|^{2B_k})^2 \partial^k\psi_m(x)\partial^k\psi_m(x)dx.$$

But we can use the recurrence relations for the derivatives of the Hermite functions and orthogonality to bound the right hand side by a polynomial in m. The convergence now follows since $\tilde{T}g_n$ is a Cauchy sequence due to the rapid decay of the $\{b_m\}$. Exactly the same argument works for T^*. □

Clearly, all of the above analysis holds in higher dimensions by considering tensor products

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

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 $\mathbb{Z}_{\geq 0}^d$ 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 [Lub08], 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

$$\Psi_n(z, T) := \min\{\sigma_1((T - zI)|_{P_n(l^2(\mathbb{N}))}), \sigma_1((T^* - \bar{z}I)|_{P_n(l^2(\mathbb{N}))})\}$$

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*

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

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

$$\left|\Psi_n(z, T)^2 - \min\{\sigma_1(W_n), \sigma_1(V_n)\}\right| \leq n\epsilon.$$

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

$$\min\{\sigma_1((T - zI)|_{P_n(l^2(\mathbb{N}))}), \sigma_1((T^* - \bar{z}I)|_{P_n(l^2(\mathbb{N}))})\}^2 = \min\{\sigma_1(\tilde{W}_n), \sigma_1(\tilde{V}_n)\}$$

and

$$\left|\min\{\sigma_1(\tilde{W}_n), \sigma_1(\tilde{V}_n)\} - \min\{\sigma_1(W_n), \sigma_1(V_n)\}\right| \leq \max\left\{\left\|W_n - \tilde{W}_n\right\|, \left\|V_n - \tilde{V}_n\right\|\right\}. \tag{3.3.3}$$

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 $\sigma_1(M)$ to arbitrary precision using finitely many arithmetic operations and comparisons via the argument in the proof of Theorem 3.2.7. 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*

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

Proof. We will use an alternative form of the total variation which holds for smooth enough functions and can be found in [Nie92]:

$$\mathrm{TV}_{[-r,r]^d}(\psi_m) = \sum_{k=1}^{d} \sum_{1 \leq i_1 < ... < i_k \leq d} \int_{-r}^{r} ... \int_{-r}^{r} \left| \frac{\partial^k \psi_m}{\partial x_{i_1} ... \partial x_{i_k}}(\tilde{x}) \right| dx_{i_1} ... dx_{i_k},$$

where $\tilde{x}$ has $\tilde{x}_j = x_j$ for $j = i_1, ..., i_k$ and $\tilde{x}_j = r$ otherwise. We can use the recurrence relation (3.3.1) and the rough bound $|\psi_m(x)| \leq 1$ (which follows from Cramér's inequality which bounds the one-dimensional Hermite functions [Ind61]) to gain the bound

$$\int_{-r}^{r} ... \int_{-r}^{r} \left| \frac{\partial^k \psi_m}{\partial x_{i_1} ... \partial x_{i_k}}(\tilde{x}) \right| dx_{i_1} ... dx_{i_k} \leq \left(2r\sqrt{2(|m|+1)} \right)^k.$$

It follows that

$$\begin{aligned} \mathrm{TV}_{[-r,r]^d}(\psi_m) &\leq \sum_{k=1}^{d} \left(2r\sqrt{2(|m|+1)} \right)^k \sum_{1 \leq i_1 < ... < i_k \leq d} 1 \\ &= \sum_{k=1}^{d} \left(2r\sqrt{2(|m|+1)} \right)^k \binom{d}{k} = \left(1 + 2r\sqrt{2(|m|+1)} \right)^d - 1. \end{aligned}$$

□

Proposition 3.3.8. *Given $T \in \Omega^1_{\mathrm{TV}}$ or $T \in \Omega^1_{\mathrm{AN}}$ 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^1_j and Ξ^3_j in §3.1.2) given to us in each class.

Proof. Let $T \in \Omega^1_{\mathrm{TV}}$ or $T \in \Omega^1_{\mathrm{AN}}$ 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

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

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

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

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 $T \in \Omega^1_{\mathrm{TV}}$ or $T \in \Omega^1_{\mathrm{AN}}$ separately, noting that it is sufficient to approximate the integral $\int_{|x_i|\leq r}a_k\overline{a_j}\psi_m\psi_ndx$ 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 $\|\psi_m\|_{\mathcal{A}_r} \leq L_r(m)$.

Case 1: $T \in \Omega^1_{\mathrm{TV}}$**.** 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

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

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.

Case 2: $T \in \Omega^1_{\mathrm{AN}}$**.** On the compact cube $|x_i| \leq r$ the double series

$$a_k(x)\overline{a_j(x)} = \sum_{t\in(\mathbb{Z}_{\geq 0})^d} \sum_{s\in(\mathbb{Z}_{\geq 0})^d} a_k^t \overline{a_j^s} x^{t+s}$$

converges uniformly (recall that $\{a_k^t\}_{t\in(\mathbb{Z}_{\geq 0})^d}$ are the power series coefficients for a_k) so we can exchange the series and integration to write

$$\int_{|x_i|\leq r} a_k\overline{a_j}\psi_m\psi_n dx = \sum_{t,s\in(\mathbb{Z}_{\geq 0})^d} a_k^t\overline{a_j^s} \int_{|x_i|\leq r} x^{s+t}\psi_m(x)\psi_n(x)dx. \tag{3.3.6}$$

But $\left|\int_{|x_i|\leq r} x^{s+t}\psi_m(x)\psi_n(x)dx\right|$ is bounded by

$$r^{|t|+|s|}\int_{x\in\mathbb{R}^d} |\psi_m|\,|\psi_n|\,dx \leq r^{|t|+|s|}\left(\int_{x\in\mathbb{R}^d} |\psi_m|^2\,dx\right)^{1/2}\left(\int_{x\in\mathbb{R}^d} |\psi_n|^2\,dx\right)^{1/2} = r^{|t|+|s|},$$

where we have used Hölder's inequality and the fact that the tensorised Hermite functions are orthonormal. Let $\tau = r/(r+1)$, then using the fact that we know d_r in (3.1.6), we can bound the tail of the series in (3.3.6) by

$$d_r^2 \sum_{|t|,|s|>M} \tau^{|t|+|s|} \leq d_r^2\left(\sum_{|t|>M} \tau^{\frac{|t_1|}{d}+...+\frac{|t_d|}{d}}\right)^2,$$

using the fact that $|x| \leq (|x_1| + ... + |x_d|)/d$. We can explicitly sum this series (as the difference of geometric series) to gain the bound

$$d_r^2\left[\frac{1-(1-\tau^{(M+1)/d})^d}{\left(1-\tau^{1/d}\right)^d}\right]^2.$$

Given r and d_r (and d) we can keep increasing M and evaluating the bound (strictly speaking an upper bound accurate to $1/M$ say), to choose M large such that the tail is smaller than $\delta/2$ for any given $\delta > 0$. It follows that it is enough to estimate integrals of the form $\int_{|x_i|\leq r} x^{s+t}\psi_m(x)\psi_n(x)dx$. Using the recurrence relations for Hermite functions and writing $\psi_m(x) = Q_m(x)\exp(-x^2/2)$, it is enough to split the multidimensional integral up as products and sums of one-dimensional integrals of the form $\int_{-r}^{r} x^a\exp(-x^2)dx$, for $a \in \mathbb{Z}_{\geq 0}$. Again, we have assumed that we can compute the coefficients of the Q_m to any given accuracy using finitely many arithmetic operations and comparisons and using this we can bound the total error of the expression by $\delta/2$. The above integral vanishes unless a is even, so integration by parts (again assuming we can evaluate $\exp(-x^2)$ to any desired accuracy) reduces this to estimating $\int_{-r}^{r}\exp(-x^2)dx$. Consider the Taylor series for $\exp(-x^2)$. The tail can be bounded by

$$\sum_{k>N}\frac{r^{2k}}{k!} \leq \frac{r^{2N}}{N!}\exp(-r^2).$$

Integrating this estimate over the interval $[-r, r]$, we can bound this by any given $\eta > 0$ by choosing N large enough. We can then explicitly compute $\int_{-r}^{r} \sum_{k \leq N} x^{2k}/k! dx$. Keeping track of all the errors is elementary and hence $\int_{|x_i| \leq r} a_k \overline{a_j} \psi_m \psi_n dx$ can be approximated with finitely many arithmetic operations and comparisons as required. □

In some cases, we can also directly compute matrix elements without the cut-off argument used in the above proof. For instance, if each $a_k(x)$ (and hence $\widetilde{a}_k(x)$) is a polynomial then we can simply integrate the power series to compute $\langle a_k(x)\psi_m, a_j(x)\psi_n \rangle$ and use the recurrence relations for Hermite functions. If we know a bound on the degree of the polynomials, then clearly we can compute

$$\langle (T - zI)\psi_m, (T - zI)\psi_n \rangle \quad \text{and} \quad \langle (T - zI)^* \psi_m, (T - zI)^* \psi_n \rangle \tag{3.3.7}$$

to within ϵ using finitely many arithmetical operations and comparisons directly. Even if we do not know the degree of the polynomials and are only promised that each $a_k(x)$ is a polynomial, then we can successively approximate by more terms of the power series and eventually compute (3.3.7) to within ϵ using finitely many arithmetical operations and comparisons. However, we do not know when the given accuracy has been reached (recall that we only know a finite portion of the coefficients $c_1, c_2, ...$ at any one time for $T \in \Omega^1_{\mathrm{AN}}$).

We can now prove the positive parts of Theorems 3.1.10 and 3.1.12.

Proof of inclusions in Theorems 3.1.10 and 3.1.12. **Step 1**: $\{\Xi^1_1, \Omega^1_{\mathrm{TV}}\}, \{\Xi^3_1, \Omega^1_{\mathrm{AN}}\} \in \Sigma^1_A$. 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

$$\left| \upsilon_n(z,T)^2 - \min\{\sigma_1((T - zI)|_{P_n(l^2(\mathbb{N}))}), \sigma_1((T^* - \bar{z}I)|_{P_n(l^2(\mathbb{N}))})\}^2 \right| \leq \frac{1}{n^2}.$$

We now set

$$\gamma_n(z, T) = \upsilon_n(z, T) + 1/n. \tag{3.3.8}$$

Then γ_n satisfies the hypotheses of Proposition 3.2.5. 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: $\{\Xi^1_2, \Omega^1_{\mathrm{TV}}\}, \{\Xi^3_2, \Omega^1_{\mathrm{AN}}\} \in \Sigma^1_A$. Consider the sequence of functions γ_n defined by equation (3.3.8). 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 3.2.6.

Step 3: $\{\Xi^2_1, \Omega^2_{\mathrm{TV}}\}, \{\Xi^2_2, \Omega^2_{\mathrm{TV}}\} \in \Delta^A_2$. Let $T \in \Omega^2_{\mathrm{TV}}$. 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

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

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

$$r^d \cdot \frac{\big(\log(M)+1\big)^d}{M} \cdot b_r^2 \cdot L_r(m) \cdot L_r(n), \tag{3.3.9}$$

where M is the number of Halton points. We can let M depend on r, n and m such that (3.3.9) 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

$$\left|\upsilon_n(z,T)^2 - \min\{\sigma_1((T-zI)|_{P_n(l^2(\mathbb{N}))}), \sigma_1((T^*-\bar{z}I)|_{P_n(l^2(\mathbb{N}))})\}^2\right| \leq \frac{n(1+|z|^2)Q(n)}{r(n,z)^2} + \frac{1}{n^3}. \tag{3.3.10}$$

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 3.2.5 and Proposition 3.2.6 since the error in (3.3.10) decays faster than $1/n^2$. We can use these propositions to build the required arithmetical algorithm.

Step 4: $\{\Xi_1^4, \Omega^2_{\mathrm{AN}}\}, \{\Xi_2^4, \Omega^2_{\mathrm{AN}}\} \in \Delta_2^A$. We argue as in step 3. To control the error in the approximation of the integral over a compact hypercube, choose the cut-off $M(r)$ such that

$$\sum_{|t|,|s|>M(r)} \left(\frac{r}{r+1}\right)^{|t|+|s|} \leq \frac{1}{b_r^2 r^2}.$$

It follows that there exists some (unknown) constant B such that we can bound the error in approximating $\int_{|x_i|\leq r} a_k\overline{a_j}\psi_m\psi_n dx$ by B/r^2 where we have absorbed the arbitrarily small error that comes from approximating the integral of the truncated power series using finitely many arithmetic operations and comparisons. The rest of the argument is the same as in step 3. □

3.3.2 Proofs of impossibility results

We first deal with Theorem 3.1.10. Recall the maps

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

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

Proof that $\{\Xi_j, \Omega^1_{\mathrm{TV}}\} \notin \Delta^G_1$. Suppose first for a contradiction that a height one tower, Γ_n, exists for the problem $\{\Xi_1, \Omega^1_{\mathrm{TV}}\}$ 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

$$(T_0 u)(x) = 0, \qquad (T_m u)(x) = \rho_m(x)u(x),$$

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 $\Gamma_n(T_0) = \Gamma_n(\epsilon T_m)$. □

Proof that $\{\Xi_j, \Omega^2_{\mathrm{TV}}\} \notin \Sigma^G_1 \cup \Pi^G_1$. Suppose first of all that a Σ^G_1 tower, Γ_n, exists for $\{\Xi_1, \Omega^2_{\mathrm{TV}}\}$. We will deal with the one-dimensional case and higher dimensions are similar. Consider the operators

$$(T_0 u)(x) = 0, \qquad (T_1 u)(x) = f(x)u(x),$$

where we define f in terms of Γ_n as follows. We choose f so that $f(x) = 1$ except for finitely many values of x where it takes the value 0 and hence T_0 and T_1 have spectra $\{0\}$ and $\{1\}$ respectively and are both self-adjoint. Note that once the zeros of f are fixed, this choice ensures that f has total variation bounded by a constant on any hypercube and hence we may take $b_r = 1$ for all $r \in \mathbb{N}$. There exists some n such that $\Gamma_n(T_0)$ contains $z_n \in B_{1/8}(0)$ with a guaranteed error estimate of $\mathrm{dist}(z_n, \mathrm{Sp}(T_0)) \leq 1/4$. But $\Gamma_n(T_0)$ can only depend on finitely many evaluations of 0 (as well as $b_r = 1$ and the trivial choice of $g_j(x) = x$). We choose f to be zero at precisely these evaluation points. By assumptions (ii) and (iii) in Definition 2.1.1, $\Gamma_n(T_1) = \Gamma_n(T_0)$, including the given error estimates, which is the required contradiction.

For $\{\Xi_2, \Omega^2_{\mathrm{TV}}\} \notin \Sigma^G_1$, given $\epsilon > 0$ we replace f by $3\epsilon f$ in the above argument and keep all other inputs the same. Hence T_0 and T_1 have ϵ-pseudospectra $[-\epsilon, \epsilon]$ and $[2\epsilon, 4\epsilon]$ respectively. We note that again there exists some n such that $\Gamma_n(T_0)$ contains $z_n \in B_{\epsilon/8}(0)$ with a guaranteed error estimate of $\mathrm{dist}(z_n, \mathrm{Sp}_\epsilon(T_0)) \leq \epsilon/4$. But $\Gamma_n(T_0)$ can only depend on finitely many evaluations of 0 (as well as $b_r = 1$ and the trivial choice of $g_j(x) = x$). We choose f to be zero at precisely these evaluation points. By assumptions (ii) and (iii) in Definition 2.1.1, $\Gamma_n(T_1) = \Gamma_n(T_0)$, including the given error estimates, which is the required contradiction.

To argue that neither problem lies in Π^G_1, we can use the same arguments in the proof that $\{\Xi_j, \Omega^1_{\mathrm{TV}}\} \notin \Delta^G_1$. The only change now is that the algorithm, Γ_n, used to derive the contradiction provides Π^G_1 information rather than Δ^G_1. For the spectrum, we consider the operators

$$(T_0 u)(x) = 0 \quad \text{and} \quad (T_m u)(x) = \rho_m(x)u(x),$$

and choose n large such that $\Gamma_n(T_0)$ produces the guarantee $\mathrm{Sp}(T_0) \cap B_{1/4}(0)^c = \emptyset$. For m sufficiently large, we argue as before to get $\Gamma_n(T_m) = \Gamma_n(T_0)$, including the guarantee, the required contradiction. Again a similar argument works for the pseudospectrum by rescaling T_m to $2\epsilon T_m$. □

We now deal with the impossibility results in Theorem 3.1.12 where

$$\Xi_j^3, \Xi_j^4 : \Omega_{\mathrm{AN}}^1, \Omega_{\mathrm{AN}}^2 \mapsto \begin{cases} \mathrm{Sp}(T) \in \mathcal{M}_{\mathrm{AW}} & j = 1 \\ \mathrm{Sp}_\epsilon(T) \in \mathcal{M}_{\mathrm{AW}} & j = 2. \end{cases}$$

Proof that $\{\Xi_j^3, \Omega_{\mathrm{AN}}^1\} \notin \Delta_1^G$. Suppose for a contradiction that a height one tower, Γ_n, exists for $\{\Xi_1^3, \Omega_{\mathrm{AN}}^1\}$ such that $d_{\mathrm{AW}}(\Gamma_n(T), \Xi_1^3(A)) \leq 2^{-n}$. Now consider the two (self-adjoint and bounded) operators

$$(T_1 u)(x) = 0 \quad \text{and} \quad (T_2 u)(x) = x^k \exp(-x^2) u(x)/s_k,$$

where k is even and will be chosen later. We choose s_k such that the range of the function $x^k \exp(-x^2)/s_k$ is $[0, 1]$ and hence T_2 has spectrum $[0, 1]$. We can take the polynomial bounding function to be the constant 1 for both operators and must show that we can use the same d_r for both operators in (3.1.6), independent of k. Simple calculus yields that $s_k = (k/(2e))^{k/2}$. It follows that such a d_r must satisfy

$$\left(\frac{2e}{k}\right)^{k/2} \frac{(r+1)^{2m+k}}{m!} \leq d_r, \quad \forall k \in 2\mathbb{N}, m \in \mathbb{Z}_{\geq 0}. \tag{3.3.11}$$

Hence it suffices to show that the function on the left hand side of (3.3.11) is bounded (as a function of m, k for all $r \in \mathbb{N}$). Using Stirling's approximation (explicitly the bounds on $m!$), this will follow if we can show that the right-hand side of

$$\frac{r^{2m+k}}{k^{k/2} m^{m+1/2}} \leq \left(\frac{r}{\sqrt{k}}\right)^k \left(\frac{r}{\sqrt{m}}\right)^{2m}$$

is bounded for all $r \in \mathbb{N}$ (now with $m > 1$). But this is obvious.

We can now choose k (which depends on the algorithm Γ_n) to gain a contradiction. Since $\mathrm{Sp}(T_1) = \{0\}$ and $1 \in \mathrm{Sp}(T_2)$ for all even k, there exists n such that $\mathrm{dist}(1, \Gamma_n(T_1)) > 1/4$ but $\mathrm{dist}(1, \Gamma_n(T_2)) < 1/4$. However, $\Gamma_n(T)$ can only depend on finitely many of the coefficients $\{c_j\}$, say $c_1, ..., c_{\tilde{N}(T,n)}$, of T (as well as the other given information). By assumption (iii) in Definition 2.1.1, we can choose k such that the coefficient corresponding to x^k, call it c_{l_k}, has $l_k > \tilde{N}(T_1, n)$ and get $\Gamma_n(T_1) = \Gamma_n(T_2)$, the required contradiction.

To show $\{\Xi_2^3, \Omega_{\mathrm{AN}}^1\} \notin \Delta_1^G$ uses exactly the same argument as above. In order to gain the necessary separation $3\epsilon \notin \mathrm{Sp}_\epsilon(T_1)$ but $3\epsilon \in \mathrm{Sp}_\epsilon(T_2)$, we rescale T_2 to $3\epsilon T_2$. Then there exists n such that $\mathrm{dist}(3\epsilon, \Gamma_n(T_1)) > \epsilon/2$ but $\mathrm{dist}(3\epsilon, \Gamma_n(T_2)) < \epsilon/2$. The rest of the contradiction follows. □

Proof that $\{\Xi_j^4, \Omega_{\mathrm{AN}}^2\}, \{\Xi_j^4, \Omega_p\} \notin \Sigma_1^G \cup \Pi_1^G$. Since $\Omega_p \subset \Omega_{\mathrm{AN}}^2$, it is enough to show the results for Ω_p.

Suppose for a contradiction that there exists a Σ_1^G algorithm, Γ_n, for $\{\Xi_1^4, \Omega_p\}$. Consider

$$(T_1 u)(x) = x u(x) \quad \text{and} \quad (T_2 u)(x) = (x - x^k) u(x),$$

where k is even and chosen later. $(T_j \pm iI) C_0^\infty(\mathbb{R})$ are dense in $L^2(\mathbb{R})$ with T_j initially defined on $C_0^\infty(\mathbb{R})$ symmetric. It follows that the closure of $T_j|_{C_0^\infty(\mathbb{R})}$ is self-adjoint and hence that $T_j \in \Omega_p$. Note that $\mathrm{Sp}(T_1) = \mathbb{R}$ but $\mathrm{Sp}(T_2) \subset (-\infty, 1]$. Now choose n such that $\Gamma_n(T_1)$ contains a point $z_n \in B_{1/4}(2)$ with a guaranteed error estimate of $\mathrm{dist}(z_n, \mathrm{Sp}(T_1)) \leq 1/4$. However, $\Gamma_n(T)$ can only depend on the

first $\tilde{N}(T,n)$ coefficients, $c_1, ..., c_{\tilde{N}(T,n)}$, of T (as well as the trivial choice $g_j(x) = x$ and the numbers $b_n = n!$). By assumption (iii) in Definition 2.1.1, we can choose k such that the coefficient corresponding to x^k, call it c_{r_k}, has $r_k > \tilde{N}(T_1, n)$ and get $\Gamma_n(T_1) = \Gamma_n(T_2)$, the required contradiction. Similarly by rescaling as above, we get $\{\Xi_2^4, \Omega_p\} \notin \Sigma_1^G$.

To show $\{\Xi_1^4, \Omega_p\} \notin \Pi_1^G$ we argue the same way, but now set $(T_1 u)(x) = 0$ and $(T_2 u)(x) = x^k u(x)$. As before, $T_j \in \Omega_p$, but now $\mathrm{Sp}(T_1) = \{0\}$ and $1 \in \mathrm{Sp}(T_2)$. Choose n such that $\Gamma_n(T_1)$ produces the guarantee $\mathrm{Sp}(T_1) \cap B_{1/4}(0)^c = \emptyset$. Again, choose k such that c_{r_k} has $r_k > \tilde{N}(T_1, n)$ and get $\Gamma_n(T_1) = \Gamma_n(T_2)$, the required contradiction. Rescaling and using the same argument shows $\{\Xi_2^4, \Omega_p\} \notin \Pi_1^G$. □

3.4 Computing Approximate States

The algorithms proposed above can also be used to gain states corresponding to elements in the spectrum in addition to the spectrum itself. For simplicity we will consider bounded operators on $l^2(\mathbb{N})$. For such an operator, not all of the spectrum is composed of eigenvalues. If the operator is normal then given $z \in \mathrm{Sp}(A)$ there exists a sequence of unit vectors $x_n \in l^2(\mathbb{N})$ such that

$$\|(A - zI)P_n x_n\| \to 0 \text{ as } n \to \infty.$$

Such a sequence is known as an approximate eigenvector sequence or an approximate eigenstate sequence. In the non-normal case, one only has the existence of x_n such that at least one of $\|(A - zI)P_n x_n\|$ and $\|(A^* - \overline{z}I)P_n x_n\|$ converge to zero. The question is whether given a z in the output $\Gamma_n(A)$ of the algorithm in §3.2 that converges to $\mathrm{Sp}(A)$ and the function $\gamma_n(z, A)$ in Theorem 3.2.7, we can find a x_n satisfying

$$\min\{\|(A - zI)x_n\|, \|(A^* - \overline{z}I)x_n\|\} \leq \gamma_n(z, A).$$

The convergence proof of the algorithm shows that $\{x_n\}$ will be an approximate eigenvector sequence.

Theorem 3.4.1 (Approximate States)**.** *Suppose A is a bounded operator with dispersion bounded by f. Given any $z \in \Gamma_n(A)$ with computed function $\gamma_n(z, A)$ of Theorem 3.2.7, we can compute a corresponding vector x_n satisfying*

$$\min\{\|(A - zI)P_n x_n\|, \|(A^* - \overline{z}I)P_n x_n\|\} \leq \gamma_n(z, A)$$

in finitely many arithmetic and square root operations.

Proof. We will deal with the normal case and note that dealing with the general case is simply a matter of applying the following argument to $(A^*, \overline{z})$ as well as (A, z). Recall that

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

where $\hat{\gamma}_n$ is a computable approximation of $\tilde{\gamma}_n(z, A) = \sigma_1(P_{f(n)}(A - zI)|_{P_n(l^2(\mathbb{N}))})$ to precision $1/n$.

Let $\epsilon = (\hat{\gamma}_n(z, A) + 1/n)^2$ and consider the matrix

$$B = P_n(A^* - \overline{z}I)P_{f(n)}(A - zI)P_n - \epsilon I$$

then B is a Hermitian matrix but not positive definite. It follows that B can be put into the form $PBP^T = LDL^*$, where L is lower triangular with 1's along its diagonal, D is block diagonal with block sizes 1 or 2 and P is a permutation matrix. This can be computed in finitely many arithmetic operations. Without loss

of generality we assume that $P = I$. Let x be an eigenvector of B with non-positive eigenvalue then set $y = L^*x$. Such an x exists by assumption. Note that

$$\langle y, Dy\rangle = \langle L^*x, DL^*x\rangle = \langle x, Bx\rangle \leq 0.$$

It follows that there exists a unit vector y_n with $\langle y_n, Dy_n\rangle \leq 0$. Such a vector is easy to spot by either considering 1 blocks or 2 blocks (where we need to extract square roots) in the block diagonal matrix D. L^* is invertible and upper triangular so we can efficiently solve for $\tilde{x}_n = (L^*)^{-1}y_n$ and then normalise to get x_n. Note that

$$\left\|P_{f(n)}(A - zI)P_n x_n\right\|^2 = \langle x_n, Bx_n\rangle + \epsilon = \frac{1}{\|\tilde{x}_n\|^2}\langle y_n, Dy_n\rangle + \epsilon \leq \epsilon.$$

It follows that

$$\|(A - zI)P_n x_n\| \leq c_n + \left\|P_{f(n)}(A - zI)P_n x_n\right\| \leq \hat{\gamma}_n(z, A) + c_n + 1/n = \gamma_n(z, A). \quad \square$$

The upshot of this is that the algorithm not only computes $\Gamma_n(A)$ converging to the spectrum of A, but it also computes approximating eigenvector sequences for the spectrum (and can do so for each point in the output of $\Gamma_n(A)$). Since not all of the spectrum is necessarily composed of eigenvalues in the infinite-dimensional case, this is the best any algorithm can hope to achieve in generality. This method is fast and can be efficiently implemented, as was done for Figure 3.3 below.

3.5 Numerical Implementation

Before demonstrating the algorithms of this chapter, we discuss their numerical implementation. We begin with simple pseudocode for the algorithms, which will be a useful reference in later chapters. We then discuss how to implement the algorithms efficiently.

3.5.1 Routines for core algorithms

The algorithm for the spectrum can be described by the routine `CompSpec`, shown as pseudocode below. Recall that this depends on the routines `Grid` and `CompInvg` described by (3.2.2) and (3.2.3) respectively. This relies on the approximation to $\|R(z, A)\|^{-1}$ in Theorem 3.2.7 given by the routine `DistSpec`.

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

Throughout we have used the fact that `DistSpec` (with a small modification - see Theorem 3.2.7 and the subsequent discussion) requires only finitely many arithmetic operations and comparisons. This is discussed further in §3.2 and details on fast implementation can be found in §3.5.2. In practice, we also replace the while loop by a much more efficient interval bisection method.

Function `CompSpec` $(A,n,\{g_m\},f(n),c_n)$
Input : $n \in \mathbb{N}$, $f(n) \in \mathbb{N}$, $c_n \in \mathbb{R}_+$ (bound on dispersion), $g_m : \mathbb{R}_+ \to \mathbb{R}_+$, $A \in \Omega_g$
Output: $\Gamma \subset \mathbb{C}$, an approximation to $\mathrm{Sp}(A)$, $E \in \mathbb{R}_+$, the error estimate
$G =$ `Grid` (n)
for $z \in G$ **do**
  $F(z) =$ `DistSpec` $(A,n,f(n),z)$
  if $F(z) \leq (|z|^2+1)^{-1}$ **then**
    for $w_j \in B_{\texttt{CompInvg}(n,F(z),g_{\lceil |z| \rceil})}(z) \cap G = \{w_1,...,w_k\}$ **do**
      $F_j =$ `DistSpec` $(A,n,f(n),w_j)$
    end
    $M_z = \{w_j : F_j = \min_q\{F_q\}\}$
  else
    $M_z = \emptyset$
  end
end
$\Gamma = \cup_{z\in G} M_z$
$E = \max_{z\in\Gamma}\{$`CompInvg` $(n,$`DistSpec` $(A,n,f(n),z) + c_n,\ g_{\lceil |z| \rceil})\}$
end

The algorithm for computing the pseudospectrum is shown in `PseudoSpec`.

Function `PseudoSpec` $(A,n,f(n),c_n,\epsilon)$
Input : $n \in \mathbb{N}$, $f(n) \in \mathbb{N}$, $c_n \in \mathbb{R}_+$, $A \in \hat{\Omega}$, $\epsilon > 0$
Output: $\Gamma \subset \mathbb{C}$, an approximation to $\mathrm{Sp}_\epsilon(A)$
$G =$ `Grid` (n)
for $z \in G$ **do**
  $F(z) =$ `DistSpec` $(A,n,f(n),z) + c_n$
end
$\Gamma = \bigcup\{z \in G \,|F(z) < \epsilon\}$
end

3.5.2 Efficient computation

Here we shall describe how to implement the algorithm for the spectrum efficiently. The main computational bottleneck is the computation of $\gamma_n(z, A)$ (or `DistSpec`) over a grid of points in Theorem 3.2.7, and we recall the algorithm outlined in its proof. The search routine for the smallest singular value can be efficiently implemented using an interval bisection method. To test for positive definiteness, we used an incomplete Cholesky decomposition. If our matrix A is sparse, we can take advantage of the fact that the matrices $B_n(z)$ and $C_n(z)$ have the same sparsity structure as we vary our test point z. We can calculate a permuta-

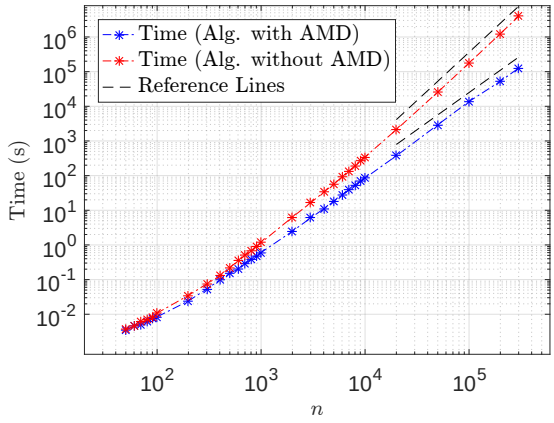


Figure 3.1: Speed-up of method when we take advantage of the structure preservation under changes in z and use AMD ordering. The AMD ordering only needs to be calculated once for each n and can subsequently be used on all test points. Both these plots are for the Laplacian H_0 on the Penrose tile discussed in §3.6.1.

tion of the indices corresponding to an Approximate Minimum Degree (AMD) ordering. This is a standard procedure to reduce the number of operations needed for Cholesky Decomposition or Gaussian Elimination. This can be computed with the MATLAB commands `[~,~,Q1]=chol(`B_n`-speye(`n`))` and `[~,~,Q2]=chol(`C_n`-speye(`n`))`. We can then replace B_n and C_n by $B_n(\texttt{Q1},\texttt{Q1})$ and $C_n(\texttt{Q2},\texttt{Q2})$ (using MATLAB's notation for matrix index ordering) in subsequent calculations. As shown in Figure 3.1 this offers considerable speed-up, especially in two-dimensional models, where the initial matrix A is not banded. For the case considered in §3.6.1, the time taken was of order $\sim \mathcal{O}(n^{2.1})$ and $\sim \mathcal{O}(n^{2.8})$ for large n with and without the AMD ordering respectively (shown as reference lines).

Of course, for large sparse rectangular truncations $P_n(A-zI)P_m$, there exist efficient iterative methods to approximate the smallest singular value. We found the partial Cholesky approach (with interval bisection and AMD ordering) slightly faster for the Penrose tile example in this chapter, but note that the user can easily use different subroutines for the computation of the smallest singular value. For the case of computing pseudospectra (e.g. Figure 3.5), using the partial Cholesky positive definite test is much more efficient since we can test levels of the resolvent norm on a logarithmic scale and we found it to be more stable for non-normal A. It is also much easier to implement the incomplete Cholesky approach when using interval arithmetic, allowing completely rigorous error bounds.

In many applications, such as finite-dimensional lattice models in condensed matter physics, we can bound f via $f(n) - n \sim Cn^{\alpha}$ for $\alpha \in [0,1)$. For example, $\alpha = 1/2$ for the Penrose lattice model considered in §3.6.1. The number of operations, pre-AMD ordering can then be bounded.

Proposition 3.5.1. *Let $A \in \hat{\Omega}$ and suppose that for large n, $f(n) - n \sim Cn^{\alpha}$, where f is the dispersion function, C a constant and $\alpha \in [0,1)$. Suppose that f is non-decreasing and also describes the off-diagonal sparsity structure of A in the sense that $A_{n,k}, A_{k,n} = 0$ if $k > f(n)$. If we use $m_1(n)$ test points and an accuracy of $1/m_2(n)$ for approximating $\tilde{\gamma}_n$ in Theorem 3.2.7, then the proposed algorithm for the spectrum can be computed in*

$$\mathcal{O}(m_1(n)n^{(\alpha+1)\alpha+1}\log(m_2(n)))$$

operations.

Proof. We first show that testing positive definiteness of $B := B_n - \epsilon I$ can be achieved in $\mathcal{O}(n^{(\alpha+1)\alpha+1})$ operations. It is then clear, using a binary search routine, that the computation of $\gamma_n(z, A)$ over the grid can be achieved in $\mathcal{O}(m_1(n)n^{(\alpha+1)\alpha+1}\log(m_2(n)))$ operations. The rest of the algorithm can be executed in $\mathcal{O}(m_1(n)n)$ operations, yielding the result.

To test positive definiteness we checked whether a Cholesky decomposition of the matrix B was possible. One can see that B also has a dispersion function $\tilde{f}(n) - n \sim \tilde{C}n^{\alpha}$ and hence without loss of generality we can assume $f = \tilde{f}$. Furthermore, B is sparse with f describing its sparsity structure. We refer the reader to [TBI97] Chapter 23 where Cholesky factorisation is explained. Following the notation there, one computes (assuming B positive definite)

$$B = R_1^*...R_m^*R_m...R_1$$

with $R = R_m...R_1$ upper triangular. Using the fact that f is non-decreasing with $f(n) \geq n$ it is straightforward to prove that all R_i's used to compute R have the same sparsity/dispersion function f. A simple operation count gives complexity of order

$$\sum_{k=1}^{n}\sum_{j=k+1}^{f(k)}(f(j)-j) \lesssim \sum_{k=1}^{n}\sum_{j=k+1}^{f(k)} j^{\alpha} \lesssim \sum_{k=1}^{n}(f(k)^{\alpha+1} - k^{\alpha+1}) \lesssim \sum_{k=1}^{n} k^{(\alpha+1)\alpha} \lesssim n^{(\alpha+1)\alpha+1}$$

and we get the result. □

Remark 3.5.2. *If we are studying a finite range Hamiltonian on the lattice $l^2(\mathbb{Z}^d)$ then one can choose $\alpha = (d-1)/d$ and in the general case of such Hamiltonians this is easily seen to be optimal. If $m_1 = Ln, m_2 = n$ then in two dimensions for a constant L this reduces to $n^{2.75}\log(n)$ which is the slope in Figure 3.1.*

Examples of f used in the numerics

We end with some examples for the graph case $l^2(V(\mathcal{G}))$. Suppose our enumeration $\{e_1, e_2, ...\}$ of the vertices obeys the following pattern. All of e_1's neighbours (including itself) are $S_1 = \{e_1, e_2, ..., e_{q_1}\}$ for some finite q_1. The set of neighbours of these vertices is $S_2 = \{e_1, ..., e_{q_2}\}$ for some finite q_2, where we continue the enumeration of S_1 and this process continues inductively enumerating S_m.

Example 3.5.3. Suppose that the bounded operator A can be written as

$$A = \sum_{v \sim_k w} \alpha(v, w) \left|v\right\rangle \left\langle w\right| \tag{3.5.1}$$

for some $k \in \mathbb{N}$ (we write $v \sim_k w$ for two vertices $v, w \in V$ if there is a path of at most k edges connecting v and w, that is, A only involves k-th nearest neighbour interactions). Suppose also that the vertex degree of $\mathcal{G}$ is bounded by M. It holds that $e_n \in S_n$ and $\{w \in V : v \sim_k w\} \subset S_{n+k}$. Inductively $|S_m| \leq (M+1)^m$ and hence we may take the upper bound

$$S(n) = (M+1)^{n+k}.$$

Example 3.5.4. Consider a nearest neighbour operator ($k = 1$ in (3.5.1)) on $l^2(\mathbb{Z}^d)$. It holds that $|S_m| \sim \mathcal{O}(m^d)$ whilst $|S_{m+1} - S_m| \sim \mathcal{O}(m^{d-1})$. It is easy to see that we can choose a suitable S such that

$$S(n) - n \sim \mathcal{O}(n^{\frac{d-1}{d}}),$$

that is, S grows at most linearly.

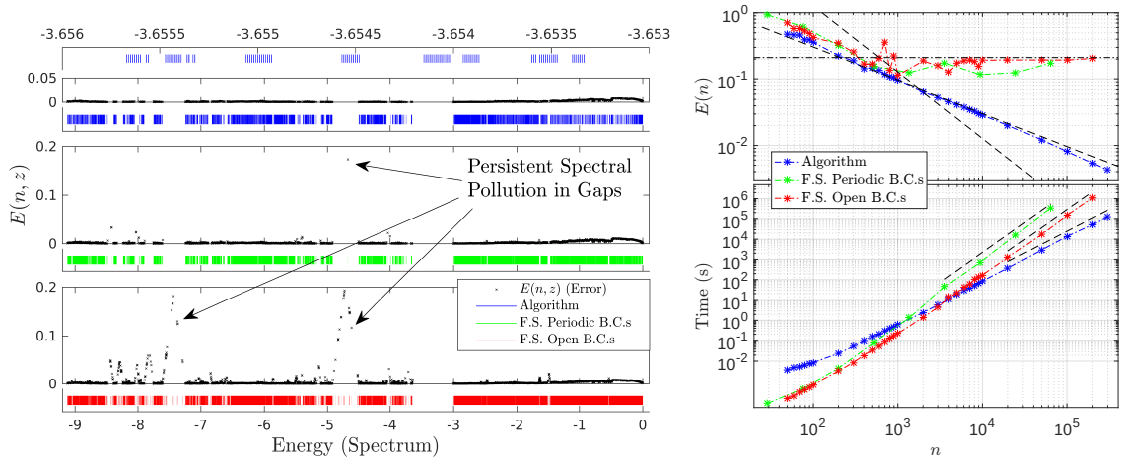


Figure 3.2: Left: Large scale experiment with $n = 10^5$ for the algorithm of this chapter and finite section with open boundary conditions and periodic approximants (see descriptions in main text), applied to the operator H_0 in (3.6.1). The top row shows a magnified section of the approximation provided 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).

3.6 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.

3.6.1 Quasicrystals

Quasicrystals,[3] and more generally aperiodic systems, have generated considerable interest due to their often exotic physical/spectral properties [SBGC84, Sta12]. We present the first rigorous spectral computational study with error bounds on a Penrose tile, the standard 2D model of a quasicrystal [VNA13, TTK15, DVET$^+$05]. No previous algorithm converges to the spectrum, nor provides error bounds on the output.

The free Hamiltonian H_0 (Laplacian) is given by

$$(H_0\psi)_i = \sum_{i\sim j} (\psi_j - \psi_i)\,, \tag{3.6.1}$$

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].

[3] Discovered in 1982 by D. Shechtman who was awarded the Nobel prize in 2011 for his discovery.

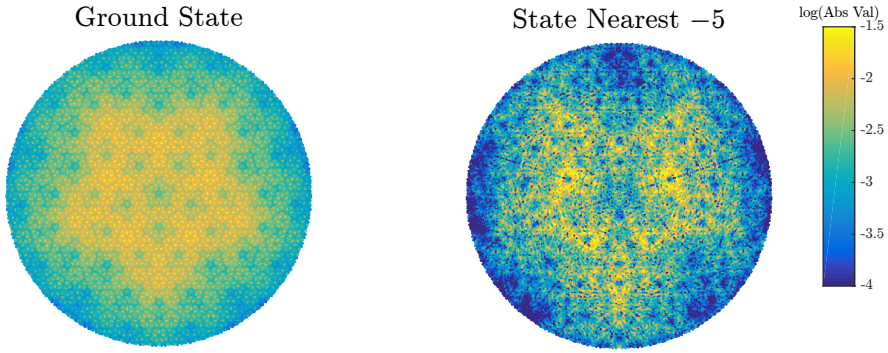


Figure 3.3: The ground 'state' for the Penrose Laplacian (from the cover of Physical Review Letters Volume 122, Issue 25 [CRH19]) and an approximate state corresponding to energy nearest -5 (demonstrating that we can deal with parts of the spectrum that are not at the edge of the spectrum). The algorithm 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.

Figure 3.2 (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 the proof of Theorem 3.2.7, 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 (run using 200 cores) and for the finite section methods (which are not parallelisable in general) to reach the final output (shown in Figure 3.2) 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. The direct diagonalisation approach is hard to parallelise[4] and so will have difficulty competing with the speed of our method for large n. It is also possible to use the methods of this chapter to locally compute approximate states corresponding to a given energy level without the need to diagonalise the whole system, as shown in Figure 3.3 and proven in §3.4.

Finally, we consider a magnetic Hamiltonian [TDGG15, HK87, VM04, Hof76]

$$(H\psi)_i = -\sum_{\langle i,j\rangle} e^{\mathrm{i}\alpha_{ji}}\psi_j.$$

A constant perpendicular magnetic field $\mathbf{B} = B\mathbf{z}$ with potential $\mathbf{A} = (0, xB, 0)$ is applied, leading to the Peierls phase factor between sites i and j: $\alpha_{ji} = \frac{2\pi}{\Phi_0}\int_{\mathbf{r}_j}^{\mathbf{r}_i} \mathbf{A}\cdot d\mathbf{l}$, where $\Phi_0 = hc/e$ is the flux quantum. Figure 3.4 shows the output for the finite section method and the algorithm of this chapter for $n = 5000$ up to the first self-similar mode B_0. The absence of spectral pollution when using our algorithm is striking.

Recently, Hofstadter's butterfly has been experimentally observed in graphene lattices [HSYY$^+$13, DWM$^+$13, PGY$^+$13]. Clearly, numerical methods that avoid spectral pollution, converge, and provide

[4] See, for example, [Cup80, TD99, NH13] for parallel computations of eigenvalues of finite square matrices.

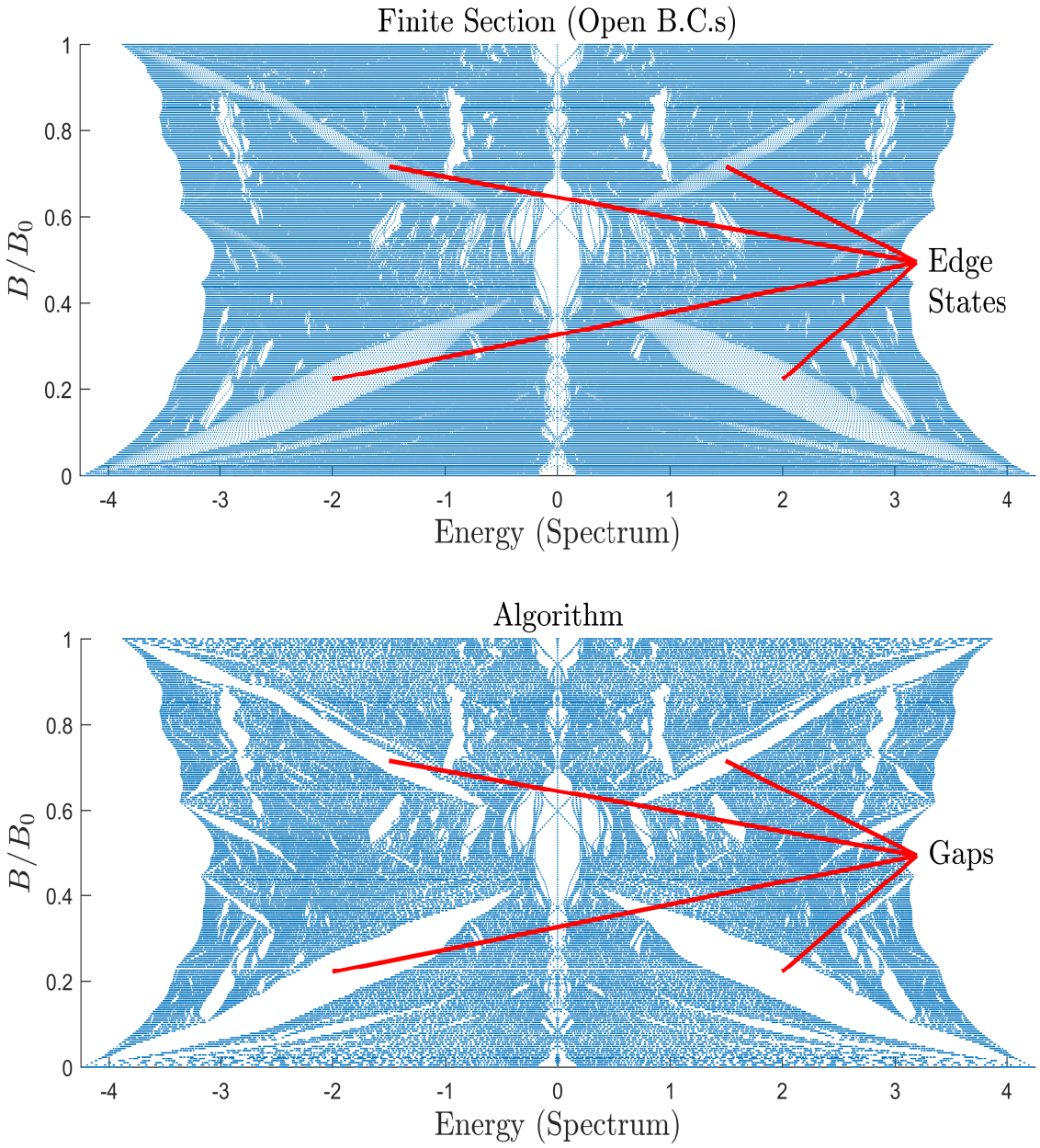


Figure 3.4: Comparison of the finite section method and the algorithm of this chapter for the magnetic Hamiltonian. The new algorithm correctly leaves out the gaps and is able to capture the complicated structure with guaranteed error maximum 0.058 for $n = 5000$, whereas the finite section method produces incorrect eigenvalues (edge states).

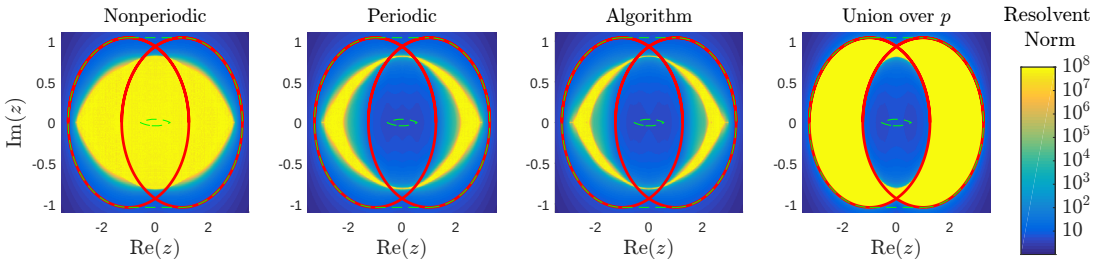


Figure 3.5: Pseudospectra of the finite section method with non-periodic boundary conditions shown as contours of the resolvent norm $\|(H_n - zI)^{-1}\|$ for $n = 10^6$. Similar plots for periodic boundary conditions, the new algorithm with and without varying p. Bounds on the spectrum are shown in green and the set $E + M$ in red.

error bounds are needed to study such operators with fractal-like spectra. The new method can also be applied to problems with arbitrary (even non-constant) magnetic fields and models with large degrees of freedom. Numerical difficulties have previously prevented theoretical modelling of many experimental results of quasicrystals in higher dimensions. The new algorithm can tackle such models, and future work will study 3D systems.

3.6.2 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.[5] 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.

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, and this is discussed further in Chapter 10. 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

[5] 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).

hull of $E+[-1,1]$. Figure 3.5 (I am indebted to Bogdan Roman who assisted me with parallelisation across multiple machines for this example) 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 Chapter 10.

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.

3.6.3 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 $\hat{p} \to -\hat{p}, i \to -i$. Further distinction can be made between exact (unbroken) PT-symmetry when H shares common 'eigenfunctions' with PT and broken PT-symmetry when they possess different eigenfunctions. 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. Such systems are of wide interest [GEB$^+$15, WRM$^+$15, EHW$^+$13, Sch13, Lon09] and can be realised in optics [MEGCM08, GSD$^+$09, RMEG$^+$10, SLZ$^+$11, RBM$^+$12, RMB$^+$13, FWM$^+$14, HMH$^+$14, ZHI$^+$15].

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 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 3.6 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.

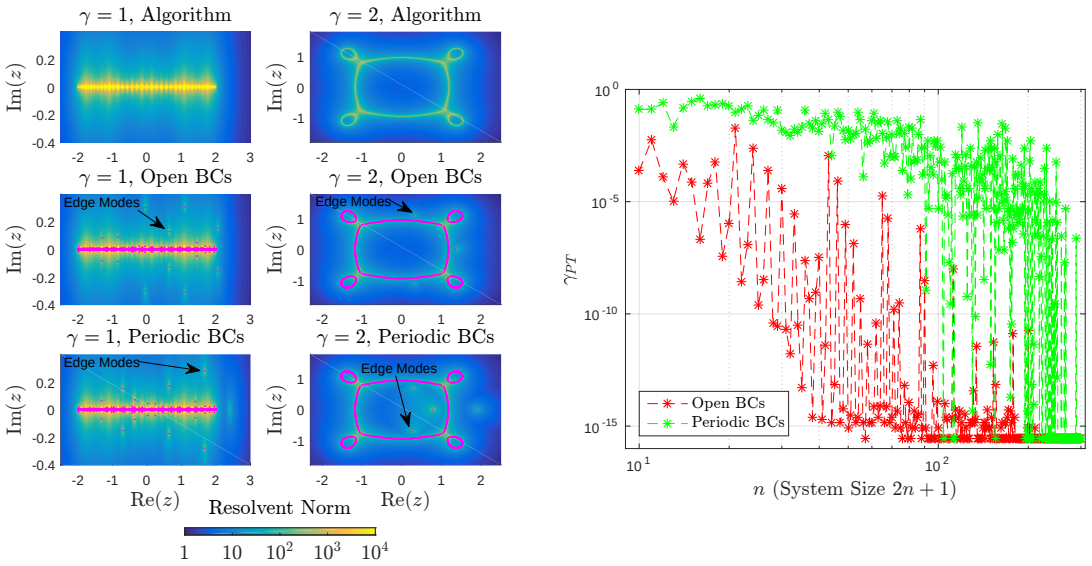


Figure 3.6: Left: Pseudospectra of H computed with the new algorithm and finite sections with different BCs (in magenta). We can easily detect edge modes with the new algorithm, whereas the finite section approach produces incorrect solutions (edge modes). In the periodic case we have no edge, and rather 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.

3.6.4 Partial differential operators

We demonstrate the algorithms of this chapter on PDOs on $L^2(\mathbb{R}^d)$. For many of the examples, we consider operators of the form

$$T = P(x_1, ..., x_d, \partial_1, ..., \partial_d),$$

with domain $\mathcal{D}(T) \subset L^2(\mathbb{R}^d)$, where P is a polynomial. In this case, the matrix representation in the Hermite basis is sparse. From the comments in Example 3.5.4 and recurrence relations for Hermite functions, we can choose a basis such that $f(n) - n \sim Cn^{(d-1)/d}$, where f is the dispersion function and $C = C(d)$ a constant, such that f also describes the off-diagonal sparsity structure of A in the sense that $A_{n,k} = A_{k,n} = 0$ if $k > f(n)$. 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 §3.6.4.

Anharmonic oscillators

First, consider operators of the form

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

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 fact that such operators are essentially self-adjoint follow from the Faris–Lavine theorem [RS75, Theorem X.28] (one can also prove that compactly supported smooth functions form a core). Anharmonic oscillators have attracted interest in quantum research for over four decades [BO13, Wen96, BW73, FMT89]

and amongst their uses are approximations of potentials near stationary points. The problem of developing efficient algorithms to compute their spectra has received renewed interest due to advances in asymptotic analysis and symbolic computing algebra [GSS15, Bar05, Tur10]. The methods in the cited works are rich and diverse, but lack uniformity.

We begin with comparisons to some known results in one dimension, calculated using super-symmetric quantum mechanics [CM91]:

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

These examples have discrete spectra and, following the physicists' convention, we list the energy levels as $E_0 \leq E_1 \leq E_2 \leq ....$ Note that other methods such as finite section (of the corresponding matrices A constructed using Hermite functions) will converge in this case (due to there being no gaps in the essential spectrum), but do not provide the sharp Σ_1 classification. 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. Note that we quickly gain convergence and that the error bounds become the precision of the search routine in `DistSpec` (namely, $10^5 n$). In this simple example, the output happens to agree precisely with the eigenvalues since they lie on the search grid.

Remark 3.6.1. *In the notation of Proposition 3.5.1 the above examples have $L(n)m_1(n)$ constant (we found the local search intervals via previous estimations with $\Gamma_k(A)$) and $m_2(n) = n$. Hence we expect the time taken to scale linearly up to* log *factors. This was found to be the case.*

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 (through `DistSpec`) 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 3.7 shows the convergence of the estimate of $\|R(z, H_1)\|^{-1}$ from above as well as finite section estimates. As expected

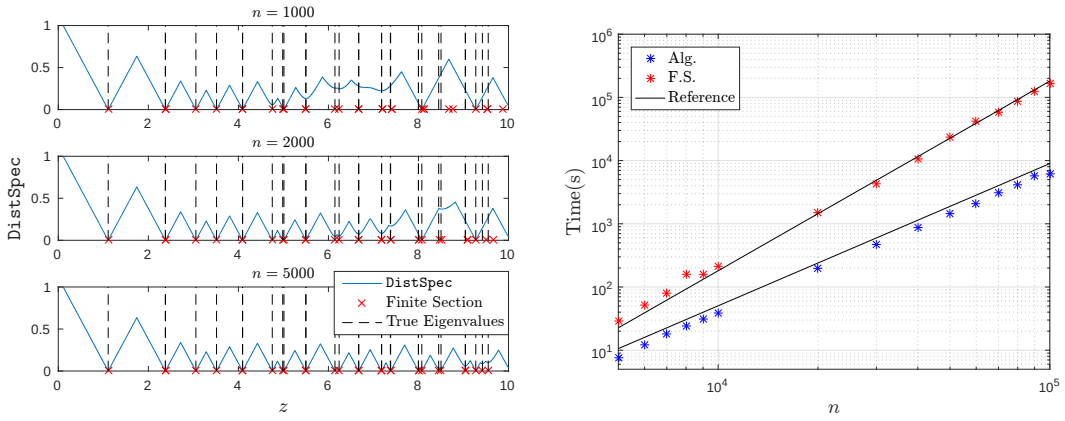


Figure 3.7: Two-dimensional example. Left: The convergence of our algorithm (shown as `DistSpec`) and finite section to the true eigenvalues on the interval $[0, 10]$. Note that points with reliable finite section eigenvalues correspond to points where the estimate of the resolvent norm is well-resolved. 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).

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 3.7, where we found that the finite section method grew roughly cubically whereas our algorithm grew roughly as $\mathcal{O}(n^{2.25})$ (both shown as reference lines). The speed-up for our algorithm, compared with $\mathcal{O}(n^{2.75})$, was due to the AMD 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 3.8 (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 3.8 (right) shows the output of our algorithm. This avoids spectral pollution whilst converging to the true spectrum.

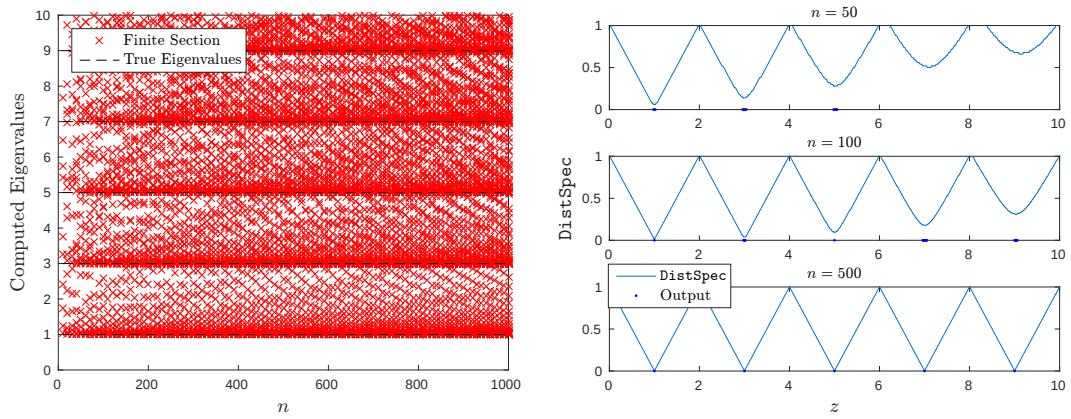


Figure 3.8: Left: Finite section for various n. Note the extremely heavy spectral pollution, although eigenvalues do appear to cluster around the true spectrum. Right: The estimates provided by `DistSpec`. The estimate converges quickly to the true value from above. The output of our algorithm can be 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 $2n+1$

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 §4.6.3, where we compute highly oscillatory bounded modes.

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 [CH19a]). 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. Each entry in the table is computed with an error bound at most 10^{-9} provided by `DistSpec`.

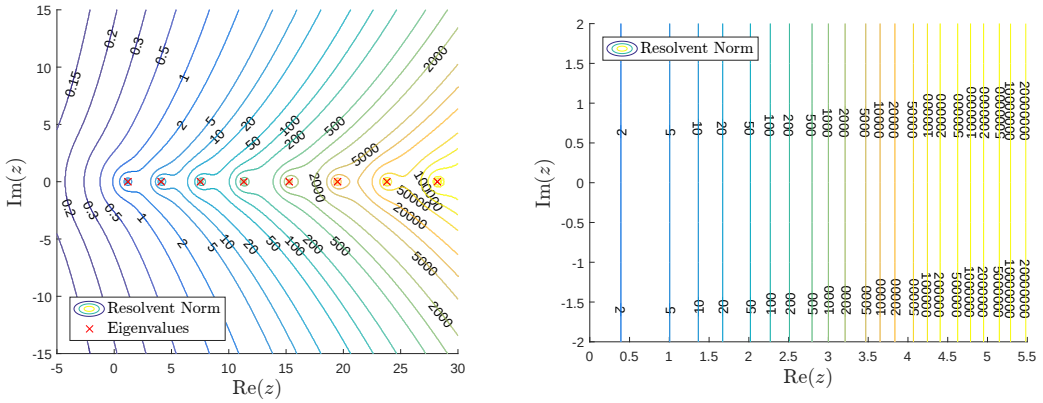


Figure 3.9: Left: Calculated pseudospectrum for the imaginary cubic oscillator. Note the clear presence of eigenvalues. Right: Calculated pseudospectrum for imaginary Airy operator. Both figures were produced with $n = 1000$.

Pseudospectra and PT-symmetry

We now turn to the pseudospectrum and consider PT-symmetric non-self-adjoint operators T (and for which it is known that compactly supported smooth functions form a core of T and T^* [EE87]). 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 §3.6.3) [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]. The eigenvectors are complete but do not form a Riesz basis [SK12]. Figure 3.9 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 $\mathrm{Re}(z) \to +\infty$. We have shown the computed pseudospectrum for $n = 1000$ in Figure 3.9.

We do not need to discretise anything to apply the above method. Up to numerical errors in the testing of positive definiteness (which can be implemented using interval arithmetic if desired [Tuc11]), all computed pseudospectra are guaranteed to be inside the correct pseudospectra. This is in contrast to the numerical experiments conducted in [Dav99], where the operator is discretised. It is also easy to construct examples where discretisations fail dramatically, either not capturing the whole spectrum or suffering from spectral pollution (even without spectral pollution - figuring out which parts of computations are trustworthy can be very difficult for finite section and related methods [Zha15]). Algorithms like `PseudoSpec` are a useful tool to test the reliability of such outputs.

Chapter 4

Computing Spectral Measures

Part of the richness and beauty that arises in infinite dimensions is the possibility of different spectral types. Given a normal operator A, there is an associated projection-valued measure (resolution of the identity), which we denote by 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 much more complicated with different types of spectra (see Chapter 5). The computation of E^A, along with its various decompositions and their supports, is of great interest, both theoretically and for practical applications. For example, spectral measures are intimately 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 one to compute many additional objects, such as the functional calculus, the Radon–Nikodym derivative of the absolutely continuous component of the measure, and spectral measures and spectral set decompositions. For instance, in §4.1.3 we discuss how the results of this chapter allow the computation of spectral measures and the functional calculus of almost arbitrary self-adjoint partial differential operators on $L^2(\mathbb{R}^d)$. An important class of examples is given by solutions of evolution equations such as the time-dependent Schrödinger equation on $L^2(\mathbb{R}^d)$ [Lub08, HO10]. An excellent and readable introduction to the spectral theorem can be found in Paul Halmos' article [Hal63].

Despite its importance, there has been no general method for computing spectral measures of normal operators, or even self-adjoint operators. Although there is a rich literature on the theory of spectral measures, most of the efforts to develop computational tools have focused on specific examples where analytical formulas are available or perturbations thereof. For example, the work of [WO17] deals with compact perturbations of tridiagonal Toeplitz operators and there are methods for computing spectral density functions of Sturm–Liouville problems.[1] In some sense, the lack of general methods is not surprising given the dif-

[1] For further discussions on applications and previous methods in the literature, we refer the reader to the discussions in [CHT20].

ficulty of rigorously computing spectra. One can consider the open problem of general computation of spectral measures as the infinite-dimensional analogue of computing projections onto eigenspaces.[2]

In this chapter, we provide algorithms for the computation of spectral measures for a large class of self-adjoint operators (and, more generally, normal operators whose spectrum lies on a regular enough Jordan curve). We classify the computation of measures, measure decompositions, functional calculus and Radon–Nikodym derivatives in the SCI hierarchy for such operators. Given a matrix representation, we show that if each matrix column decays at infinity at a known asymptotic rate, then it is possible to compute spectral measures. The central ingredient of the new algorithm is the computation of the resolvent operator with error control. We also discuss how to improve the convergence rates of the new algorithms for smooth enough measures (locally) by using different rational kernels. Under regularity assumptions, this allows the computation of spectral measures (specifically local computation of the Radon–Nikodym derivative of the absolutely continuous part) with error control.

We also demonstrate the applicability of the new algorithms. These algorithms are parallelisable, allowing large scale computations. Examples include orthogonal polynomials on the real line (obtaining the measure from the recurrence relations), a model of magneto-graphene that demonstrates high-resolution computation and the avoidance of spectral pollution, fractional diffusion on a quasicrystal and the solution of infinite-dimensional evolution equations with error control. Partial differential operators on the continuum are also studied, and the results of this chapter carry over by employing spectral methods to solve the PDEs corresponding to the resolvent. As an example, we study a Dirac operator with radially symmetric potential, which models relativistic quantum electrons in an external field. This example corresponds to a coupled first-order system on the half-line, and we show how the resolvent gives rise to a very efficient numerical method to compute highly oscillatory bound states, whilst avoiding spectral pollution. In all cases, the challenging numerical aspect is the computation of the resolvent close to the real-axis, and this is one reason why the high-order methods developed in §4.5 are particularly useful. 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 also collectively highlight an important point, that is, the new algorithms can easily be used in tandem with any numerical procedure that computes the action of the resolvent in an adaptive manner with asymptotic error control. This gives great flexibility to the methods.[3]

4.1 Background and Summary

We begin with the relevant background for the rest of Part I of this thesis. As in Chapter 3, 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.[4] In particular, we can

[2] Of course eigenvectors exist in the infinite-dimensional case, but not all of the spectrum consists of eigenvalues. The projection-valued measure generalises the notion of projections onto eigenspaces.

[3] This flexibility is explored further in the paper [CHT20], and the theory and methods developed in this chapter form much of the foundations of the software package `SpecSolve` written by the author and Andrew Horning:

https://github.com/SpecSolve/SpecSolve

This chapter is based mainly on [CRH19], the theory of high-order rational kernels in [CHT20] (§4.5 of this chapter) was developed by the author, and the spectral method implementation of high-order kernels in [CHT20] (in particular the numerical examples of §4.5.1 and §4.6.3) was developed jointly by the author and Andrew Horning.

[4] It should be pointed out, however, that not all bases are created equal. For example, for a given operator it may be the case that there is a basis which gives a diagonal representation, whereas another basis may generate a dense matrix. We will see that a good basis is one for which we know the function f appearing in (4.1.3). Moreover, for the methods of this chapter, one only needs to be able to compute the resolvent to sufficient accuracy. One does not need a basis representation, and hence a wide number of tools from scientific computation become available when solving the shifted linear systems.

handle partial differential operators through spectral methods as discussed in §4.1.3. Indeed, 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 $\mathrm{Sp}_{\mathrm{p}}(A)$, which is not always closed. 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 subclasses $\subset \Omega_{\mathrm{N}}$ of self-adjoint (again allowing unbounded operators) and unitary operators will be denoted by Ω_{SA} and Ω_{U} respectively. Recall that for $A \in \Omega_{\mathrm{SA}}$ and $A \in \Omega_{\mathrm{U}}$, $\mathrm{Sp}(A) \subset \mathbb{R}$ and $\mathrm{Sp}(A) \subset \mathbb{T}$ respectively, where $\mathbb{T}$ denotes the unit circle.

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

$$\mu^A_{x,y}(B) = \langle E^A_B x, y\rangle.$$

Via the Lebesgue decomposition theorem [Hal50], the spectral measure $\mu^A_{x,y}$ can be decomposed into three parts

$$\mu^A_{x,y} = \mu^A_{x,y,\mathrm{ac}} + \mu^A_{x,y,\mathrm{sc}} + \mu^A_{x,y,\mathrm{pp}},$$

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

$$\rho^A_{x,y}(\lambda) = \frac{d\mu^A_{x,y,\mathrm{ac}}}{dm}(\lambda), \tag{4.1.1}$$

the Radon–Nikodym derivative of $\mu^A_{x,y,\mathrm{ac}}$ 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 $\mu^A_{x,x}$ 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^A_{ac}, P^A_{sc} and P^A_{pp} 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

We now define the various subclasses of operators and the evaluation set used in this chapter and the next. Given an operator $A \in \mathcal{C}(l^2(\mathbb{N}))$, we can view it as an infinite matrix

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

through the inner products $a_{ij} = \langle Ae_j, e_i \rangle$. All of the algorithms constructed can also be adapted to operators on $l^2(\mathbb{Z})$, either through the use of a suitable reordering of the basis, or though considering truncations of matrices in two directions, which is useful numerically since it preserves bandwidth. To be precise about the information needed to compute spectral properties, we define the class of evaluation functions $\Lambda_1 = \{\langle Ae_j, e_i \rangle : i, j \in \mathbb{N}\}$. This is entirely analogous to §3.1.1. 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. An example of this was given in Chapter 3 and we discuss in §4.1.3 how to extend the results of this chapter to partial differential operators (see also an appendix of [Col19a]).

Remark 4.1.1. *As usual, all of the algorithms that follow can be easily extended to the case of inexact input and restrictions to arithmetic operations with rational numbers.*

We will be concerned with operators whose matrix representation has a known asymptotic rate of column/off-diagonal decay. Namely, let $f : \mathbb{N} \to \mathbb{N}$ with $f(n) > n$ and let $\alpha = \{\alpha_n\}_{n\in\mathbb{N}}, \beta = \{\beta_n\}_{n\in\mathbb{N}}$ be null sequences[5] of non-negative real numbers. We then define for $\mathrm{X} = \mathrm{SA}$ or $\mathrm{X} = \mathrm{U}$,

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

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

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

When discussing $\Omega^{\mathrm{SA}}_{f,\alpha,\beta}$ and $\Omega^{\mathrm{SA}}_{f,\alpha}$ we will use the notation $\Omega_{f,\alpha,\beta}$ and $\Omega_{f,\alpha}$. The collection of vectors in $l^2(\mathbb{N})$ satisfying $\|P_n x - x\| = \mathcal{O}(\beta_n)$ 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. Other additions are sometimes needed such as data regarding open sets as inputs for computations of measures, but this will always be made clear. When considering the general case of $\Omega_{f,\alpha}$, the function f and sequence α can also be considered as inputs to the algorithm - in other words, the same algorithm works for each class.

4.1.2 A motivating example

As a motivating example, consider the case of a Jacobi operator with matrix

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

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

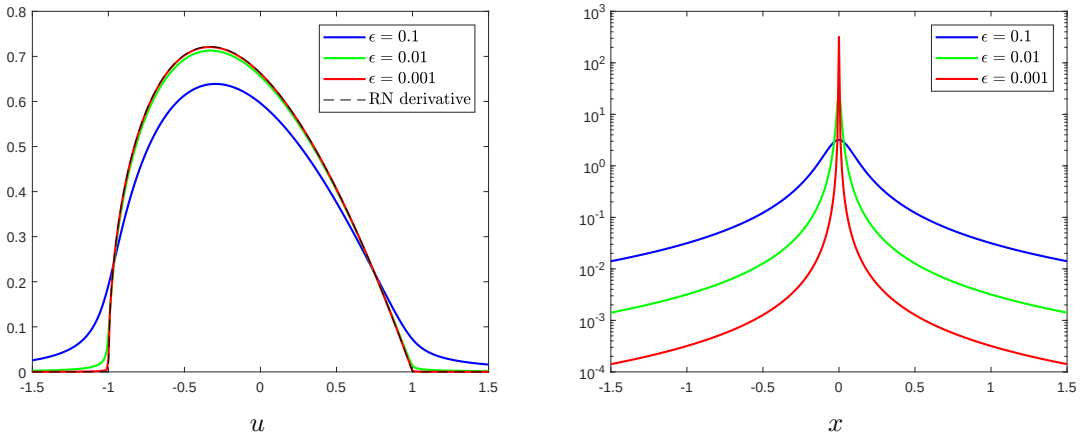


Figure 4.1: Smoothed approximations of the Radon–Nikodym derivative for the Jacobi operator associated to Jacobi polynomials with $\alpha = 1$, $\beta = 1/2$. Here the measure is absolutely continuous and supported on $[-1, 1]$. Left: Convolutions $K_H(u + i\epsilon; J, e_1)$ for different ϵ using the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

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, GS97] and one can develop the discrete analogue of what is known as Weyl–Titchmarsh–Kodaira theory for Sturm–Liouville operators. Going back to the work of Stieltjes [Sti94] (see also [Akh65, Wal48]), there is a representation of G as a continued fraction:

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

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

However, there are two significant obstacles to overcome when using (4.1.4) and its variants as a means to compute measures. First of all, 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. As a by-product, we compute the m-function of such operators with error control. Specific cases where this can be written explicitly do exist, such as periodic Jacobi matrices or perturbations of Toeplitz

operators [DE15]. However, there has been no general method proposed to compute the resolvent with error control. This consideration is crucial to allow the computation of measures in one limit.

To see how we might compute the measure using the resolvent, consider the Poisson kernel for the half-plane and the unit disk, defined respectively by

$$P_H(x,y) = \frac{1}{\pi}\frac{y}{x^2+y^2} \text{ and } P_D(x,y) = \frac{1}{2\pi}\frac{1-(x^2+y^2)}{(x-1)^2+y^2} = P_D(r,\theta) = \frac{1}{2\pi}\frac{1-r^2}{1-2r\cos(\theta)+r^2},$$

where (x,y) and (r,θ) denote the usual Cartesian and polar coordinates respectively. 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

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

Similarly, if A is unitary, $z = r\exp(i\psi) \in \mathbb{C}\backslash\mathbb{T}$ (with $z \neq 0$) and $x \in l^2(\mathbb{N})$ we define

$$K_D(z;A,x) := \frac{1}{2\pi i}[R(z,A) - R(1/\overline{z},A)]x = \frac{1}{2\pi i}\int_{\mathbb{T}}\left[\frac{1}{\lambda - z} - \frac{1}{\lambda - 1/\overline{z}}\right] dE^A(\lambda)x.$$

We change variables $\lambda = \exp(i\theta)$ and with an abuse of notation, write $dE^A(\lambda) = i\exp(i\theta)dE^A(\theta)$. A simple calculation then gives

$$K_D(z;A,x) = \int_0^{2\pi} P_D(r, \psi - \theta) dE^A(\theta)x. \tag{4.1.5}$$

Returning to our example, 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 and also the computation of measures with error control under certain regularity assumptions.

Finally, we remark on a similar, though different, object studied in the mathematical physics literature: the density of states [Kir07, CL90, KM07], which we mention for completeness and to avoid potential confusion. This is related to the finite section method and often used when considering random Schrödinger operators. This object is defined via the 'thermodynamic limit', where instead of considering the full infinite-dimensional operator A, one considers finite truncations, say P_nAP_n, and the limit as $n \to \infty$ (in the weak* sense) of the measure $\sum_{x_j \in \mathrm{Sp}(P_nAP_n)} \delta_{x_j}/n$. To see why the density of states is different from the spectral measure of A (and why the averaging leads to a less refined measure than the full spectral measure), consider A with discrete spectra below the essential spectrum. The contribution of these eigenvalues to the density of states vanishes as $n \to \infty$. The spectral measure, on the other hand, does not ignore these eigenvalues and allows the computation of spectral decompositions, as we demonstrate in this thesis.

4.1.3 Summary of results of chapter and extensions to partial differential operators

We prove that with respect to Λ_1 (and with convergence in the corresponding metric spaces):

- **Proposition 4.2.1 and Corollary 4.2.2:** Computation of $R(z, A)x \in l^2(\mathbb{N})$ for $(A, x) \in \Omega_{f,\alpha,\beta}$ lies in Δ_1^A. In other words, the resolvent can be computed in one limit using an arithmetic algorithm with error control.

- **Theorem 4.3.1:** Computation of $E_U^A x \in l^2(\mathbb{N})$ for $(A, x) \in \Omega_{f,\alpha,\beta}$ and U open lies in Δ_2^A but not Δ_1^G. In other words, this can be done in one limit using an arithmetic algorithm but, in general, error control is impossible. This can be extended to other types of sets such as closed intervals or singletons, and can be extended to unitary and much more general operators (see Theorem 4.2.4). Through taking inner products, these results extend to the computation of the scalar measures $\mu_{x,y}^A(U)$.

- **Theorem 4.3.3:** Computation of the decompositions $\mu_{x,y,\mathcal{I}}^A(U)$ for $(A, x, y) \in \Omega_{f,\alpha,\beta} \times V_\beta$ and U open lies in Δ_3^A but not Δ_2^G for $\mathcal{I} = \mathrm{ac}, \mathrm{sc}$ and pp. In other words, this can be done in two limits using arithmetical algorithms but not one limit.

- **Theorem 4.4.1:** Computation of $F(A)x \in l^2(\mathbb{N})$ for $(A, x) \in \Omega_{f,\alpha,\beta}$ and F a continuous bounded function on $\mathrm{Sp}(A)$ lies in Δ_2^A. Error control is sometimes possible. For example, if $\mathrm{Sp}(A)$ is bounded and F is holomorphic on an open neighbourhood of $\mathrm{Sp}(A)$.

- **Theorem 4.4.2:** Computation of $\rho_{x,y}^A \in L^1(U)$ for $(A, x) \in \Omega_{f,\alpha,\beta} \times l^2(\mathbb{N})$ (where U is an open set separated from the singular parts of the measure $\mu_{x,y}^A$) lies in Δ_2^A (i.e. we can compute Radon–Nikodym derivatives in one limit). Without the separation condition, our algorithm converges (Lebesgue) almost everywhere.

Higher-order kernels

In §4.5, we consider the use of higher-order kernels replacing the Poisson kernel in the computation of spectral quantities. Under local regularity assumptions on the measure, we prove, in **Theorem 4.5.3** and **Theorem 4.5.7** respectively, pointwise and L^p convergence rates of arbitrary order in the smoothing parameter ϵ. In **Corollary 4.5.9**, this is translated to Δ_1^A classifications for computing the Radon–Nikodym derivative of the absolutely continuous part of the spectral measures in L^p spaces for $1 \leq p \leq \infty$. A similar conclusion holds for the functional calculus. The high-order kernels constructed in this chapter form the basis of the software package `SpecSolve` for computing spectral measures of self-adjoint operators.

Partial differential operators

The techniques of this chapter can be extended to partial self-adjoint partial differential operators on $L^2(\mathbb{R}^d)$ of the form

$$Lu(x) = \sum_{k \in \mathbb{Z}_{\geq 0}^d, |k| \leq N} a_k(x) \partial^k u(x), \tag{4.1.6}$$

with polynomially bounded coefficients of locally bounded total variation. This is done explicitly in an appendix of [Col19a], using some of the techniques of Chapter 3 to compute the resolvent with asymptotic error control. For the sake of brevity, we have not repeated these results here. As an important example, consider Schrödinger operators $L = -\Delta + V$ with polynomially bounded potentials of locally bounded total variation. Hence, in this case, we can compute the spectral properties (measures, functional calculus etc.) of L by point sampling the potential V, if we have an asymptotic bound on the total variation of V

over finite rectangles. In particular, we can solve the Schrödinger equation

$$\frac{du}{dt} = -iLu, \quad u_{t=0} = u_0 \tag{4.1.7}$$

on $L^2(\mathbb{R}^d)$ by computing $\exp(-itL)u_0$ with guaranteed convergence. These kinds of results can be extended to other domains such as the half-line, and can be adapted to cope with other types of potentials or coefficients that are not of locally bounded total variation (for instance Coulombic potentials for Dirac or Schrödinger operators). We also remark that the software package `SpecSolve` makes use of state-of-the-art spectral methods for unbounded domains.

4.2 Approximating the Resolvent

The algorithms built in this chapter and the next rely (adaptively) 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. Given this, one can then compute the projections E_S^A for a wide range of sets S (Theorem 4.3.1 and its generalisations), and hence the measures $\mu_{x,y}^A$. In this section, we discuss the computation of the resolvent with error control and how this can be used to compute measures via generalisations of Stone's formula. This will also form the basis of high-order rational kernels in §4.5.

4.2.1 Approximating the resolvent operator

The key proposition for computing the action of the resolvent operator is the following, where we use σ_1 to denote the injection modulus of an operator:

$$\sigma_1(A) := \min\{\|Ax\| : x \in \mathcal{D}(A), \|x\| = 1\}.$$

The proof boils down to a careful computation of a least-squares solution of a rectangular linear system. Similar results and other SCI classifications of computing inverses of linear systems are developed by the author in collaboration in [BACH$^+$19].

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 $n = \operatorname{rank}(P_n) = \operatorname{rank}((A - zI)P_n) = \operatorname{rank}(P_{f(n)}(A - zI)P_n)$ for large n since $\sigma_1(A - zI) > 0$ and $\|(I - P_{f(n)})(A - zI)P_n\| \leq C_1\alpha_n \to 0$. Hence we can define

$$\widetilde{\Gamma}_n(A, x, z) := \begin{cases} 0, & \text{if } \sigma_1(P_n(A^* - \overline{z}I)P_{f(n)}(A - zI)|_{P_n(l^2(\mathbb{N}))}) \leq \frac{1}{n} \\ [P_n(A^* - \overline{z}I)P_{f(n)}(A - zI)P_n]^{-1}P_n(A^* - \mathrm{z}I)P_{f(n)}x, & \text{otherwise.} \end{cases}$$

Suppose that n is large enough so that $\sigma_1(P_n(A^* - \overline{z}I)P_{f(n)}(A - zI)|_{P_n(l^2(\mathbb{N}))}) > 1/n$. Then $\widetilde{\Gamma}_n(A, x, z)$ is a (least-squares) solution of the optimisation problem $\operatorname{argmin}_y \|P_{f(n)}(A - zI)P_n y - x\|$. The linear space $\operatorname{span}\{e_n : n \in \mathbb{N}\}$ 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

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

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 $(\|\widetilde{C}_n\| + v_n)w_n \leq n^{-1}$. From Corollary 3.2.9 and a simple search routine, we can also compute $\sigma_1(P_n(A^* - \overline{z}I)P_{f(n)}(A -$

$zI)|_{P_n(l^2(\mathbb{N}))})$ 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,

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

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

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

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

since A is normal so that $\|R(z,A)\| = \operatorname{dist}(z,\operatorname{Sp}(A))^{-1}$. 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]. The following corollary of Proposition 4.2.1 will be used repeatedly in the following proofs.

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 $z \in \mathbb{C}\backslash\mathbb{R}$,*

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

Proof. Let $\Gamma_n(A,x,z) = \widehat{\Gamma}_{m(n,A,x,z)}(A,x,z)$ where $\widehat{\Gamma}_k$ are the algorithms from the statement of Proposition 4.2.1 and $m(n,A,x,z)$ is a subsequence diverging to infinity as $n \to \infty$. Clearly statement (1) holds so we must show how to choose the sequence $m(n,A,x,z)$ such that (2) holds (and hence our algorithms will be adaptive). From (4.2.1), it is enough to show that $m = m(n,A,x,z)$ can be chosen such that

$$\beta_{f(m)} + \alpha_m\|\widehat{\Gamma}_m(A,x,z)\| + \|P_{f(m)}(A-zI)\widehat{\Gamma}_m(A,x,z) - P_{f(m)}x\| \lesssim \alpha_n + \beta_n.$$

The left-hand side can be approximated to arbitrary accuracy using finitely many arithmetic operations and comparisons and hence by repeatedly computing approximations to within $\alpha_n + \beta_n$, we can choose the minimal m such that these approximate bounds are at most $2(\alpha_n + \beta_n)$. □

4.2.2 Stone's formula and Poisson kernels

We next show how the computation of the resolvent with error control allows the computation of the convolution of spectral measures with Poisson kernels, as mentioned in §4.1.2. Moreover, this can be done with a certain sense of error control. This is related to Stone's famous formula [Sto90, CL55, RS80] to compute the pointwise action of the projection-valued measures associated with an operator $A \in \Omega_{\mathrm{SA}}$. However, Stone's formula can be generalised to unitary operators and a much larger class of normal operators (see Proposition 4.2.4). We will assume the reader is familiar with standard results from spectral theory and harmonic analysis, which, for example, can be found in [Dur70, RS80]. The following proposition is the celebrated Stone's formula, and we include a short proof for the benefit of the reader since the ideas in the proof will be used elsewhere.

Proposition 4.2.3 (Stone's formula [Sto90])**.** *Recalling the definitions of K_H and K_D in §4.1.2, the following boundary limits hold.*

(i) Let $A \in \Omega_{\mathrm{SA}}$. Then for any $-\infty \leq a < b \leq \infty$ and $x \in l^2(\mathbb{N})$,

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

(ii) Let $A \in \Omega_{\mathrm{U}}$. Then for any $0 \leq a < b < 2\pi$ and $x \in l^2(\mathbb{N})$,

$$\lim_{\epsilon \downarrow 0} \int_a^b i \exp(i\psi) K_D((1-\epsilon)\exp(i\psi); A, x) d\psi = E^A_{(a,b)_{\mathbb{T}}} x + \frac{1}{2} E^A_{\{\exp(ia), \exp(ib)\}} x,$$

where $(a,b)_{\mathbb{T}}$ denotes the image of (a,b) under the map $\theta \to \exp(i\theta)$.

Proof. To prove (i), we can apply Fubini's theorem to interchange the order of integration and arrive at

$$\int_a^b K_H(u + i\epsilon; A, x) du = \int_{-\infty}^{\infty} \int_a^b P_H(u - \lambda, \epsilon) du \, dE^A(\lambda) x$$

But

$$\int_a^b P_H(u - \lambda, \epsilon) du = \frac{1}{\pi} \left[\tan^{-1}\left(\frac{b - \lambda}{\epsilon} \right) - \tan^{-1}\left(\frac{a - \lambda}{\epsilon} \right) \right]$$

is bounded and converges pointwise as $\epsilon \downarrow 0$ to $\chi_{(a,b)}(\lambda) + \chi_{\{a,b\}}(\lambda)/2$, where χ_S denotes the indicator function of a set S. Part (i) now follows from the dominated convergence theorem.

To prove (ii), we apply Fubini's theorem again, now noting that

$$\int_a^b i \exp(i\psi) P_D((1-\epsilon), \psi - \theta) d\psi = \frac{i \exp(i\theta)}{2\pi} \int_{a-\theta}^{b-\theta} \frac{(2\epsilon - \epsilon^2)\exp(i\psi)}{\epsilon^2 + 2(1-\epsilon)(1 - \cos(\psi))} d\psi. \tag{4.2.2}$$

We can split the interval into small intervals of width $\mathcal{O}(\rho)$ $(0 < \rho < 1)$ around each point where $\cos(\psi) = 1$ and a finite union of intervals on which $1 - \cos(\psi)$ is positive, bounded away from 0. On these later intervals, the limit vanishes as $\epsilon \downarrow 0$. Hence by periodicity and considering odd and even parts, we are left with considering

$$I_1(\rho, \epsilon) = \int_0^{\rho} \frac{(2\epsilon - \epsilon^2)\cos(\psi)}{\epsilon^2 + 2(1-\epsilon)(1 - \cos(\psi))} d\psi, \quad I_2(\rho, \epsilon) = \int_0^{\rho} \frac{(2\epsilon - \epsilon^2)\sin(\psi)}{\epsilon^2 + 2(1-\epsilon)(1 - \cos(\psi))} d\psi.$$

Explicit integration yields $I_2(\epsilon, \rho) = \mathcal{O}(\epsilon \log(\epsilon))$ and hence the contribution vanishes in the limit. We also have

$$I_1(\rho, \epsilon) = \frac{(\epsilon^2 - 2\epsilon)\rho + 2(2 + \epsilon^2 - 2\epsilon)\tan^{-1}\left(\frac{(2-\epsilon)\tan\left(\frac{\rho}{2}\right)}{\epsilon} \right)}{2(1-\epsilon)}.$$

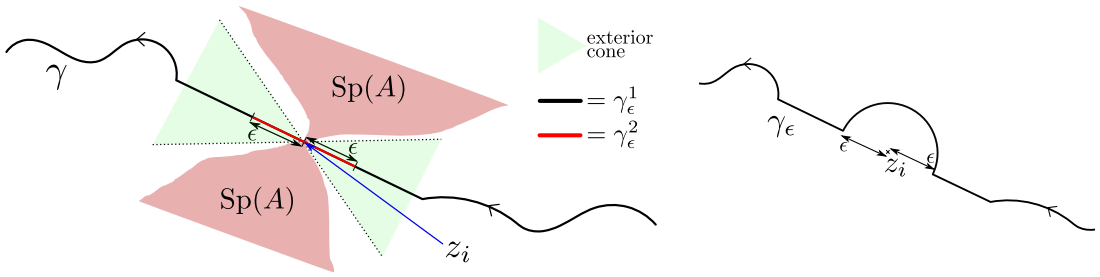


Figure 4.2: Left: Exterior cone condition for Proposition 4.2.4. Right: Deformed contour γ_ϵ to compute the quantity $f_\epsilon(z_i)$.

This converges to π as $\epsilon \downarrow 0$. Considering the contributions of I_1 and I_2 in (4.2.2), we see that (4.2.2) converges pointwise as $\epsilon \downarrow 0$ to

$$i \exp(i\theta)\big\{\chi_{(a,b)}(\theta) + [\chi_{\{a\}}(\theta) + \chi_{\{b\}}(\theta)]/2\big\}.$$

Since the integral is also bounded, part (ii) now follows from the dominated convergence theorem and change of variables. □

This type of construction can be generalised to $A \in \Omega_{\mathrm{N}}$ whose spectrum lies on a regular enough curve. However, it is much more straightforward in the general case to use the analytic properties of the resolvent. The next proposition does this and also holds for operators whose spectrum does not necessarily lie on a curve.

Proposition 4.2.4 (Generalised Stone's formula)**.** *Let γ be a rectifiable positively oriented Jordan curve. Suppose that $A \in \Omega_{\mathrm{N}}$ is such that* Sp(A) *intersects γ at finitely many points $z_1, ..., z_m$. Suppose also that in a neighbourhood of each of the z_i, γ is formed of a line segment meeting* Sp(A) *only at z_i, at which point* Sp(A) *has a local exterior cone condition with respect to γ (see Figure 4.2). Let $x \in l^2(\mathbb{N})$. We can then define the Cauchy principal value integral of the resolvent $R(z, A)x$ along γ and have*

$$\frac{-1}{2\pi i}\mathrm{PV}\int_\gamma R(z,A)x\,dz = E^A_{\mathrm{Sp}(A;\gamma)}x - \frac{1}{2}\left[\sum_{j=1}^m E^A_{\{z_j\}}x\right], \tag{4.2.3}$$

where Sp$(A;\gamma)$ *is the closure of the intersection of* Sp(A) *with the interior of γ.*

Proof. We will argue for the case $m = 1$, and the general case follows in exactly the same manner. Let $\epsilon > 0$ be small so that in a neighbourhood of the $\epsilon-$ball around z_1, γ is given by a straight line. We then decompose γ into two disjoint parts

$$\gamma = \gamma_\epsilon^1 \cup \gamma_\epsilon^2$$

as shown in Figure 4.2. We set

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

We then consider the inner integral

$$f_\epsilon(\lambda) = \int_{\gamma_\epsilon^1} \frac{1}{\lambda - z}dz.$$

If λ is inside γ then $\lim_{\epsilon\downarrow 0} f_\epsilon(\lambda) = -2\pi i$ via Cauchy's residue theorem. Similarly, if λ is outside γ then $\lim_{\epsilon\downarrow 0} f_\epsilon(\lambda) = 0$. To calculate $f_\epsilon(z_1)$, consider the contour integral along γ_ϵ in Figure 4.2. We see that

$$f_\epsilon(z_1) - i\pi = -2i\pi$$

and hence $f_\epsilon(z_1) = -i\pi$. We would like to apply the dominated convergence theorem. Clearly, away from z_1, f_ϵ is bounded as $\epsilon \downarrow 0$. Now let $0 < \delta < \epsilon$ then

$$f_\delta(\lambda) - f_\epsilon(\lambda) = \int_\delta^\epsilon \frac{1}{\frac{\lambda - z_1}{w} - s} + \frac{1}{\frac{\lambda - z_1}{w} + s} ds = \log\left(\frac{\epsilon + \frac{\lambda - z_1}{w}}{-\epsilon + \frac{\lambda - z_1}{w}}\right) - \log\left(\frac{\delta + \frac{\lambda - z_1}{w}}{-\delta + \frac{\lambda - z_1}{w}}\right)$$

for some $w \in \mathbb{T}$. Taking the pointwise limit $\delta \downarrow 0$, we see that we can prove that $f_\epsilon(\lambda)$ is bounded for $\lambda \in \mathrm{Sp}(A)$ in a neighbourhood of z_1 as $\epsilon \downarrow 0$ if we can prove the same for

$$g_\epsilon(\lambda) = \log\left(\frac{\epsilon + \frac{\lambda - z_1}{w}}{-\epsilon + \frac{\lambda - z_1}{w}}\right).$$

By rotating and translating, we can assume that $w = 1$ and $z_1 = 0$ without loss of generality. Let $\lambda_1 = \mathrm{Re}(\lambda)$ and $\lambda_2 = \mathrm{Im}(\lambda)$. Using the cone condition gives $\alpha |\lambda_1| \leq |\lambda_2|$ for some $\alpha > 0$. Assume $\lambda_1 \neq 0$ then

$$\left|\frac{\epsilon + \lambda}{-\epsilon + \lambda}\right|^2 = \frac{(\epsilon + \lambda_1)^2 + \lambda_2^2}{(\epsilon - \lambda_1)^2 + \lambda_2^2} = 1 + \frac{4x}{(x-1)^2 + y^2},$$

where $x = \epsilon/\lambda_1$ and $y = \lambda_2/\lambda_1$. Note that $y^2 \geq \alpha^2$ and without loss of generality we take $y \geq \alpha$. Define

$$h(x, y) = \frac{4x}{(x-1)^2 + y^2}$$

Note that $h(x, y) \to 0$ as $|x|^2 + |y|^2 \to \infty$. We must show that $h(x, y)$ is bounded above -1 for $y \geq \alpha$. It is enough to consider points where $\partial h/\partial x = 0$ which occur when $x_\pm = \pm\sqrt{1 + y^2}$. We have

$$h(x_\pm, y) = \frac{\pm 2}{\sqrt{1 + y^2} \mp 1} \geq \frac{-2}{\sqrt{1 + \alpha^2} + 1} > -1,$$

and hence we have proved the required boundedness. We then define

$$\mathrm{PV} \int_\gamma R(z, A) x dz = \lim_{\epsilon\downarrow 0} F_\epsilon(x, A).$$

The relation (4.2.3) now follows from the dominated convergence theorem. □

4.3 Computation of Measures

For the sake of brevity, the rest of this chapter will, unless otherwise stated, consider the self-adjoint case $A \in \Omega_{\mathrm{SA}}$, which is the case most encountered in applications (see [Col19a] for numerical examples involving unitary operators). However, the algorithms built are based on Proposition 4.2.1 (and Corollary 4.2.2) and the link with Poisson kernels/Cauchy transforms. Given the relation (4.1.5) and Proposition 4.2.4, the results can be straightforwardly extended to the unitary case and more general cases where conditions similar to that of Proposition 4.2.4 hold.

4.3.1 Full spectral measure

We start by considering the computation of $E_U^A x$ 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

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

Note that such a decomposition always exists. 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*

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

Then $\{\Xi_{\mathrm{meas}}, \Omega_{f,\alpha,\beta} \times \mathcal{U}, \widetilde{\Lambda}_1\} \in \Delta_2^A$*. In other words, we can construct a convergent sequence of arithmetic algorithms for the problem.*

Remark 4.3.2. *What this theorem essentially tells us is that if we can compute the action of the resolvent operator with asymptotic error control, then we can compute the spectral measures of open sets in one limit. In the unitary case, this can easily be extended to relatively open sets of* $\mathbb{T}$*. For any* $U \in \mathcal{U}$ *the approximation of* $E_U^A x$ *has finite support, and hence we can take inner products to compute* $\mu_{x,y}^A(U)$*.*

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 $\widetilde{\Lambda}_1$). □

This theorem can clearly be extended to cover the more general case of Proposition 4.2.4 if γ is regular enough to allow approximation of

$$\mathrm{PV}\int_\gamma R(z,A)x\,dz$$

given the ability to compute $R(z,A)x$ with asymptotic error control. Note that when it comes to numerically computing the integrals in Propositions 4.2.3 and 4.2.4, it is advantageous to deform the contour away from the spectrum. This ensures that the resolvent has a smaller Lipschitz constant. The proof can also be adapted to compute $E_I x$ where I is a closed interval by considering intervals shrinking to $[a,b]$ (a,b finite). A special case of this is the computation of the spectral measure of singleton sets. However, for these it much easier to directly use the formulae

$$E^A_{\{u\}}x = \lim_{\epsilon\downarrow 0}\epsilon\cdot\pi K_H(u+i\epsilon;A,x),\quad E^A_{\{\exp(i\theta)\}}x = \lim_{\epsilon\downarrow 0}\epsilon\cdot\pi i\exp(i\theta)K_D((1-\epsilon)\exp(i\theta);A,x),$$

for $A\in\Omega_{\mathrm{SA}}$ and $A\in\Omega_{\mathrm{U}}$ respectively. This idea will be used numerically in §4.6.3.

In the setting of Theorem 4.3.1, it is possible to compute the convolutions with error control. One may also wonder whether it is possible to upgrade the convergence of the algorithm in Theorem 4.3.1 from Δ_2 to Δ_1. In other words, whether it is possible to compute the measure with error control. However, this is difficult because the measure may be singular. Theorem 5.3.2 shows this is impossible even for singleton sets and discrete Schrödinger operators acting on $l^2(\mathbb{N})$.

4.3.2 Measure decompositions and projections

Recall from §4.1 that $P^A_{\mathcal{I}}$ denotes the orthogonal projection onto the space $\mathcal{H}^A_{\mathcal{I}}$, where $\mathcal{I}$ denotes a generic type ($\mathrm{ac},\mathrm{sc},\mathrm{pp},\mathrm{c}$ 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. In this section, we prove the following theorem.

Theorem 4.3.3. *Given the set-up in §§4.1 and 4.3.1, consider the map*

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

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.

Remark 4.3.4. *In the special case that $x = y$ so that the measure is real-valued, we can define the Σ and Π classes as in §2.2. In this case, the proof shows that some of these computation problems lie in the relevant Σ and Π classes, which will be used in §5.3.*

Proof of inclusion in Theorem 4.3.3

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, by Theorem 4.3.1 and Remark 4.3.2, 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

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

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

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

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,[6]

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

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

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

[6] 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

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

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 $t \in \text{supp}(f_s)$,

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

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

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

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

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

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

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

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

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

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

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

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

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

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 4.3.3. □

Proof of exclusion in Theorem 4.3.3

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 5.3.4). 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 4.3.3. 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 and Remark 4.3.2, 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

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

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

$$\widetilde{\Gamma}_n(H_v) \to \frac{\langle P^{H_v}_{\mathrm{pp}} E^{H_v}_{(0,4)} e_1, e_1\rangle}{\mu^{H_v}_{e_1,e_1}((0,4))}.$$

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 $\langle P^{H_{v_1}}_{\mathrm{pp}} E^{H_{v_1}}_{(0,4)} e_1, e_1\rangle = \mu^{H_{v_1}}_{e_1,e_1}((0,4))$. 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, $\langle P^{H_{v_2}}_{\mathrm{pp}} E^{H_{v_2}}_{(0,4)} e_1, e_1\rangle = 0$ but $\mu^{H_{v_2}}_{e_1,e_1}((0,4)) \neq 0$, 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 $\mu^{H_v}_{e_1,e_1}((0,4)) \neq 0$ 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 [Col19b], 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 5.3.4, $\langle P_{\mathrm{ac}}^{H_v} E_{(0,4)}^{H_v} e_1, e_1\rangle = \mu_{e_1,e_1}^{H_v}((0,4))$ 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

4.4.1 Computation of the functional calculus

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

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

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 $\mu^A_{x,y}$ 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 $n^2 C_2$ for some (possibly unknown) constants C_1 and C_2, we can approximate this integral with an error that vanishes in the limit $n \to \infty$. □

4.4.2 Computation of the Radon–Nikodym derivative

Recall the definition of the Radon–Nikodym derivative in (4.1.1) and note that $\rho^A_{x,y} \in L^1(\mathbb{R})$ 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*

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

We restrict this map to the quadruples (A, x, y, U) *such that* U *is strictly separated from* $\mathrm{supp}(\mu^A_{x,y,\mathrm{sc}}) \cup \mathrm{supp}(\mu^A_{x,y,\mathrm{pp}})$ *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.*

Remark 4.4.3. *What this theorem essentially tells us is that, if we can compute the action of the resolvent operator with asymptotic error control, then we can compute the Radon–Nikodym derivative of the absolutely continuous part of the measures on open sets which are a positive distance away from the singular support of the measure. The assumption that U is separated from $\operatorname{supp}(\mu^A_{x,y,\mathrm{sc}}) \cup \operatorname{supp}(\mu^A_{x,y,\mathrm{pp}})$ may seem unnatural but is needed to gain L^1 convergence of the approximation. However, without it, the proof still gives almost everywhere pointwise convergence. We will see in §4.5.2 how to compute the Radon–Nikodym derivative in L^p spaces with error control when it is sufficiently regular (see Corollary 4.5.9).*

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

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

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 $\operatorname{supp}(\mu^A_{x,y,\mathrm{sc}}) \cup \operatorname{supp}(\mu^A_{x,y,\mathrm{pp}})$, 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

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

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

4.5 High-order Kernels/Convergence and Error Control

As $\epsilon \downarrow 0$, typically we need to take larger n in Proposition 4.2.1 as the linear system approaches becoming singular and we adaptively compute the resolvent to the required accuracy. Hence, it is beneficial to design algorithms with faster convergence rates in terms of the smoothing parameter ϵ. In this section, we show that under regularity assumptions on the spectral measure[7], convergence can be obtained with error control and with arbitrarily high orders of convergence through convolutions with rational kernels (which are shown in Table 4.1 and Figure 4.4). This is done in both the pointwise and L^p senses, and the convolutions can be computed via the resolvent. A collocation approach was also constructed by the author in [Col19a] and found to speed up the computation of the spectral measures if they are smooth enough. Our approach here is *local*, compared to our *global* collocation approach in [Col19a]. There are several advantages of the approach adopted here - we do not need to know the support of the spectrum, and the convergence rate is only affected by how smooth the measure is locally (the global collocation approach, on the other hand, is adversely affected by singular behaviour, even at large distances from where the measure is singular). Furthermore, high-order kernels allow computation of the functional calculus with high orders of convergence, regardless of the regularity properties of the spectral measure.

Remark 4.5.1. *The use of the Poisson kernel allows the computation of the pure point part of the spectral measure (see Chapter 5 and the example in §4.6.3). For further examples connected with orthogonal polynomials, see [Col19a]. For eigenvalues separated from the rest of the spectrum, the use of high-order kernels makes the 'spikes' in such plots (see Figure 4.12) more pronounced because the convolution of the spectral measure with the kernel decays more rapidly to zero off the spectrum.*

4.5.1 A motivating integral operator example

As motivation for high-order methods, 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

[7]For example, when considering PDEs on the real line or half-line, it is sometimes known apriori how smooth the measure is locally. The kernels presented here can be used in this case also.

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. This example was developed in collaboration between the author and Andrew Horning, and appears in [CHT20].

There are two natural limits to take: $N \to \infty$ and $\epsilon \downarrow 0$. These two limits must be taken with considerable care [Col19a]. For this example, which uses a square discretisation, fixing N and taking $\epsilon \downarrow 0$ would (ignoring roundoff errors in the computation) simply recover the spectral measure of the discretisation - a series of Dirac measures located at the eigenvalues. Instead, as one takes $\epsilon \downarrow 0$, one must appropriately increase N too. Proposition 4.2.1 (or rather its generalisation to other methods where an error bound can be approximated) 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.[8]

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 $\mu^{\mathcal{L}}_{f,f}$ 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 4.5.3 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

In this subsection, we derive convergence rates and error bounds for convolutions with high-order kernels. The question of implementation is delayed until §4.5.3. In this subsection, we use the notation x to denote a point in $\mathbb{R}$. 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 (similar to definition 1.3 of [Tsy08]).

Definition 4.5.2 (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:*

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

[8] Two reasons for this, explored in more detail in [CHT20], 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).

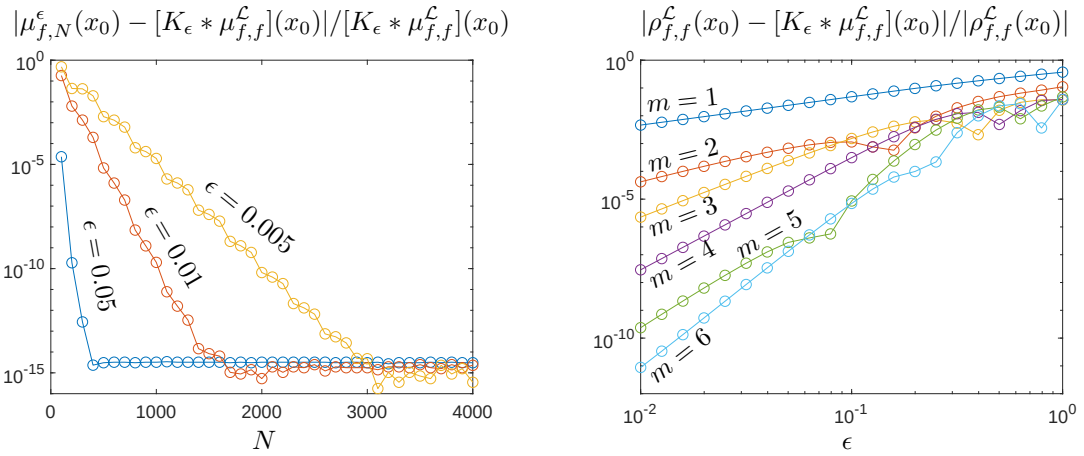


Figure 4.3: Left: The relative error in the numerical approximation, denoted by $\mu^{\epsilon}_{f,N}$, corresponding to discretisation size N, of the smoothed measure $[K_\epsilon * \mu^{\mathcal{L}}_{f,f}](x_0)$ (K_ϵ denotes the rescaled Poisson kernel) for the operator in (4.5.1) with $\epsilon = 0.05$, $\epsilon = 0.01$, 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.

(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 $h_\epsilon * \mu^A_{x,y}$. 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 $\mu^A_{x,y}$ 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 (which also takes into account the distance of a point to where the measure may be singular) describes the pointwise convergence rates.

Theorem 4.5.3. *Let K be an mth 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 4.5.4. *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).*

Remark 4.5.5. *Similar results to Theorem 4.5.3, without the first term $C_K\epsilon^m/(\epsilon + \frac{\eta}{2})^{m+1}$, for absolutely continuous probability measures with globally Hölder continuous density functions are used in kernel density estimation in statistics (see, for example, proposition 1.2 of [Tsy08]).*

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

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

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 4.5.6. *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 $\alpha \in (0,1)$, $\widehat{K} \in C^{m-1,\alpha}(\mathbb{R})$.*

Proof. The first part follows from (4.5.2) since we can use the dominated convergence theorem to differentiate $(m-1)$ times under the integral in the definition of $\widehat{K}$. The fact that $(\widehat{K})^{(j)}(0) = 0$ for $j = 1, ..., m-1$ follows from the vanishing moment criterion since

$$(\widehat{K})^{(j)}(0) = (2\pi i)^j \int_{\mathbb{R}} K(x)x^j dx.$$

For $(\widehat{K})^{(m-1)} \in C^{\alpha}(\mathbb{R})$, we argue when $m = 1$ and the general case is similar. Let $\tau \neq 0$ and note that by Hölder's inequality

$$\begin{aligned}\left|\frac{\widehat{K}(\omega+\tau) - \widehat{K}(\omega)}{|\tau|^{\alpha}}\right| &\leq \int_{\mathbb{R}} |K(x)| \left|\frac{\exp(2\pi i\tau x) - 1}{\tau^{\alpha}}\right| dx \\ &\leq \left(\int_{\mathbb{R}} \left|K(x)x^{\alpha+1/2}\right|^2 dx\right)^{1/2} \left(\int_{\mathbb{R}} \left|\frac{\exp(2\pi i\tau x) - 1}{x^{\alpha+1/2}|\tau|^{\alpha}}\right|^2 dx\right)^{1/2}\end{aligned}$$

Note that the first of these integrals is finite by (4.5.2). After a change of variables, the second integral becomes

$$\int_{\mathbb{R}} \left|\frac{\exp(2\pi ix) - 1}{x^{\alpha+1/2}}\right|^2 dx$$

which is bounded, independent of ω and τ. □

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 4.5.6 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 in terms of $G_{m,K}$.

Theorem 4.5.7. *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

$$\|\rho_I - [K_\epsilon * \mu]\|_{L^p,[a,b]} = \mathcal{O}(\epsilon^m). \tag{4.5.8}$$

Remark 4.5.8. *Similar results to Theorem 4.5.7, for $p = 2$ without the first term on the right-hand side of (4.5.6) and for absolutely continuous probability measures with $W^{m,2}(\mathbb{R})$ density function, are used in kernel density estimation in statistics (see, for example, proposition 1.5 of [Tsy08]). In this context, the L^2 error is used to bound the bias term in the mean integrated squared error. The case of L^1 convergence requires a different proof technique due to being on the 'boundary of integrability'.*

Proof of Theorem 4.5.7. 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 4.5.6, 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

$$\|K_\epsilon * g - g\|_{L^p,\mathrm{supp}(g)} \leq \left\| [\chi_{[-L,L]}\phi_{m,K_\epsilon}] * g^{(m)} \right\|_{L^p} \tag{4.5.13}$$

$$\leq \int_{-L}^{L} |\phi_{m,K_\epsilon}(x)|\, dx \cdot \|g^{(m)}\|_{L^p}. \tag{4.5.14}$$

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 $\mathcal{O}(\epsilon^m)$. □

The constants in Theorems 4.5.3 and 4.5.7 can be made explicit if desired and hence Theorems 4.5.3 and 4.5.7 applied with the Poisson kernel (by altering the proof of Theorem 4.4.2) immediately imply the following corollary.

Corollary 4.5.9 (Computation of the Radon–Nikodym derivative with error control)**.** *Let $I = (a, b) \subset \mathbb{R}$ be a fixed bounded open interval, $p \in [1, \infty]$ (note that we allow the value $p = \infty$), $\delta \in (0, 1)$, $\eta > 0$ and $M > 0$. Consider the map*

$$\begin{aligned} \Xi^p_{\mathrm{RN}} : \Omega_{f,\alpha} \times l^2(\mathbb{N}) \times l^2(\mathbb{N}) &\to L^p(I) \\ (A, x, y) &\to \rho^A_{x,y}|_I. \end{aligned}$$

We restrict this map to the quadruples (A, x, y) such that $\mu^A_{x,y}$ is absolutely continuous on $(a - \eta, b + \eta)$ and so that bounds on $\|(I - P_{f(n)})AP_n\|$, $\|P_n x - x\|$ and $\|P_n y - y\|$ are known explicitly (i.e. bounded by a known null sequence). We also restrict so that

- *In the case that $p = \infty$, $\rho^A_{x,y}|_{(a-\eta,b+\eta)}$ is at least $C^{0,\delta}((a - \eta, b + \eta))$ with*

$$\left\| \rho^A_{x,y}|_{(a-\eta,b+\eta)} \right\|_{C^{0,\delta}((a-\eta,b+\eta))} \leq M.$$

- *In the case that $p < \infty$, $\rho^A_{x,y}|_{(a-\eta,b+\eta)} \in W^{1,p}((a - \eta, b + \eta))$ with*

$$\left\| \rho^A_{x,y}|_{(a-\eta,b+\eta)} \right\|_{W^{1,p}((a-\eta,b+\eta))} \leq M.$$

and denote this subclass by Ω^p_f (where for notational convenience we have dropped the dependence on I, η, δ, M). Then $\{\Xi^p_{\mathrm{RN}}, \Omega^p_f, \widetilde{\Lambda}_1\} \in \Delta^A_1$. In other words, we can compute the Radon–Nikodym derivative, in the suitable $L^p(I)$ space, with error control.

Finally, 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 4.5.3 and 4.5.7 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, and an example of this is presented in §4.6.2.

4.5.3 Constructing rational kernels

Theorems 4.5.3 and 4.5.7 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 $\mu^A_{x,y} * K_\epsilon$ with error control through the formula

$$\mu^A_{x,y} * K_\epsilon(u) = \frac{-1}{2\pi i} \left[\sum_{j=1}^{n_1} \alpha_j \langle R(u - \epsilon a_j, A)x, y\rangle - \sum_{j=1}^{n_2} \beta_j \langle R(u - \epsilon b_j, A)x, y\rangle \right]. \tag{4.5.18}$$

Remark 4.5.10. *Throughout this section (as in the rest of the numerical examples), we used the adaptive method in Proposition 4.2.1 to tell us, given ϵ, how large n should be when solving for the resolvents $R(u - \epsilon a_j, A)x$.*

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

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_{n_1} \\ a_1^2 & a_2^2 & \cdots & a_{n_1}^2 \\ \vdots & \vdots & & \vdots \\ a_1^{m-1} & a_2^{m-1} & \cdots & a_{n_1}^{m-1} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{n_1} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{4.5.19}$$

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

$$\mu^A_{x,x} * K_\epsilon(u) = \frac{-1}{\pi} \mathrm{Im} \left[\sum_{j=1}^{m} \alpha_j \langle R(u - \epsilon a_j, A)x, x\rangle \right], \tag{4.5.20}$$

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

$$a_j = \frac{2j}{m+1} - 1 + i, \qquad b_j = \overline{a_j}, \qquad 1 \le j \le m. \tag{4.5.21}$$

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

m	$\pi K(u)\prod_{j=1}^{m}(u-a_j)(u-\overline{a_j})$	$\{\alpha_1,\ldots,\alpha_{\lceil m/2\rceil}\}$
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 $\alpha_{m+1-j}=\overline{\alpha_j}$.

useful when one wishes to sample the smoothed measure $K_\epsilon * \mu^A_{x,y}$ 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

In this subsection, we demonstrate accelerated convergence for locally smooth measures using rational kernels. 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 $a_j > 0$. In this case, under suitable conditions, the probability measure $\mu_J := \mu^J_{e_1,e_1}$ is exactly the probability measure associated with the orthonormal polynomials defined by

$$\begin{aligned} xP_k(x) &= a_{k+1}P_{k+1}(x)+b_{k+1}P_k(x)+a_kP_{k-1}(x),\\ P_{-1}(x) &= 0, \quad P_0(x)=1, \end{aligned}$$

and the spectral measure that appears in the (multiplicative version of the) spectral theorem (see, for example, [Tes00, Dei99, Sto90]) .

Classically, one usually first considers the measure and then constructs the orthogonal polynomials (and the corresponding J). In some sense, the algorithms constructed in this chapter compute the inverse problem. In other words, we compute the measure μ_J given the recurrence coefficients defining the orthogonal polynomials. For a host of numerical examples of the methods introduced in this chapter, including singular measures and quindiagonal unitary matrices corresponding to measures on the unit circle (here we can simply take $f(n)=n+2$ and use the Poisson kernel for the unit disk[9]), we refer the reader to [Col19a]. A collocation approach is also constructed in [Col19a] and found to speed up the computation of the spectral measures if they are smooth enough. Instead of repeating the numerical examples of the author in [Col19a], we indicate how the rational kernels constructed in §4.5.3 (with the choice of poles in (4.5.21)) can be used to increase convergence rates for smooth enough measures when taking convolutions.

[9] Using similar techniques, one can also develop high-order rational kernels for convolution on the unit circle and unitary operators.

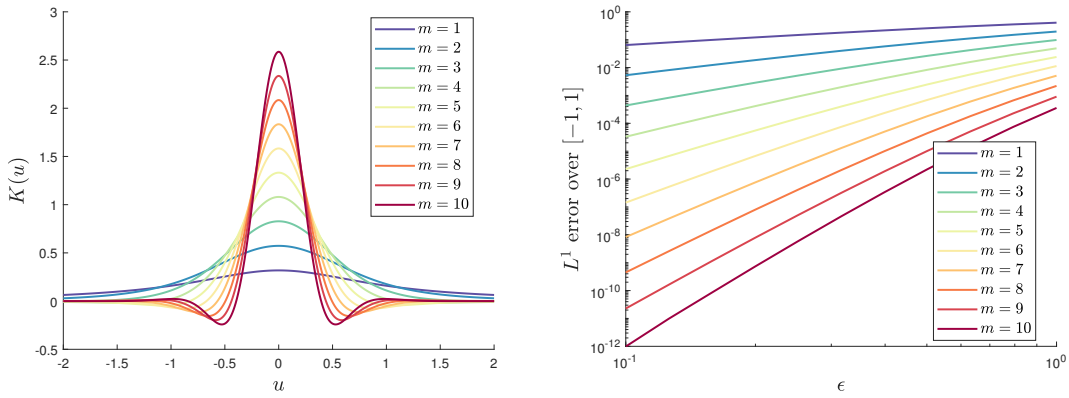


Figure 4.4: Left: Kernels used for convolution. Right: Convergence for Gaussian measure.

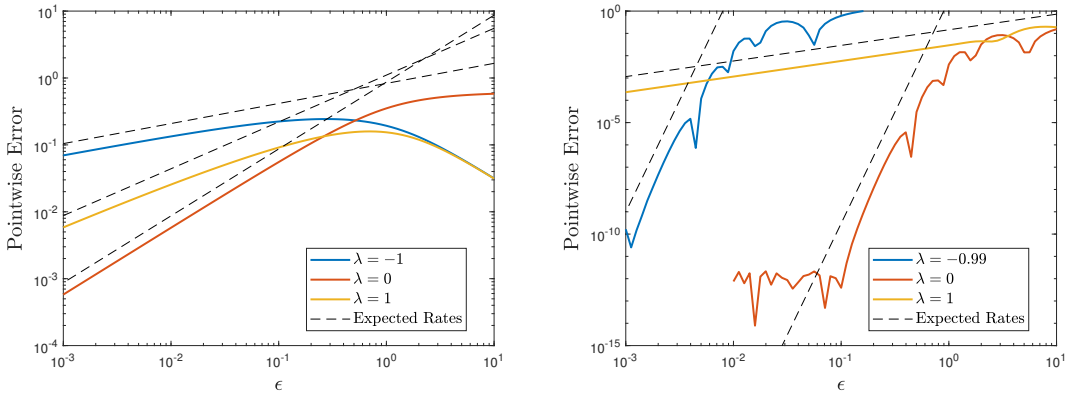


Figure 4.5: Left: Pointwise errors for $\lambda = -1, 0, 1$ for $m = 1$ and $\alpha = 0.7$, $\beta = 0.3$. Right: Pointwise errors for $\lambda = -0.99, 0, 1$ for $m = 10$ and $\alpha = 0.7$, $\beta = -0.3$.

As a simple example, we consider the case when $a_k = \sqrt{k/2}$ 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)[10] from Theorems 4.5.3 and 4.5.7.

As a second example, we consider the case of 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 4.5.3. 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

[10] 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 $\prod_{j=1}^m a_j = \prod_{j=1}^m b_j$.

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 4.5.3.

4.6 Numerical Examples and Applications

4.6.1 Magneto-graphene Schrödinger operator

For the first example of this section, we apply our 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 (see Figure 4.6), whose unusual properties are studied in condensed-matter physics [NGP$^+$09, Nov11]. The magnetic properties of graphene are important because of the experimental observation of the quantum Hall effect and Hofstadter's butterfly [PGY$^+$13], and the exciting new area of twistronics [Cha19, LSY$^+$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 $T_1, T_2, T_3 : \ell^2(\mathbb{Z}^2; \mathbb{C}^2) \to \ell^2(\mathbb{Z}^2; \mathbb{C}^2)$ 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 is absolutely continuous, and hence we have plotted the Radon–Nikodym derivative of the measure $\mu^{H_0}_{e_1,e_1}$. 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 $\rho^{H_0}_{e_1,e_1}$ 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.[11] The support of $K_\epsilon * \mu^{H_0}_{e_1,e_1}$ is the whole real line due to the non-compact support of the kernel K. However, if $\lambda \notin \mathrm{Sp}(H_0)$, then $|[K_\epsilon * \mu^{H_0}_{e_1,e_1}](\lambda)| \leq C_K\epsilon^m(\epsilon + \mathrm{dist}(\lambda, \mathrm{Sp}(H_0)))^{-(m+1)}$, where C_K is the constant in (4.5.2) and m is the order of the kernel, so $|[K_\epsilon * \mu^{H_0}_{e_1,e_1}](\lambda)|$ decays rapidly off of the spectrum. We also

[11] 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.

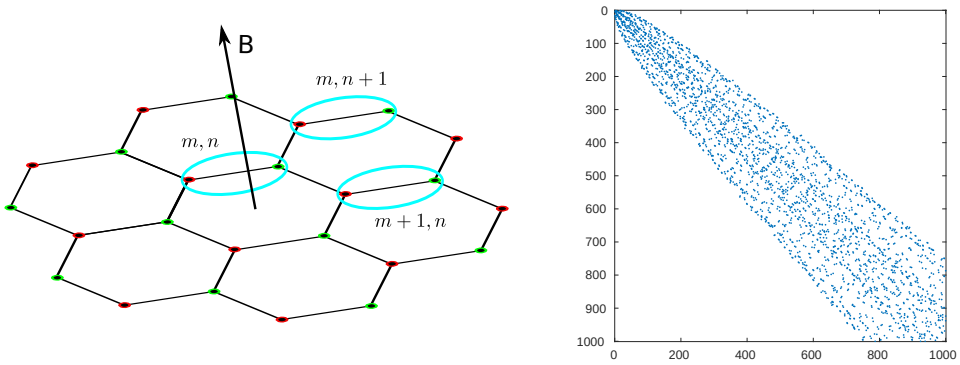


Figure 4.6: Left: Honeycomb structure of graphene as a bipartite graph. We have shown the spinor structure via the circled lattice vertices and the indexing via (m, n). The arrow shows the perpendicular magnetic field **B**. Right: Sparsity structure of the first $10^3 \times 10^3$ block of the infinite matrix, and the corresponding growing local bandwidth.

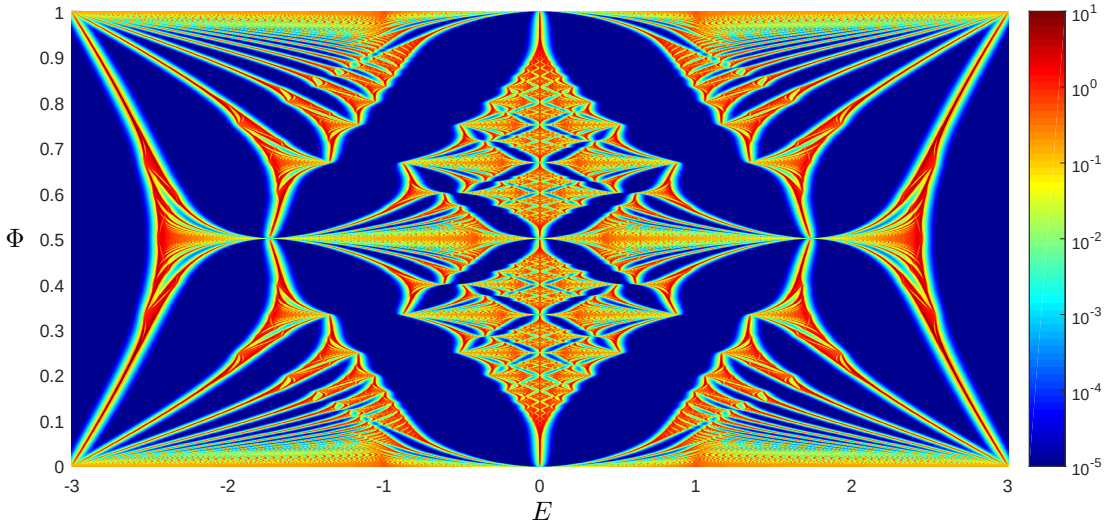


Figure 4.7: Radon–Nikodym derivative $\rho^{H_0}_{e_1,e_1}$ (`log10` scale) of the measure for various magnetic field strengths Φ. The axis label E (energy) stands for the spectral parameter. The Radon–Nikodym derivative is computed to high precision using $\epsilon = 0.01$ and a fourth-order kernel 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 $\rho^{H_0}_{e_1,e_1}$ (which vanishes in these gaps).

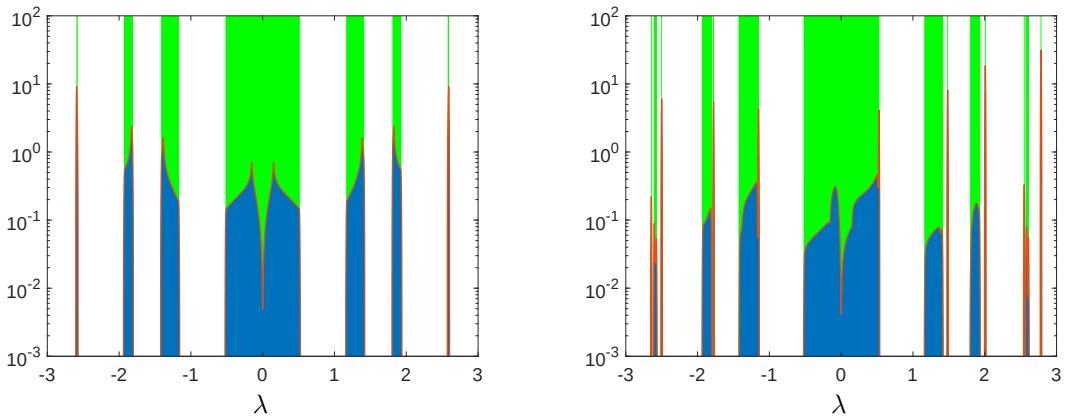


Figure 4.8: Left: Smoothed measure with no potential. We show the algorithm from Chapter 3 as shaded strips (green) for comparison. Right: The same computation but with the added potential in (4.6.1). The additional eigenvalues correspond to spikes in the smoothed measure.

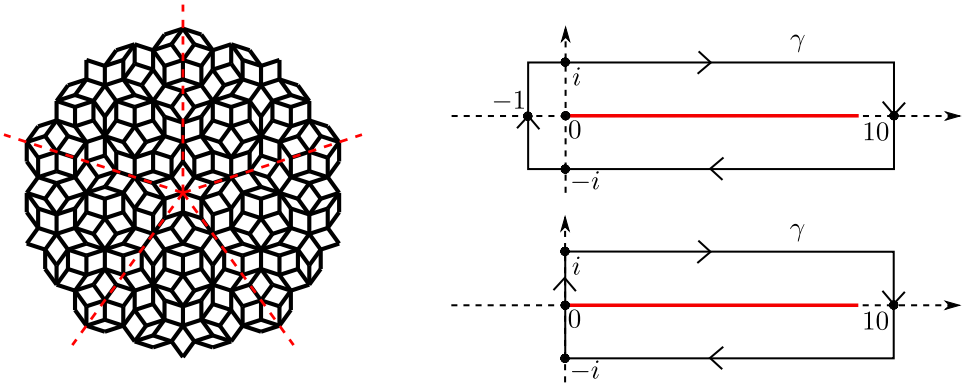


Figure 4.9: Left: Finite portion of Penrose tile showing the fivefold rotational symmetry. Vertices are ordered from the centre in a spiral outwards in increasing distance from the origin. Right: Contours used for the fractional diffusion on the Penrose tile ($\alpha \in \mathbb{N}$ top, $\alpha \notin \mathbb{N}$ bottom). The red line represents the interval containing the spectrum, the branch cut for z^α is taken to be $\mathbb{R}_{\leq 0}$.

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 Fractional diffusion on a quasicrystal

In this example, we demonstrate computation of the functional calculus and consider the transport Hamiltonian on a Penrose tile (shown in the left of Figure 4.9) discussed in §3.6.1 of Chapter 3. Recall that the

free Hamiltonian H_0 (Laplacian) is given by

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

with summation over nearest neighbour sites (vertices). We chose an ordering of the vertices explained in the caption to Figure 4.9, which leads to an operator H_0 acting on $l^2(\mathbb{N})$. The local bandwidth grows for this operator (our ordering is asymptotically optimal) and hence computation of powers H_0^m are infeasible for $m \gtrsim 50$, rendering polynomial approximations of the functional calculus intractable. In the above notation, $H_0 \in \Omega_{f,0}$ with $f(n) - n = \mathcal{O}(\sqrt{n})$. Throughout, we take $u_0 = e_1$, though different initial conditions are handled in the same manner.

The ability to compute the functional calculus allows the solution of linear evolution equations. Given $A \in \Omega_{\mathrm{N}}$, a function F (continuous and bounded on $\mathrm{Sp}(A)$) and $u_0 \in l^2(\mathbb{N})$, consider the evolution equation

$$\frac{du}{dt} = F(A)u, \quad u_{t=0} = u_0. \tag{4.6.2}$$

The solution of this equation is

$$u(t) = \exp(F(A)t)u_0$$

and can be computed via the algorithm outlined in §4.4.1. We consider fractional diffusion governed by

$$\frac{du}{dt} = -(-H_0)^{\alpha}u, \quad u_{t=0} = u_0,$$

for $\alpha > 0$. If α is an integer, then the solution can be represented via contour deformation as

$$u(t) = \frac{1}{2\pi i}\int_{\gamma} \exp(-z^{\alpha}t)R(z, -H_0)u_0 dz, \tag{4.6.3}$$

where γ is a closed contour looping once (clockwise due to $R(z, A) = (A - zI)^{-1}$) around the spectrum. We took the rectangular contour shown in Figure 4.9 and approximated the integral via Gauss–Legendre quadrature along each line segment. This allows us to compute the solution with error control (we known the minimal distance between γ and $\mathrm{Sp}(-H_0)$ so can bound the Lipschitz constant of the resolvent) and clearly, this holds for other functions F, holomorphic on a neighbourhood of $\mathrm{Sp}(-H_0)$. Note that other methods, such as direct diagonalisation of finite square truncations or discrete time stepping (which is difficult if $\alpha \notin \mathbb{N}$), do not give error control and are slower. In fact, for this example, direct diagonalisation was impractical. Our approach deals directly with the infinite-dimensional operator (the 'bulk' operator in physicists' terminology), and hence the numerical results are free from finite size effects. When $\alpha \notin \mathbb{N}$, we can still deform the contour but not at 0 since $0 \in \mathrm{Sp}(-H_0)$, so we deform the contour to that shown in Figure 4.9. For a discussion of contour methods applied to finite matrices (in the case that the spectrum is strictly positive), we refer the reader to [HHT08]. Unfortunately, such methods cannot be applied here since $0 \in \mathrm{Sp}(-H_0)$.

Figure 4.10 shows the convergence of the algorithm for $\alpha = 1/2$ and $\alpha = 1$. For $\alpha = 1/2$, error control is not given by our algorithm, so we computed an error by comparing to a 'converged' solution using larger n (note that this is not a 100% verifiable and rigorous error bound, unlike the error bound provided by the algorithm for the holomorphic case). The l^2 error refers to the error in $l^2(\mathbb{N})$. The method converges algebraically for $\alpha = 1/2$ (owing to the contour touching the spectrum at 0) but converges exponentially for $\alpha = 1$ with similar convergence observed over a large range of times t. Figure 4.11 shows the magnitude (log scale) of the computed solution for various times. As expected, a smaller α

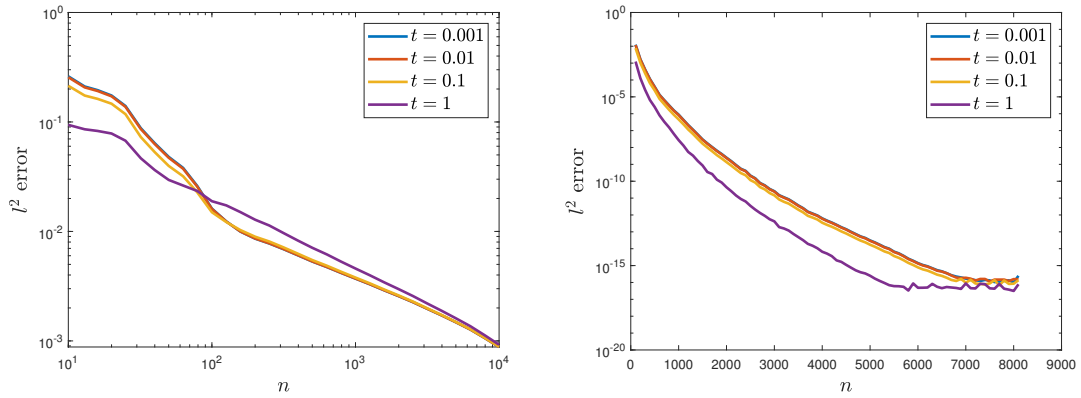


Figure 4.10: Left: Convergence for $\alpha = 1/2$. Right: Convergence for $\alpha = 1$. We have plotted the errors as a function of the matrix size (number of matrix columns in the rectangular truncations) used.

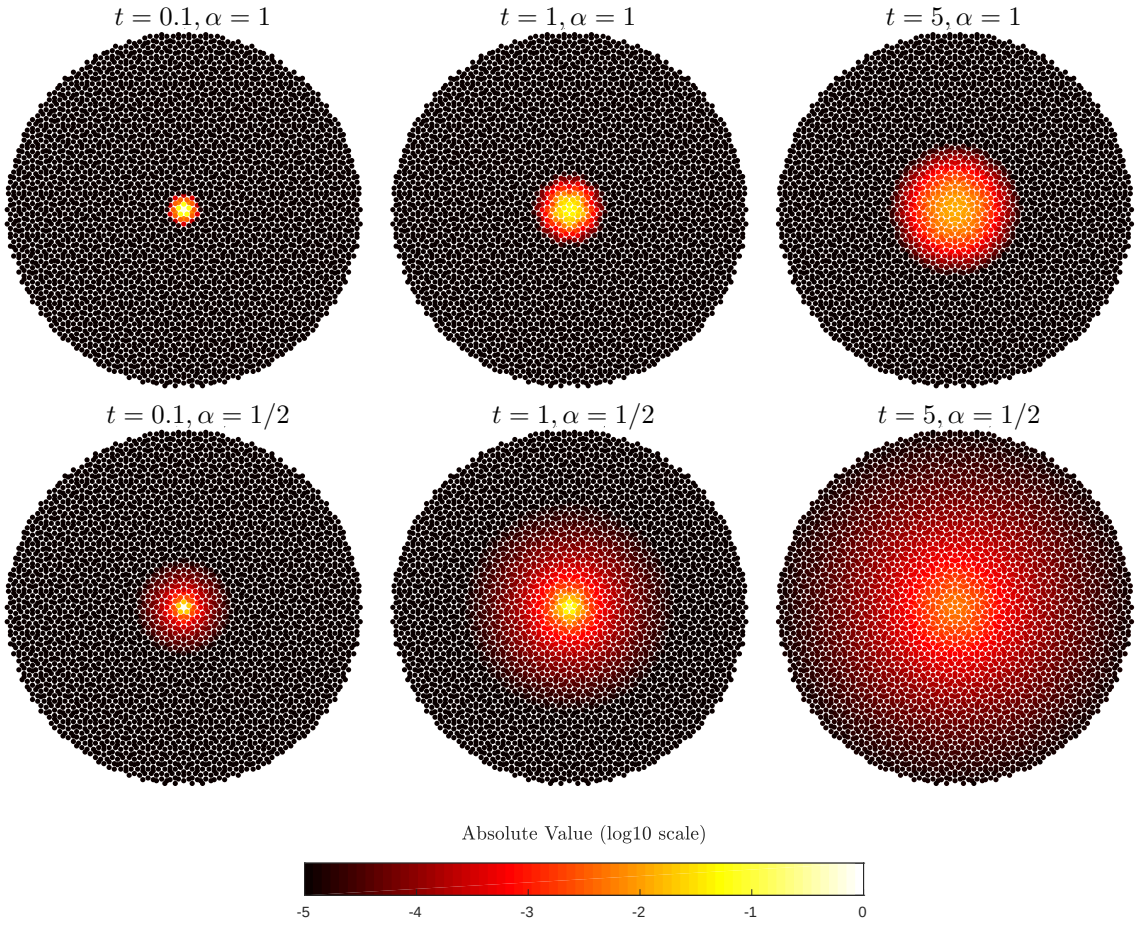


Figure 4.11: Evolution of initial wavepacket under fractional diffusion.

corresponds to more spreading out of the initial wavepacket. Similar results were found for other α. Note that the techniques presented here can be applied to any evolution equation of the form (4.6.2) on *infinite-dimensional* Hilbert spaces. They may also be useful for splitting methods/exponential integrators which require fast computation of matrix/operator exponentials (see [HO10, MQ02] and the references therein) and more generally in the field of infinite-dimensional ODE/PDE systems.

4.6.3 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 (see also the discussion in §4.1.3).

We consider the case of 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 is a first-order differential operator acting 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

$$\mathrm{Sp}(D_V) = \mathrm{cl}\left(\bigcup \mathrm{Sp}\left(D_V^{k_j}\right)\right), \quad \mathrm{Sp}_d(D_V) = \bigcup \mathrm{Sp}_d\left(D_V^{k_j}\right).$$

From now on, we will 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. The numerical code for this example was developed in collaboration between the author and Andrew Horning. A full discussion of the ultraspherical method is beyond the scope of this thesis. 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[12]

$$\lim_{\epsilon\downarrow 0}\nu_f^\epsilon(\lambda) = \begin{cases} \|P_{E_j} f\|^2, & \text{if } \lambda = E_j \\ 0, & \text{otherwise} \end{cases}.$$

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

$$E_j = \left(1 + \frac{\gamma^2}{\left(j+\sqrt{1-\gamma^2}\right)^2}\right)^{-1/2}, \quad j \in \mathbb{Z}_{\geq 0}.$$

[12] 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 $\mathcal{O}(\epsilon)$.

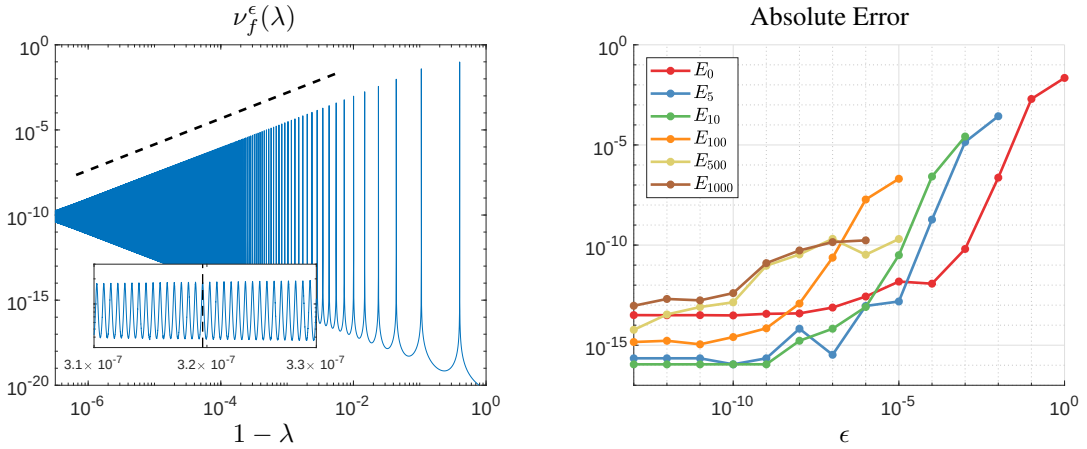


Figure 4.12: Left: The function $\nu_f^\epsilon(x)$ for λ near 1. We have plotted the function against $1-\lambda$ to aid visability of the accumulation at $\lambda = 1$. The sloped dashed line shows the algebraic decay of $\|P_{E_j} f\|^2$ (approximately $\mathcal{O}(j^{-3})$). The magnified region 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 4.12 (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 4.12 (right), we show the absolute error in the computed eigenvalues E_j for $j = 0, 5, 10, 100, 500$, and 1000 as $\epsilon \downarrow 0$. We can resolve hundreds of eigenvalues, even when highly clustered, to an accuracy of essentially machine precision. Finally, we remark that the methods of Chapters 3 and 6 can also be used to compute discrete spectra of Dirac operators.

Chapter 5

Computing Spectral Type

This chapter 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]. We also prove a theorem of independent interest regarding finite rank perturbations of Anderson models.

5.1 Computing Spectral Types as Sets - the Main Result

Define the problem functions $\Xi^{\mathbb{C}}_{\mathcal{I}}(A) = \mathrm{Sp}_{\mathcal{I}}(A)$ 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*

$$\Delta^G_2 \not\ni \{\Xi^{\mathbb{C}}_{\mathrm{ac}}, \Omega_{f,\alpha}, \Lambda_1\} \in \Sigma^A_2, \quad \Delta^G_2 \not\ni \{\Xi^{\mathbb{C}}_{\mathrm{sc}}, \Omega_{f,\alpha}, \Lambda_1\} \in \Sigma^A_3, \quad \Delta^G_2 \not\ni \{\Xi^{\mathbb{C}}_{\mathrm{pp}}, \Omega_{f,\alpha}, \Lambda_1\} \in \Sigma^A_2.$$

If $f(n) - n \geq \sqrt{2n} + \frac{1}{2}$, then the sharp lower bound $\{\Xi^{\mathbb{C}}_{\mathrm{sc}}, \Omega_{f,0}, \Lambda_1\} \notin \Delta^G_3$ also holds.

Remark 5.1.2. *In [Col19a] it was shown that, with slightly more information, $\Xi^{\mathbb{C}}_{\mathrm{pp}}$ can be computed in two limits for the class $\mathcal{C}(l^2(\mathbb{N}))$ in such a way that the set after the first limit is contained in the point spectrum (i.e. we recover a portion of the eigenvalues after one limit).*

The difficulty encountered when computing the singular continuous spectrum is partly due to the negative definition of the singular continuous part of a measure. It is the 'leftover' part of the measure, the part that is not continuous with respect to Lebesgue measure and does not contain atoms. The challenge of studying $\mathrm{Sp}_{\mathrm{sc}}$ analytically also reflects this difficulty - singular continuous spectra were once thought to be rather rare or exotic. However, they are quite generic; see, for example, [Sim95].

One might at first expect computational results to be independent of the function f due to tridiagonalisation. However, the infinite-dimensional case is much more subtle than the finite-dimensional case. Using Householder transformations on a bounded sparse self-adjoint operator A leads to a tridiagonal operator, but, in general, this operator is A restricted to a strict subspace of $l^2(\mathbb{N})$. Part of the operator may be lost in the strong operator limit. Instead, one must consider a sum of possibly infinitely many tridiagonal operators (see [Han08a] chapters 2 and 8). Hence some spectral problems may have different SCI classifications for different f which describe the off-diagonal decay structure of the matrix.

Theorem 5.1.1 shows that despite the results of Chapter 4, in general, it is very hard to compute the decomposition of the spectrum in the sense of (4.1.2). The proof of the negative result for point spectra uses the fractional moment method to prove a certain result connected to Anderson localisation (Theorem 5.2.1), which was also used in §4.3.2. As a by-product, we also answer the question posed in §4.2.2 and prove that the spectral measures, while computable in one limit, cannot be computed with error control (see Theorem 5.3.2), unless one has regularity assumptions as was the case in §4.5. We begin with some preliminary results concerning Anderson localisation and then deal with each spectral type.

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. Since P. W. Anderson's introduction of his famous model 60 years ago [And58] (which led to the Nobel prize in 1977), there has been a considerable amount of work by both physicists and mathematicians aiming to understand the suppression of transport due to disorder (Anderson localisation). A full discussion of Anderson localisation is beyond the scope of this thesis, and we refer the reader to [CL90, CFKS87, Kir07] for broader surveys. We will use the fractional moment method [AM93, Gra94, AW15] to prove Anderson localisation in the multi-dimensional setting under finite rank perturbations. The notation used is the same as that in §4.3.2. In particular, 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.

When considering Anderson localisation, we will 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. We assume that 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 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. We have included a short proof since the result may be of independent interest.

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*

$$W = \sum_{j=1}^{M} \alpha_j \left|x_{m_j}\right\rangle \left\langle x_{n_j}\right|. \tag{5.2.1}$$

In other words, the operator W's matrix with respect to the canonical basis has only finitely many non-zeros.

Remark 5.2.2. *We do not discuss the property of exponentially localised eigenfunctions. For this and cases such as less regular probability distributions, dependent potential sites, slowly decaying off-diagonal terms, and off-diagonal randomness, we refer the reader to the seminal paper [AM93].*

5.2.1 Proof of Theorem 5.2.1

Throughout this section we will fix a graph G as discussed in §5.2 and an operator W of the form (5.2.1). Recall that there exists some $N \in \mathbb{N}$ such that $\langle Wx, y\rangle = 0$ if $|x| \geq N$ or $|y| \geq N$. Given a (bounded) self-adjoint operator h acting on $l^2(V(G))$, we recall the following definition of the Green's function for $x, y \in V(G)$ and $z \in \mathbb{C}\backslash \mathrm{Sp}(h)$:

$$\mathcal{G}(x, y; z) = \langle (h - z)^{-1}\delta_y, \delta_x\rangle,$$

where $\delta_x(y) = \delta_{xy}$. We will use a subscript $\mathcal{G}_\omega$ when referring to a particular sampled operator $h = H_0 + v_\omega + W$. As mentioned in §5.2, our proof strategy will use the fractional moment method.

We follow the set-up and notation of Graf [Gra94]. We recall Ruelle's criterion (or the RAGE theorem) as follows. Let P_{c}^h be the projection onto the continuous spectral subspace of the operator h and let P_D denote the orthogonal projection onto wave functions with support inside the subset $D \subset V(G)$. Then for any $\psi \in l^2(V(G))$

$$\begin{aligned} \|P_{\mathrm{c}}^h \psi\|^2 &= \lim_{R\to\infty} \lim_{t\to\infty} \frac{1}{t} \int_0^t \|P_{|x|\geq R} e^{-ihs}\psi\|^2 ds \\ &= \lim_{R\to\infty} \lim_{\epsilon\downarrow 0} \frac{\epsilon}{\pi} \int_{-\infty}^{\infty} \|P_{|x|\geq R}(h - E - i\epsilon)^{-1}\psi\|^2 dE. \end{aligned} \tag{5.2.2}$$

The strategy is to bound the fractional moments of the Green's function (which can be used to prove dynamical localisation) via the following series of technical lemmas.

Lemma 5.2.3. *Let $0 < s < 1$. Then there exists some constant $C_1 = C_1(s, G)$ such that*

$$\mathbb{E}_\omega(|\mathcal{G}_\omega(x, y; z)|^s) \leq C_1 \|\rho\|_\infty^s$$

for all $z \in \mathbb{C}\backslash\mathbb{R}$ and $x, y \in V(G)$ with $|x|, |y| \geq N$.

Proof. The proof is virtually identical to that of Lemma 5 from [Gra94]. The only difference is that we need $|x|, |y| \geq N$ to apply the argument in order to avoid the perturbation caused by W. □

We also recall the following decoupling Lemma from [Gra94]:

Lemma 5.2.4 ([Gra94])**.** *Let $0 < s < 1$. Then there exists some constant $c > 0$ such that*

$$\frac{\int_{\mathbb{R}} \rho(\nu)(|\nu - \eta|^s / |\nu - \beta|^s) d\nu}{\int_{\mathbb{R}} \rho(\nu)(1/|\nu - \beta|^s) d\nu} \geq c \frac{\|\rho\|_1^s}{\|\rho\|_\infty^s}$$

for all $\rho \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$, $0 \not\equiv \rho \geq 0$ and all $\beta, \eta \in \mathbb{C}$.

Lemma 5.2.5. *Let $0 < s < 1$. Then if $\|\rho\|_\infty$ is sufficiently small (independent of W and only dependent on C_G) then there exists some constant $m > 0$ independent of W such that*

$$\mathbb{E}_\omega(|\mathcal{G}_\omega(x,y;z)|^s) \leq C_1\|\rho\|_\infty^s \exp(-m\min\{|x-y|, |y|-N\})$$

for all $z \in \mathbb{C}\backslash\mathbb{R}$ and $x, y \in V(G)$ with $|x|, |y| \geq N$.

Proof. The proof is easier and similar if $x = y$ so we assume that $x \neq y$. Let $\widehat{\omega}$ be obtained from ω by setting $v_x = v_y = 0$. Then we can write

$$H^\omega := H_0 + v_\omega + W = H_0 + v_{\widehat{\omega}} + W + v_x P_x + v_y P_y = H^{\widehat{\omega}} + v_x P_x + v_y P_y,$$

that is, we consider the final two terms as a rank two perturbation. Let $P = P_x + P_y = P_{\{x,y\}}$. Then an application of the second resolvent identity yields Krein's formula:

$$P(H^\omega - z)^{-1}P = (B + v_x P_x + v_y P_y)^{-1}, \tag{5.2.3}$$

where $B = [P(H^{\widehat{\omega}} - z)^{-1}P]^{-1}$ if it exists acts on $\mathcal{R}(P)$ (the range of P) and is independent of v_x, v_y. The inverse exists since $\mathrm{Im}(z)^{-1}\mathrm{Im}((H^{\widehat{\omega}} - z)^{-1})$ is positive definite. Explicitly, with respect to the basis $\{\delta_x, \delta_y\}$ of $\mathcal{R}(P)$, write

$$B = \begin{pmatrix} b_{xx} & b_{xy} \\ b_{yx} & b_{yy} \end{pmatrix},$$

then from (5.2.3),

$$\mathcal{G}_\omega(x,y;z) = -\frac{b_{xy}}{(v_x + b_{xx})(v_y + b_{yy}) - b_{xy}b_{yx}} = \frac{\alpha}{v_y - \beta} \tag{5.2.4}$$

where α, β are independent of v_y.

Assume that $|y| > N$ then we also have by considering the matrix elements of $(H^\omega - z)^{-1}(H^\omega - z) = I$ that

$$\sum_{e\sim y} \mathcal{G}_\omega(x,e;z) = (v_y + \zeta(y) - z)\mathcal{G}_\omega(x,y;z)$$

which implies that

$$\sum_{e\sim y} |\mathcal{G}_\omega(x,e;z)|^s \geq |v_y + \zeta(y) - z|^s \, |\mathcal{G}_\omega(x,y;z)|^s .$$

Using (5.2.4) and Lemma 5.2.4, this implies

$$\begin{aligned}\mathbb{E}_\omega\left(\sum_{e\sim y} |\mathcal{G}_\omega(x,e;z)|^s\right) &\geq \mathbb{E}_\omega(|v_y + \zeta(y) - z|^s \, |\mathcal{G}_\omega(x,y;z)|^s) \\ &\geq c\|\rho\|_\infty^{-s}\mathbb{E}_\omega(|\mathcal{G}_\omega(x,y;z)|^s).\end{aligned}$$

We assumed that any vertex degree of the graph G is bounded by C_G. We can iterate the above argument at least $\min\{|x-y|, |y|-N\}$ times where we iterate at most $|y|-N$ times to be able to apply Lemma 5.2.3. In particular, this implies that

$$\begin{aligned}\mathbb{E}_\omega(|\mathcal{G}_\omega(x,y;z)|^s) &\leq (C_G c^{-1}\|\rho\|_\infty^s)^{\min\{|x-y|,|y|-N\}} C_1\|\rho\|_\infty^s \\ &= C_1\|\rho\|_\infty^s \exp(-m\min\{|x-y|, |y|-N\}),\end{aligned}$$

where $e^{-m} = C_G c^{-1}\|\rho\|_\infty^s$. If $\|\rho\|_\infty$ is small enough then $m > 0$. □

Lemma 5.2.6. *Let ρ be as in Lemma 5.2.5 and also of compact support. Then there exists some constant $C_2(C_G)$ and $m > 0$ independent of W such that*

$$|\mathrm{Im}(z)|\,\mathbb{E}_\omega(|\mathcal{G}_\omega(x,y;z)|^2) \leq C_2 \exp(-m\min\{|x-y|\,,|y|-N\}) \tag{5.2.5}$$

for all $z \in \mathbb{C}\backslash\mathbb{R}$ and $x, y \in V(G)$ with $|x|\,, |y| \geq N$.

Proof. The proof is virtually identical to that of the proof of Lemma 3 in [Gra94]. The only difference is that the proof uses the estimate in Lemma 5.2.5 instead of the analogous estimate in [Gra94]. □

Proof of Theorem 5.2.1. Fix W of the form (5.2.1) and let $\delta(C_G) > 0$ be such that if $\|\rho\|_\infty \leq \delta(C_G)$ then Lemma 5.2.6 holds. We also need the constant m to be large enough (by making $\delta(C_G)$ smaller if necessary) such that $e^{-m}C_G < 1$. Suppose that the compact interval I contains the spectrum of H^ω in its interior. Then

$$\epsilon \int_{\mathbb{R}\backslash I} \|P_{|x|\geq R}(H^\omega - E - i\epsilon)^{-1}\psi\|^2 dE \leq \epsilon\|\psi\|^2 \sup_{\lambda\in\mathrm{Sp}(H^\omega)} \int_{\mathbb{R}\backslash I} \|(\lambda - E - i\epsilon)^{-1}\|^2 dE \to 0, \tag{5.2.6}$$

as $\epsilon \downarrow 0$. Since ρ has compact support, $\|H^\omega\|$ is bounded independent of ω almost surely, so we can fix any interval I such that the above holds almost surely. Let $x \in V(G)$ have $|x| \geq N$ then (5.2.6) and (5.2.2) imply that

$$\begin{aligned}\|P_{\mathrm{c}}^{H^\omega}\delta_x\|^2 &= \lim_{R\to\infty}\lim_{\epsilon\downarrow 0}\frac{\epsilon}{\pi}\int_I \|P_{|y|\geq R}(H^\omega - E - i\epsilon)^{-1}\delta_x\|^2 dE \\ &= \lim_{R\to\infty}\lim_{\epsilon\downarrow 0}\frac{\epsilon}{\pi}\int_I \sum_{|y|\geq R} |\mathcal{G}_\omega(x,y;E-i\epsilon)|^2\, dE,\end{aligned}$$

almost surely, where we have used $\mathcal{G}(x,y;z) = \overline{\mathcal{G}(y,x;\overline{z})}$ in the last line. From Fatou's lemma and (5.2.5) we obtain

$$\begin{aligned}\mathbb{E}_\omega(\|P_{\mathrm{c}}^{H^\omega}\delta_x\|^2) &\leq C\liminf_{R\to\infty}\sum_{|y|\geq R}\exp(-m\min\{|x-y|\,,|y|-N\}) \\ &\leq C\exp(m\max\{N,|x|\})\liminf_{R\to\infty}\sum_{j=R}^{\infty}[C_G e^{-m}]^j = 0,\end{aligned}$$

where the second line is obtained by summing over $|y| = j$ and since $e^{-m}C_G < 1$. It follows that $P_{\mathrm{c}}^{H^\omega}\delta_x = 0$ almost surely. But this then implies that $P_{\mathrm{c}}^{H^\omega}$ has finite rank almost surely (recall we assumed $|x| \geq N$) and hence is 0 almost surely since a finite rank self-adjoint operator has pure point spectrum. □

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].*

A similar argument gives the following theorem, where $\mathbb{V}$ is used to denote bounded real-valued potentials on $\mathbb{N}$ and Λ_2 denotes the pointwise evaluations of such potentials.

Theorem 5.3.2 (Impossibility of computing spectral measures with error control)**.** *Consider the problem function*

$$\begin{aligned} \widehat{\Xi} : \mathbb{V} \times \mathbb{N} &\to \mathbb{R}_{\geq 0} \\ (v, j) &\to \langle E^{H_v}_{\{1\}} e_j, e_j \rangle \end{aligned}.$$

Then $\{\widehat{\Xi}, \mathbb{V} \times \mathbb{N}, \Lambda_2\} \in \Delta_2^A$ but $\{\widehat{\Xi}, \mathbb{V} \times \mathbb{N}, \Lambda_2\} \notin \Delta_1^G$. In other words, $\widehat{\Xi}$ can be computed in one limit, but it cannot be computed with error control.

Proof. The positive result $\{\widehat{\Xi}, \mathbb{V} \times \mathbb{N}, \Lambda_2\} \in \Delta_2^A$ follows directly from the remarks after Theorem 4.3.1 and Proposition 4.2.1. Suppose for a contradiction that $\{\widehat{\Xi}, \mathbb{V} \times \mathbb{N}, \Lambda_2\} \in \Delta_1^G$ and that Γ_n is a sequence of general algorithms solving the problem with error control. It follows that for each $j \in \mathbb{N}$, there exists a sequence of general algorithms Γ_n^j such that

$$\lim_{n \to \infty} \Gamma_n^j(v) = \begin{cases} 1, & \text{if } \widehat{\Xi}(v, j) > 0 \\ 0, & \text{otherwise} \end{cases}.$$

Informally, these are described as follows. Fix j and consider the lower bound on $\Xi(v,j)$ computed by $\{\Gamma_m(v,j) : m \leq n\}$. If this is greater than 0 then set $\Gamma_n^j(v) = 1$, otherwise set $\Gamma_n^j(v) = 0$. It follows that $\Gamma_n^j(v)$ also converges from below. It holds that $1 \in \mathrm{Sp}_{\mathrm{p}}(H_v)$ if and only if $\widehat{\Xi}(v,j) > 0$ for some $j \in \mathbb{N}$. Now define

$$\widehat{\Gamma}_n(v) = \sup_{j \leq n} \Gamma_n^j(v).$$

It is clear that this is a general algorithm using Λ_2. Furthermore,

$$\lim_{n\to\infty} \widehat{\Gamma}_n(v) = \begin{cases} 1, & \text{if } 1 \in \mathrm{Sp}_{\mathrm{p}}(H_v) \\ 0, & \text{otherwise} \end{cases},$$

with convergence from below.

Now we may choose a v such that $1 \in \mathrm{Sp}_{\mathrm{p}}(H_v)$ (this can be achieved for example by taking a potential which induces pure point spectrum and shifting the operator accordingly). It follows that for large n we have $\widehat{\Gamma}_n(v) = 1$. But the computation of $\widehat{\Gamma}_n(v)$ is only dependent on $v(j)$ for $j < N$ for some $N \in \mathbb{N}$. Define $v_0 \in \mathbb{V}$ by $v_0(j) = v(j)$ if $j < N$ and $v_0(j) = 0$ otherwise. It follows that $\widehat{\Gamma}_n(v_0) = 1$. But since the potential has compact support, $1 \notin \mathrm{Sp}_{\mathrm{p}}(H_{v_0})$ and hence $\widehat{\Gamma}_n(v_0) = 0$, the required contradiction. □

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 5.3.3. *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 4.3.3 yields a height two arithmetical tower $\widehat{\Gamma}_{n_2,n_1}^j(A)$ for the computation of $\mu_{e_j,e_j,\mathrm{c}}^A((a,b))$. Note that the final limit is from above and using the fact that $\mu_{e_j,e_j,\mathrm{c}}^A(\{a,b\}) = 0$ we obtain a height two tower for $\mu_{e_j,e_j,\mathrm{c}}^A([a,b])$. We can then use the height one tower for $\mu_{e_j,e_j}^A([a,b])$ constructed in §4.2.2, denoted by $\widetilde{\Gamma}_{n_1}^j(A)$, and define

$$a_{j,n_2,n_1}(A) = \widetilde{\Gamma}_{n_1}^j(A) - \widehat{\Gamma}_{n_2,n_1}^j(A).$$

This provides a height two arithmetical tower for $\mu_{e_j,e_j,\mathrm{pp}}^A([a,b])$ 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

$$\lim_{n_2\to\infty} \Upsilon_{n_2}(A) = \sup_{n\in\mathbb{N}} \mu_{e_n,e_n,\mathrm{pp}}^A([a,b]),$$

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 5.3.3 repeatedly. Let $\widehat{\Gamma}_{n_2,n_1}(\cdot, I)$ denote the height two tower constructed in the proof of Lemma 5.3.3 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 $\operatorname{dist}(z_{m_j}, \Xi^{\mathbb{C}}_{\mathrm{pp}}(A)) \geq \delta$ 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 $I_{m_j} \cap \Xi^{\mathbb{C}}_{\mathrm{pp}}(A) \neq \emptyset$ and $\left|I_{m_j}\right| \to 0$. But this contradicts z_{m_j} being positively separated from $\Xi^{\mathbb{C}}_{\mathrm{pp}}(A)$.

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 $I \cap \Xi^{\mathbb{C}}_{\mathrm{pp}}(A) \neq \emptyset$ but $I_1 \cap \Xi^{\mathbb{C}}_{\mathrm{pp}}(A) = I_2 \cap \Xi^{\mathbb{C}}_{\mathrm{pp}}(A) = \emptyset$ 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

$$\min\left\{1, \left|\operatorname{dist}(x, \widetilde{\Gamma}_{n_2}(A)) - \operatorname{dist}(x, \mathrm{Sp}_{\mathrm{pp}}(A))\right|\right\} = 0 \leq \hat{\delta}_m(n_2),$$

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

$$\min\left\{1, \left|\operatorname{dist}(x, \widetilde{\Gamma}_{n_2}(A)) - \operatorname{dist}(x, \mathrm{Sp}_{\mathrm{pp}}(A))\right|\right\} \leq \hat{\delta}_m(n_2),$$

since the final limit of the algorithm from Lemma 5.3.3 is from below. This implies that

$$\begin{aligned}
&\min\left\{1, \sup_{|x| \leq m} \left|\operatorname{dist}(x, \widetilde{\Gamma}_{n_2}(A)) - \operatorname{dist}(x, \mathrm{Sp}_{\mathrm{pp}}(A))\right|\right\} \\
&\leq \min_{m \leq k \leq n_2}\left\{1, \sup_{|x| \leq k} \left|\operatorname{dist}(x, \widetilde{\Gamma}_{n_2}(A)) - \operatorname{dist}(x, \mathrm{Sp}_{\mathrm{pp}}(A))\right|\right\} \leq \delta_m(n_2).
\end{aligned}$$

It follows that we must have

$$d_{\mathrm{AW}}(\widetilde{\Gamma}_{n_2}(A), \mathrm{Sp}_{\mathrm{pp}}(A)) \leq E_{n_2}(A), \tag{5.3.1}$$

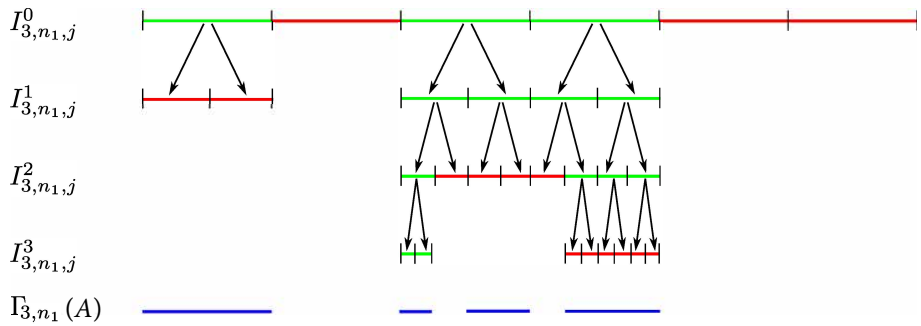


Figure 5.1: Example of tree structure used to compute the point spectrum for $n_2 = 3$. Each tested interval is shown in green ($\widehat{\Gamma}_{n_2,n_1}(A, I) = 1$) or red ($\widehat{\Gamma}_{n_2,n_1}(A, I) = 0$). The arrows show the bisections and the final output is shown in blue.

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, 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_2^A 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 $\lim_{n_1\to\infty} n_2(\epsilon, n_1) = n_2^\epsilon \in \mathbb{N}$ exists and

$$d_{\mathrm{AW}}(\widetilde{\Gamma}_{n_2^\epsilon}(A), \mathrm{Sp}_{\mathrm{pp}}(A)) \leq \epsilon.$$

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

$$n_2^\epsilon = \lim_{n_1\to\infty} n_2(\epsilon, n_1) = \min\{k : k \geq \lceil \epsilon^{-1} \rceil, E_k(A) \leq \epsilon\}.$$

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 5.3.4 (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 $\sum_{j\in\mathbb{N}} g_j^2 < \infty$ then $H_0 + v$ is purely absolutely continuous on $(0, 4)$.

(b) If $\sum_{j\in\mathbb{N}} g_j^2 = \infty$ 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 [Col19b], 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 $\{\Xi^{\mathbb{C}}_{\mathrm{ac}}, \Omega_{f,\alpha}, \Lambda_1\} \notin \Delta^G_2$. 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 $\{\Xi^{\mathbb{C}}_{\mathrm{ac}}, \Omega_{f,0}, \Lambda_1\}$. 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 5.3.4, $[0,4] \subset \mathrm{Sp}_{\mathrm{ac}}(H)$ 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^A_2 tower for $\Xi^{\mathbb{C}}_{\mathrm{ac}}$ we will need the following lemma.

Lemma 5.3.5. *Let $a < b$ with $a, b \in \mathbb{R}$ and consider the decision problem*

$$\Xi_{a,b,\mathrm{ac}} : \Omega_{f,\alpha} \to \{0,1\}$$
$$A \to \begin{cases} 1, & \textit{if } \mathrm{Sp}_{\mathrm{ac}}(A) \cap [a,b] \neq \emptyset \\ 0, & \textit{otherwise}. \end{cases}$$

Then there exists a height two arithmetical tower Γ_{n_2,n_1} (with evaluation functions Λ_1) for $\Xi_{a,b,\mathrm{ac}}$. 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 $\rho^A_{e_m,e_m}(u)$ (the Radon–Nikodym derivative of the absolutely continuous part of the measure $\mu^A_{e_m,e_m}$), it follows by the dominated convergence theorem that

$$\lim_{n_1\to\infty} a_{m,n_2,n_1}(A) =: a_{m,n_2}(A) = \int_a^b \rho^A_{e_m,e_m}(u)\chi_{n_2}(\rho^A_{e_m,e_m}(u))du.$$

We now use the fact that the χ_n are increasing and the dominated convergence theorem again to deduce that

$$\lim_{n_2\to\infty} a_{m,n_2}(A) = \mu^A_{e_m,e_m,\mathrm{ac}}([a, b]),$$

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

$$\Upsilon_{n_2,n_1}(A) = \max_{1\le j\le n_2} \widetilde{a}_{j,n_2,n_1}(A).$$

The proof now follows that of Lemma 5.3.3 exactly. □

Proof that $\{\Xi^{\mathbb{C}}_{\mathrm{ac}}, \Omega_{f,\alpha}, \Lambda_1\} \in \Sigma^A_2$. This is exactly the same construction as in the above proof of the inclusion $\{\Xi^{\mathbb{C}}_{\mathrm{pp}}, \Omega_{f,\alpha}, \Lambda_1\} \in \Sigma^A_2$. We simply replace the tower constructed in the proof of Lemma 5.3.3 by the tower constructed in the proof of Lemma 5.3.5. □

5.3.3 Singular continuous spectra

We will first prove the lower bound for the singular continuous spectrum via Theorem 5.3.4. Note that the impossibility result $\{\Xi^{\mathbb{C}}_{\mathrm{sc}}, \Omega_{f,\alpha}, \Lambda_1\} \notin \Delta^G_2$ follows from the same argument that was used to show $\{\Xi^{\mathbb{C}}_{\mathrm{ac}}, \Omega_{f,\alpha}, \Lambda_1\} \notin \Delta^G_2$. To show that two limits will not suffice for $f(n) - n \ge \sqrt{2n} + 1/2$, our strategy will be to reduce a certain decision problem to the computation of $\Xi^{\mathbb{C}}_{\mathrm{sc}}$. 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 2.4.7 in Chapter 2 §2.4 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 $\{\Xi^{\mathbb{C}}_{\mathrm{sc}}, \Omega_{f,\alpha}, \Lambda_1\} \notin \Delta^G_3$ *if* $f(n) - n \ge \sqrt{2n} + 1/2$. Assume that the function f satisfies $f(n) - n \ge \sqrt{2n} + 1/2$. 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:

$$v_{(j)}(n) = \sum_{k=1}^{\infty} a_{k,j}\delta_{n,k!}.$$

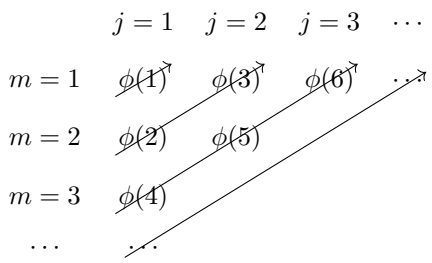
Using Theorem 5.3.4, $[0, 4] \subset \mathrm{Sp}_{\mathrm{sc}}(H_j)$ if $\sum_k a_{k,j} = \infty$ (that is, if the j-th column has infinitely many 1s) and $\mathrm{Sp}_{\mathrm{sc}}(H_j) \cap (0, 4) = \emptyset$ otherwise. Now consider an effective bijection (with effective inverse) between the canonical 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 5.3.6. *Let $a < b$ with $a, b \in \mathbb{R}$ and consider the decision problem*

$$\Xi_{a,b,\mathrm{sc}} : \Omega_{f,\alpha} \to \{0,1\}$$
$$A \to \begin{cases} 1, & \textit{if } \mathrm{Sp}_{\mathrm{sc}}(A) \cap [a,b] \neq \emptyset \\ 0, & \textit{otherwise}. \end{cases}$$

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 5.3.5 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 5.3.6. Note that we can write

$$\mu^A_{e_m,e_m,\mathrm{sc}}([a,b]) = \mu^A_{e_m,e_m}([a,b]) - \mu^A_{e_m,e_m,\mathrm{pp}}([a,b]) - \mu^A_{e_m,e_m,\mathrm{ac}}([a,b]).$$

From this and the proofs of Lemmas 5.3.3 and 5.3.5, it is clear that we can construct a height two arithmetical tower, $a_{m,n_2,n_1}(A)$, for $\mu^A_{e_m,e_m,\mathrm{sc}}([a,b])$ 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 $J^m_1 = [0, 1/m]$, and $J^m_2 = [2/m, \infty)$. Let $k(m,n,n_1) \leq n_1$ be maximal such that $\Upsilon_{m,n,n_1}(A) \in J^m_1 \cup J^m_2$. If no such k exists or $\Upsilon_{m,n,k}(A) \in J^m_1$ 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. □

Part II
Beyond Spectra

Chapter 6

Discrete Spectra and Spectral Gap

Computing the discrete spectrum of a normal operator is a problem encountered in many areas of applied mathematics and theoretical physics, as well as being of interest from a purely theoretical point of view. However, there has been no known algorithm that converges to the discrete spectrum and separates it from the essential spectrum. We provide and numerically demonstrate such an algorithm that yields a sharp classification in the SCI hierarchy. Furthermore, we show how multiplicities (and eigenspaces) can also be calculated. This problem is subtly different to that of computing the point spectrum, namely the eigenvalues of the operator, 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 by Cubitt, Perez–Garcia and Wolf [CPGW15], it was shown that the spectral gap problem is undecidable (i.e. the problem $\notin \Delta_1^T$) 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.

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. An alternative method for computing point spectra (and discrete spectra) is discussed in §4.6.3, where an example for highly oscillatory states of Dirac operators is given.

6.1.1 Computing discrete spectra and multiplicities

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.

We also want to compute multiplicities. Suppose that we have $z_{n_2,n_1} \in \Gamma_{n_2,n_1}(A)$ with

$$\lim_{n_1\to\infty} z_{n_2,n_1} = z_{n_2} = z \in \mathrm{Sp}_d(A)$$

(the limit independent of n_2). Our tower also computes a function $h_{n_2,n_1}(A,\cdot)$ over the output $\Gamma_{n_2,n_1}(A)$ such that

$$\lim_{n_2\to\infty} \lim_{n_1\to\infty} h_{n_2,n_1}(A, z_{n_2,n_1}) = h(A,z)$$

(where $h(A,z)$ denotes the multiplicity of the eigenvalue z) in $\mathbb{Z}_{\geq 0}$ with the discrete metric.

Remark 6.1.2. *Suppose that we equip $\mathbb{N} \cup \{+\infty\}$ with the metric inherited from the natural compactification of $\mathbb{R}_{\geq 0}$. One can alter the proof slightly to show that we can compute h_{n_2,n_1} such that*

$$\lim_{n_1\to\infty} h_{n_2,n_1}(A, z_{n_2,n_1}) =: h_{n_2}(A,z) \geq h(A,z),$$

with convergence monotonic from above, thus generalising the notion of Π_2^A to the multiplicity problem.

An easy corollary of the proof of Theorem 6.1.1 is as follows. 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 $\{0,1\}$)

$$\Xi_2^d : \Omega_{\mathrm{N}}^f, \Omega_{\mathrm{D}} \ni A \mapsto \text{ Is } \mathrm{Sp}_d(A) \neq \emptyset?$$

Corollary 6.1.3. *Let Ξ_2^d, Ω_{N}^f and Ω_{D} be as above. Then,*

$$\Delta_2^G \not\ni \{\Xi_2^d, \Omega_{\mathrm{N}}^f\} \in \Sigma_2^A, \quad \Delta_2^G \not\ni \{\Xi_2^d, \Omega_{\mathrm{D}}\} \in \Sigma_2^A.$$

Finally, we remark that for points approximating the discrete spectrum, the algorithm in §3.4 of Chapter 3 can be used to compute eigenvectors. This is discussed further in §6.4, where we consider numerical examples.

What happens when we cannot bound the dispersion?

Whilst Theorem 6.1.1 shows that computing the discrete spectrum requires two limits, the constructed algorithm has

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

This is demonstrated in §3.4, and we can still effectively approximate eigenspaces with error control. 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 we 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 6.1.4. *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 also shows that without additional structure, it requires three limits to compute the discrete spectrum of self-adjoint matrices. It also requires three limits to check if there are any isolated eigenvalues of finite multiplicity.

6.1.2 The spectral gap problem

The spectral gap problem has a long tradition and is linked to many important conjectures and problems such as the Haldane conjecture [GJL94] or the Yang–Mills mass gap problem in quantum field theory [BCD$^+$06]. The spectral gap question is fundamental in physics, and in the seminal paper by Cubitt, Perez–Garcia and Wolf [CPGW15], it was shown that the spectral gap problem is undecidable when considering the thermodynamic limit of finite-dimensional Hamiltonians.

In this section, we consider the general infinite-dimensional 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 $[a, a+\epsilon] \subset \mathrm{Sp}(A)$.

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 $\{0,1\}$)

$$\Xi_{\mathrm{gap}} : \widehat{\Omega}_{\mathrm{SA}}, \widehat{\Omega}_{\mathrm{D}} \ni A \mapsto \text{ Is the spectrum of } A \text{ gapped?} \tag{6.1.1}$$

Theorem 6.1.5 (Spectral gap)**.** *Let Ξ_{gap} be as in (6.1.1) and $\widehat{\Omega}_{\mathrm{SA}}, \widehat{\Omega}_{\mathrm{D}}$ as above. Then*

$$\Delta_2^G \not\ni \{\Xi_{\mathrm{gap}}, \widehat{\Omega}_{\mathrm{SA}}\} \in \Sigma_2^A, \quad \Delta_2^G \not\ni \{\Xi_{\mathrm{gap}}, \widehat{\Omega}_{\mathrm{D}}\} \in \Sigma_2^A.$$

Remark 6.1.6 (Diagonal vs. full matrix)**.** *It is worth noting that Theorem 6.1.5 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 $\widetilde{\Omega}^f_{\mathrm{SA}}$ 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 $\mathrm{Sp}(A)$.

We consider the classification problem Ξ_{class} which maps $\widetilde{\Omega}^f_{\mathrm{SA}}$ (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 $\widetilde{\Omega}^f_{\mathrm{SA}}$.

Theorem 6.1.7 (Spectral Classification)**.** *Let Ξ_{class}, $\widetilde{\Omega}^f_{\mathrm{SA}}$ and $\widetilde{\Omega}_{\mathrm{D}}$ be as above. Then*

$$\Delta_2^G \not\ni \{\Xi_{\mathrm{class}}, \widetilde{\Omega}^f_{\mathrm{SA}}\} \in \Pi_2^A, \quad \Delta_2^G \not\ni \{\Xi_{\mathrm{class}}, \widetilde{\Omega}_{\mathrm{D}}\} \in \Pi_2^A.$$

6.2 Proofs of Theorems on Discrete Spectra

Here we prove our results related to the discrete spectrum. We need some results on finite section approximations to the discrete spectrum of a Hermitian operator below the essential spectrum. There are two cases to consider; either there are infinitely many eigenvalues below the essential spectrum, or there are only finitely many. 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 $B_n = P_n B P_n \in \mathbb{C}^n$ 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 $j \leq m$, $\mu_j^n \downarrow \lambda_j$ 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 $n \geq N$, $\mu_k^n \leq a + \epsilon$.*

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+19]. For completeness, we write out the algorithm here.[1] Let P_n be the usual projection onto the first n basis elements and set $Q_n = I - P_n$. Define

$$\begin{aligned}
\mu_{m,n}(A) &:= \min\{\sigma_1(P_{f(n)}(A - zI)|_{Q_m P_n(l^2(\mathbb{N}))}), \sigma_1(P_{f(n)}(A - zI)^*|_{Q_m P_n(l^2(\mathbb{N}))})\},\\
G_n &:= \min\left\{\frac{s + it}{2^n} : s, t \in \{-2^{2n}, ..., 2^{2n}\}\right\},\\
\Upsilon_m(z) &:= z + \{w \in \mathbb{C} : |\mathrm{Re}(w)|, |\mathrm{Im}(w)| \leq 2^{-(m+1)}\}.
\end{aligned}$$

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

$$\zeta_{n_2,n_1}(A) = \{z \in \hat{\Gamma}_{n_1}(A) : E(n_1, z) < \mathrm{dist}(z, \tilde{\Gamma}_{n_2,n_1}(A) + B_{1/n_2}(0))\}.$$

[1] The actual algorithm is slightly more complicated to avoid the empty set, but its listed properties still hold.

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 $\operatorname{dist}(z, \tilde{\Gamma}_{n_2,n_1}(A))^2$ 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 $\operatorname{dist}(z, \tilde{\Gamma}_{n_2,n_1}(A))^2$. 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: $\operatorname{Sp}_d(A) \cap (\tilde{\Gamma}_{n_2}(A) + B_{1/n_2}(0))^c = \emptyset$. 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 $E(n_1, z) \geq \operatorname{dist}(z, \tilde{\Gamma}_{n_2,n_1}(A))$ for all $z \in \hat{\Gamma}_{n_1}(A)$, that $\zeta_{n_2,n_1}(A) = \emptyset$.

Case 2: $\operatorname{Sp}_d(A) \cap (\tilde{\Gamma}_{n_2}(A) + B_{1/n_2}(0))^c \neq \emptyset$. In this case, this set is a finite subset of $\operatorname{Sp}_d(A)$, $\{\hat{z}_1, ..., \hat{z}_{m(n_2)}\}$, 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 $\operatorname{Sp}(A)$ and $E(n, z) \geq \operatorname{dist}(z, \operatorname{Sp}(A))$, 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 $\operatorname{Sp}_d(A) \cap (\tilde{\Gamma}_{n_2}(A) + B_{1/n_2}(0))^c \neq \emptyset$ 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 $\operatorname{Sp}(A)$, 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 $\operatorname{Sp}_d(A) \cap (\tilde{\Gamma}_{n_2}(A) + B_{1/n_2}(0))^c \neq \emptyset$ 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 $z \in \operatorname{Sp}_d(A)$ and $\epsilon > 0$ with $\operatorname{Sp}(A) \cap B_{2\epsilon}(z) = \{z\}$. 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 $\operatorname{Sp}_d(A) \cap (\tilde{\Gamma}_N(A) + B_{1/N}(0))^c \neq \emptyset$ (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.

Step 3: **Multiplicities.** Suppose that $z_{n_2,n_1} \in \Gamma_{n_2,n_1}(A)$ converges as $n_1 \to \infty$ to some $z_{n_2} = z \in \Gamma_{n_2}(A) \subset \mathrm{Sp}_d(A)$, where Γ_{n_2} is the first limit of the height two tower constructed in step 2. Consider the following operator, viewed as a finite matrix acting on $\mathbb{C}^n$, $A_n = P_n(A - zI)^*(A - zI)P_n$. This is a truncation of the operator $(A - zI)^*(A - zI)$. The key observation is that 0 lies in the discrete spectrum of $(A - zI)^*(A - zI)$ with $h((A - zI)^*(A - zI), 0) = h(A, z)$, the multiplicity of the eigenvalue z. To see this, note that $\ker(A - zI) = \ker((A - zI)^*(A - zI))$ and that if $\|x\| = 1$ then

$$\|(A - zI)x\| \leq \sqrt{\|(A - zI)^*(A - zI)x\|}.$$

Since $(A - zI)$ is bounded below on $\ker(A - zI)^\perp$, the same must be true for $(A - zI)^*(A - zI)$. Now set

$$\begin{aligned} &h_{n_2,n_1}(A, z_{n_2,n_1}) \\ &= \min\{n_2, \left|\{w \in \mathrm{Sp}(P_{n_1}(A - z_{n_2,n_1}I)^* P_{f(n_1)}(A - z_{n_2,n_1}I)P_{n_1}) : |w| < 1/n_2 - d_{n_1}\}\right|\}, \end{aligned}$$

where d_{n_1} is some non-negative sequence converging to 0 that we define below. As usual we consider the relevant operator as a matrix acting on $\mathbb{C}^{n_1}$ and we count eigenvalues according to their multiplicity. Via shifting by $(1/n_2 - d_{n_1})I$ and assuming d_{n_1} can be computed with finitely many arithmetic operations and comparisons, Lemma 3.2.8 shows that this is a general algorithm and can be computed with finitely many arithmetic operations and comparisons. Consider the similar function (that we cannot necessarily compute since we do not know z),

$$q_{n_2,n_1}(A, z) = \min\{n_2, |\{w \in \mathrm{Sp}(A_{n_1}) : |w| < 1/n_2\}|\},$$

where

$$A_{n_1} = P_{n_1}(A - zI)^*(A - zI)P_{n_1}.$$

We set $B = (A - zI)^*(A - zI)$ and list $\lambda_1 \leq \lambda_2 \leq ...$ as in Lemmas 6.2.1 and 6.2.2, then

$$\lim_{n_1\to\infty} q_{n_2,n_1}(A, z) = \min\{n_2, |\lambda_j : \lambda_j < 1/n_2|\}.$$

It is then clear from the same lemmas that

$$\lim_{n_2\to\infty} \lim_{n_1\to\infty} q_{n_2,n_1}(A, z) = h((A - zI)^*(A - zI), 0) = h(A, z).$$

We will have completed the proof if we can choose d_{n_1} such that

$$\lim_{n_1\to\infty} |h_{n_2,n_1}(A, z_{n_2,n_1}) - q_{n_2,n_1}(A, z)| = 0.$$

It is straightforward to show that

$$\begin{aligned}
&\left\|A_{n_1} - P_{n_1}(A - z_{n_2,n_1}I)^* P_{f(n_1)}(A - z_{n_2,n_1}I)P_{n_1}\right\| \\
&\quad \leq \big(|z - z_{n_2,n_1}| + c_{n_1}\big)\big(\|AP_{n_1}\| + |z - z_{n_2,n_1}| + |z_{n_2,n_1}| + \left\|P_{f(n_1)}(A - z_{n_2,n_1}I)P_{n_1}\right\|\big) \\
&\quad \leq \big(|z - z_{n_2,n_1}| + c_{n_1}\big)\big(2\left\|P_{f(n_1)}AP_{n_1}\right\| + |z - z_{n_2,n_1}| + 2\,|z_{n_2,n_1}| + c_{n_1}\big),
\end{aligned}$$

where $D_{f,m}(A) \leq c_m$ is the dispersion bound. Choose

$$d_{n_1} = \big(E(n_1, z_{n_2,n_1}) + c_{n_1}\big)\big(E(n_1, z_{n_2,n_1}) + 2\,|z_{n_2,n_1}| + 2k_{n_1} + c_{n_1}\big),$$

where k_{n_1} overestimates $\left\|P_{f(n_1)}AP_{n_1}\right\|$ by at most 1. k_{n_1} can be computed using a similar positive definiteness test as in `DistSpec` (see §3.5.1). Since z_{n_2,n_1} converges to $z \in \mathrm{Sp}_d(A)$, it is clear that

$$\left\|A_{n_1} - P_{n_1}(A - z_{n_2,n_1}I)^* P_{f(n_1)}(A - z_{n_2,n_1}I)P_{n_1}\right\| \leq d_{n_1}$$

eventually and that d_{n_1} converges to 0. Weyl's inequality for eigenvalue perturbations of Hermitian matrices implies the needed convergence. □

Proof of Corollary 6.1.3. 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 6.1.4. **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 §2.4 which was proven in Theorem 2.4.7 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 $\{\Xi_1^d, \Omega_1^d\}, \Gamma_{n_2,n_1}$, 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 $l^j_{M_j} = \infty$ 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$^+$19] 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$^+$19] 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 $z \in \hat{\Gamma}_{n_2}(A)$, $\operatorname{dist}(z, \operatorname{Sp}(A)) \leq 2^{-n_2}$.

Using these, we initially define

$$\zeta_{n_3,n_2,n_1}(A) = \{z \in \hat{\Gamma}_{n_2,n_1}(A) : 2^{-n_3} - 2^{-n_2} \leq \operatorname{dist}(z, \tilde{\Gamma}_{n_3,n_2,n_1}(A))\}.$$

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

$$\zeta_{n_3,n_2,n_1}(A) = \{z \in \hat{\Gamma}_{n_2}(A) : 2^{-n_3} - 2^{-n_2} \leq \operatorname{dist}(z, \tilde{\Gamma}_{n_3,n_2}(A))\} =: \zeta_{n_3,n_2}(A).$$

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: $\operatorname{Sp}_d(A) \cap (\tilde{\Gamma}_{n_3}(A) + D_{2^{-n_3}}(0))^c = \emptyset$. 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, $z_{m_j} \to z \in \operatorname{Sp}(A)$. We also have that

$$\operatorname{dist}(z_{m_j}, \tilde{\Gamma}_{n_3,m_j}(A)) \geq 2^{-n_3} - 2^{-m_j},$$

which implies that $\operatorname{dist}(z, \tilde{\Gamma}_{n_3}(A)) \geq 2^{-n_3}$ and hence $z \in \operatorname{Sp}_d(A) \cap (\tilde{\Gamma}_{n_3}(A) + D_{2^{-n_3}}(0))^c$, the required contradiction. It follows that $\zeta_{n_3,n_2}(A)$ is empty for large n_2.

Case 2: $\operatorname{Sp}_d(A) \cap (\tilde{\Gamma}_{n_3}(A) + D_{2^{-n_3}}(0))^c \neq \emptyset$. In this case, this set is a finite subset of $\operatorname{Sp}_d(A)$, $\{\hat{z}_1, ..., \hat{z}_{m(n_3)}\}$. 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

$$\begin{aligned}\operatorname{dist}(z_{n_2}, \tilde{\Gamma}_{n_3,n_2}(A)) &\geq \operatorname{dist}(z_{n_2}, \tilde{\Gamma}_{n_3}(A)) \\ &\geq \operatorname{dist}(\hat{z}_1, \tilde{\Gamma}_{n_3}(A)) - 2^{-n_2} \geq 2^{-n_3} - 2^{-n_2},\end{aligned}$$

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 $\operatorname{Sp}_d(A) \cap (\tilde{\Gamma}_{n_3}(A) + D_{2^{-n_3}}(0))^c$. It follows that in this case $\zeta_{n_3,n_2}(A)$ converges to $\operatorname{Sp}_d(A) \cap (\tilde{\Gamma}_{n_3}(A) + B_{1/n_3}(0))^c \neq \emptyset$ in the Hausdorff metric as $n_2 \to \infty$.

Let $N(A) \in \mathbb{N}$ be minimal such that $\operatorname{Sp}_d(A) \cap (\tilde{\Gamma}_N(A) + D_{2^{-N}}(0))^c \neq \emptyset$ (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 6.1.5. **Step 1**: $\{\Xi_{\mathrm{gap}}, \widehat{\Omega}_{\mathrm{SA}}\} \in \Sigma_2^A$. Let $A \in \widehat{\Omega}_{\mathrm{SA}}$. Using Corollary 3.2.9 we can compute all n eigenvalues of P_nAP_n to arbitrary precision in finitely many arithmetic operations and comparisons. Note that it is not completely straightforward to deduce this with the QR algorithm, as one has to deal with halting criteria in order to achieve the correct precision. Moreover, one must approximate roots in order to extract the approximate eigenvalues from a potential 2×2 matrix block. Thus we use Corollary 3.2.9 instead. 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 $|\epsilon_n| \leq 1/n$ using Corollary 3.2.9 with $B = P_nAP_n$. 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 $J_{n_2}^1$ and $J_{n_2}^2$, 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, $l_{n_1} \in J_{n_2}^2$. 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_2^A tower.

Step 2: $\{\Xi_{\mathrm{gap}}, \widehat{\Omega}_{\mathrm{D}}\} \notin \Delta_2^G$. 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 $i,j \leq N(A,n)$ 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

$$\zeta_p := \min\{d_r : 1 \leq r \leq k_p\},$$

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 $n_m \geq m$ such that if $n \geq n_m$ then

$$\Gamma_n(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C) = 1.$$

Now let $k_{m+1} \geq \max\{N(B_{k_1} \oplus A_{k_1} \oplus ... \oplus B_{k_m} \oplus A_{k_m} \oplus C, n_m), k_m + 1\}$ 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 6.1.7. 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 6.1.5 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 $B_n = P_n A P_n$ and in the notation of Lemmas 6.2.2 and 6.2.1, let

$$0 \leq l^j_n := \mu^n_{j+1} - \mu^n_1 + \epsilon^j_n, \text{ for } j < n.$$

where we again have computed $\mu^n_{j+1} - \mu^n_1$ to accuracy $\left|\epsilon^j_n\right| \leq 1/n$ using only finitely many arithmetic operations and comparisons by Corollary 3.2.9. $\Xi_{\mathrm{class}}(A) = 1$ 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 $A \in \widetilde{\Omega}^f_{\mathrm{SA}}$ 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 6.1.5. 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 $\Xi_{\mathrm{class}}(A) \neq 4$. Finally, note that if $\Xi_{\mathrm{class}}(A) = 4$ but there exists n_2 with $\Gamma_{n_2}(A) \neq 4$ then the above implies the contradiction $\Xi_{\mathrm{class}}(A) \neq 4$. The partial converses proven above imply Γ_{n_2,n_1} realises the Π^A_2 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 $\lim_{n_1\to\infty}\Gamma_{n_2,n_1}(A)$ 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 routine `DistSpec` (see §3.5.1) 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 $l^2(\mathbb{Z})$, related to a wealth of mathematical and physical problems such as the Ten Martini Problem (see [Jit99, Dam09, AJ09]), given by

$$(H_\alpha x)_n = x_{n-1} + x_{n+1} + 2\lambda\cos(2\pi n\alpha + \nu)x_n,$$

where we set $\lambda = 1$ (critical coupling). The choice of $\lambda = 1$ was studied in Hofstadter's classic paper [Hof76] on what has become known as the Hofstadter butterfly (union of the spectra over ν as α varies). In this case, 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 (Ten Martini Problem) and does not depend on ν. Hence it follows that there is no discrete spectrum. In general, we cannot hope to work with infinite precision, and so will have to 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.

The error bounds provided by `DistSpec` (applied to the output of `DiscreteSpec`) are shown in Figure 6.2 for a representative selection of eigenvalues. We have estimated the true error by taking n_1 large

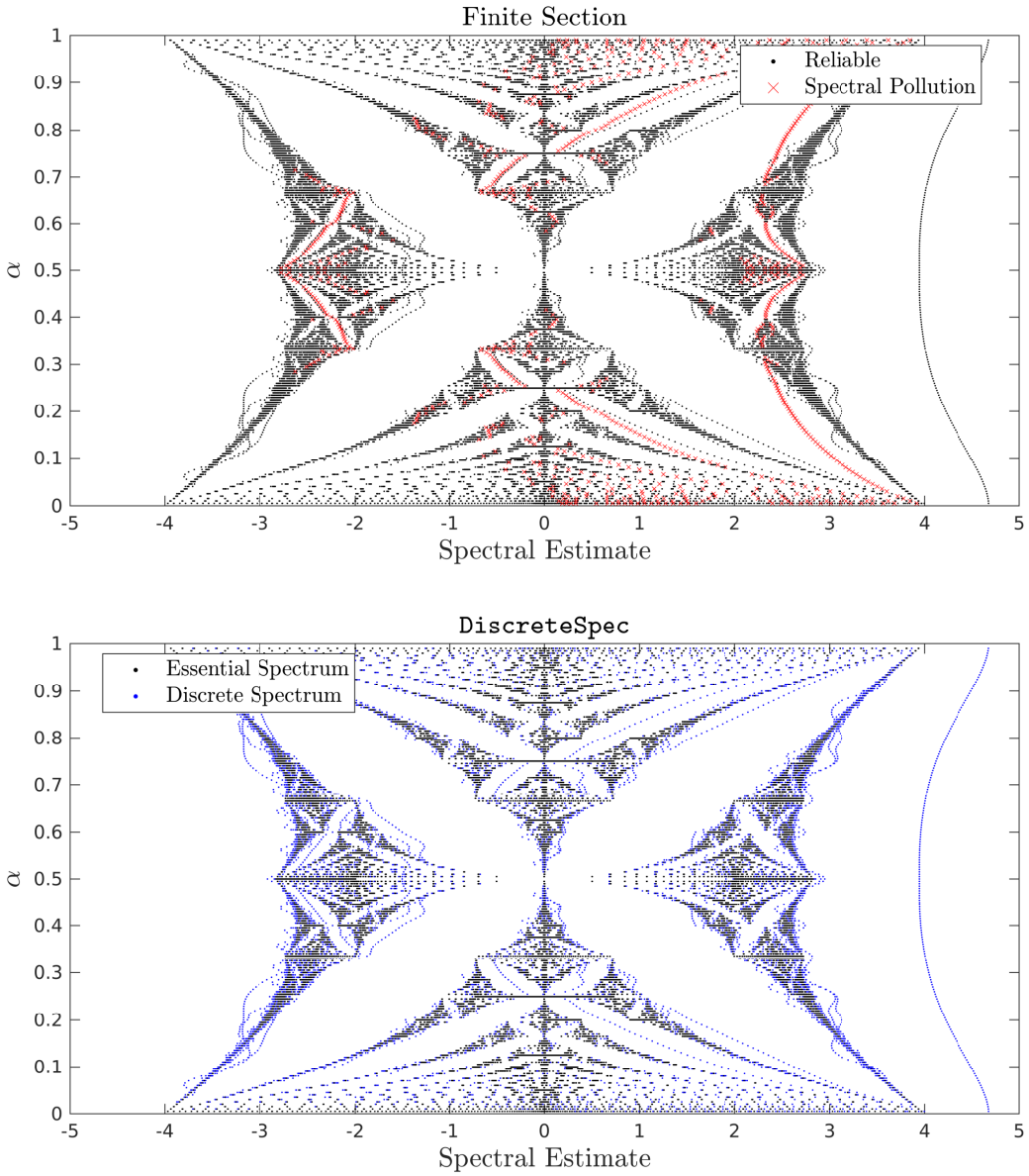


Figure 6.1: Top: Output of finite section. Spectral pollution detected by the algorithm of Chapter 3 is shown as red crosses. Bottom: Output of `DiscreteSpec` and the splitting into the essential spectrum and the discrete spectrum. The output captures the discrete spectrum down to a distance ≈ 0.01 away from the essential spectrum, which can be made smaller for larger n_2.

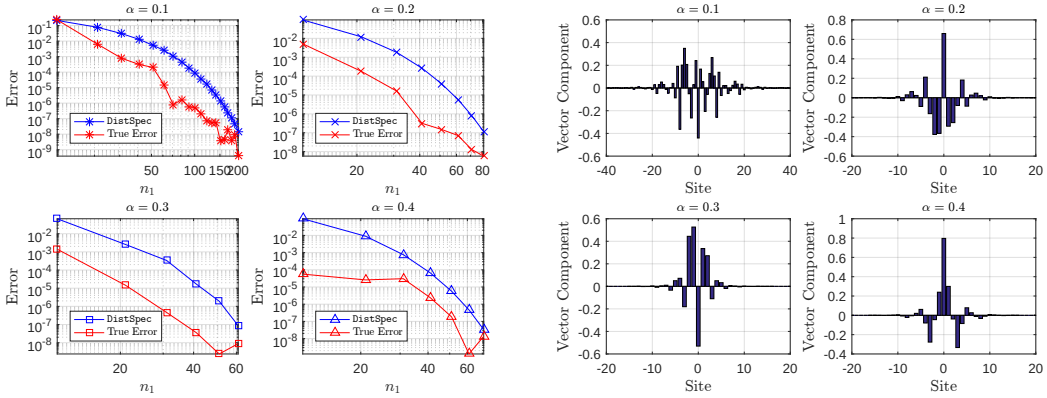


Figure 6.2: Left: Errors for approximating a typical eigenvalue over a range of α. Note that `DistSpec` always overestimates the true error and that we quickly obtain machine precision $\sqrt{\epsilon_{\text{mach}}}$. Right: The approximation of the eigenvectors (truncated for plot) using 201 basis sites.

and have also shown the estimates produced by `DistSpec`. As expected, the routine `DistSpec` gives an upper bound on the true error, which converges to zero. It is clear that only a small number of matrix values are required to gain high precision in this example.[2] The error bounds can also be translated into computing approximates of the eigenvectors of an operator A, corresponding to the discrete spectrum, with an error bound in the following manner. The algorithm in §3.4 computes a vector x_{n_1} of norm ≈ 1 such that (in this case taking $\delta \downarrow 0$, $c_n = 0$)

$$\|(A - z_{n_1} I)x_{n_1}\| \leq \texttt{DistSpec}(A, n_1, f(n_1), z_{n_1}).$$

We write

$$x_{n_1} = x^d_{n_1} + y_{n_1},$$

where $x^d_{n_1}$ is an eigenvector of A with eigenvalue z, y_{n_1} is perpendicular to the eigenspace associated with z and $z_{n_1} \to z$. It follows that

$$\|(A - zI)y_{n_1}\| \leq |z - z_{n_1}| + \texttt{DistSpec}(A, n_1, f(n_1), z_{n_1}) \leq 2 \times \texttt{DistSpec}(A, n_1, f(n_1), z_{n_1}),$$

for large n_1. But $A - zI$ is bounded below on the orthogonal complement of the eigenspace, with lower bound $\text{dist}(z, \text{Sp}(A)\backslash\{z\})$. Hence,

$$\|y_{n_1}\| \leq \frac{2 \times \texttt{DistSpec}(A, n_1, f(n_1), z_{n_1})}{\text{dist}(z, \text{Sp}(A)\backslash\{z\})}$$

for large n_1. This also bounds the l^2 distance of x_{n_1} to the eigenspace and can be estimated by approximating the spectrum of A. We have shown the value of vector components of approximate eigenvalues in Figure 6.2 for $n_1 = 201$ (corresponding to sites $n = -100, ..., 100$). It is also straightforward to adjust this procedure to eigenvalues of multiplicity greater than 1 and approximate the whole eigenspace. We note that for this example, all eigenvalues were found to have multiplicity 1 as expected for a random perturbation. Finally, the method of computing eigenvectors and error bounds can also be used for the unbounded case when z lies in the discrete spectrum.

[2] For the particular implementation, the method of using Cholesky decompositions to test for positive definiteness means that we cannot expect precision greater than $\sqrt{\epsilon_{\text{mach}}}$. For these tests this corresponds to approximately 10^{-8}, although we have tested the method using higher precision arithmetic and found the error plots to be similar, decreasing to the corresponding $\sqrt{\epsilon_{\text{mach}}}$. One can also gain higher accuracy by using iterative methods to approximate the smallest singular value of the rectangular truncations.

Chapter 7

Geometric Features and Detecting Finite Section Failure

In this chapter and the next, 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 detecting the failure of finite section (computing an error flag) is harder than computing the spectrum itself (the problem solved in Chapter 3). This also settles the open 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?

Remark 7.1.1. *An effective way to avoid spectral pollution is discussed in §4.6.3, where we compute highly oscillatory bound states of the Dirac operator. The algorithms in Chapter 3 converge to the spectrum for a large general class of operators, whilst avoiding spectral pollution. They also provide quantitative estimates through Σ_1^A error control.*

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 7.1.2 ([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 7.1.3 ([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. As discussed in the first footnote of §4.1, some bases may be preferable to others, and one can view a different choice of basis as changing the evaluations functions Λ defined below. The classes of operators we discuss are basis independent, apart from Ω_f and Ω_{D} (see definitions below).

[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 7.1.2 says that, up to a different choice of projections, this can be avoided on $W_e(A)$.

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 have included Λ_2 in these two chapters since it is natural for problems posed in variational form.

The proofs of lower bounds in this chapter and Chapter 8 make clear that all results still hold if we replace the respective sub-class $\Omega \subset \mathcal{B}(l^2(\mathbb{N})) =: \Omega_{\mathrm{B}}$ by the restriction to operators in Ω having operator norm at most $M \in \mathbb{R}_{>0}$, adding such a value M (constant function) to the evaluation set Λ. 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. As usual, our results extend to general separable Hilbert spaces $\mathcal{H}$ once one is given an orthonormal basis $\{e_1, e_2, ...\}$ and matrix values of the operators with respect to this basis. This allows computations with operators naturally defined on lattices such as $\mathbb{Z}^d$ or, more generally, on graphs. Such operators are abundant in mathematical physics.

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 (recall this notion from §3.1.1). However, sometimes the sequence $\{c_n\}$ is not needed and we will explicitly mention when this is the case. As a special case, if we know our matrix is sparse with finitely many non-zero entries in each column and row (and we know their positions) then we know an f with $c_n = 0$ and clearly Λ_1 and Λ_2 are equivalent. We can then compute matrix elements of products of A and A^* using finitely many arithmetic operations. In the more general non-sparse case, f and $\{c_n\}$ can be used to compute matrix elements of products A and A^* with error control. Conversely, given Λ_2 we can compute an f and $\{c_n\}$ simply by considering norms of truncated rows and columns. Hence knowledge of f (with or without $\{c_n\}$) and use of evaluation functions in Λ_2 are subtly different. 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:*

$$\Delta_1^G \not\ni \{\Xi_r, \Omega_{\mathrm{D}}, \Lambda_1\} \in \Sigma_1^A, \qquad \Delta_1^G \not\ni \{\Xi_r, \Omega_{\mathrm{N}}, \Lambda_1\} \in \Sigma_1^A, \qquad \Delta_1^G \not\ni \{\Xi_r, \Omega_f \cap \Omega_g, \Lambda_1\} \in \Sigma_1^A,$$
$$\Delta_2^G \not\ni \{\Xi_r, \Omega_g, \Lambda_1\} \in \Sigma_2^A, \qquad \Delta_2^G \not\ni \{\Xi_r, \Omega_f, \Lambda_1\} \in \Pi_2^A, \qquad \Delta_3^G \not\ni \{\Xi_r, \Omega_{\mathrm{B}}, \Lambda_1\} \in \Pi_3^A.$$

When considering the evaluation set Λ_2, the only changes are the following classifications:

$$\Delta_1^G \not\ni \{\Xi_r, \Omega_g, \Lambda_2\} \in \Sigma_1^A, \qquad \Delta_2^G \not\ni \{\Xi_r, \Omega_{\mathrm{B}}, \Lambda_2\} \in \Pi_2^A.$$

Remark 7.3.2. *The Π_2^A algorithm for $\{\Xi_r, \Omega_f\}$ does not need a sequence $\{c_n\}$ (converging to zero) that bounds the dispersion, $D_{f,n}(A) \leq c_n$, to be a SCI-sharp algorithm since this is absorbed in the first limit.*

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 7.3.3. *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 7.3.4. *We have the following classifications for $i = 1, 2$ and $\hat{\Omega} = \Omega_{\mathrm{D}}, \Omega_f$:*

$$\Delta_1^G \ni \{\Xi_{r,p}, \hat{\Omega}, \Lambda_i\} \in \Sigma_1^A, \qquad \Delta_2^G \ni \{\Xi_{cap}, \hat{\Omega}, \Lambda_i\} \in \Pi_2^A.$$

Whereas for $\tilde{\Omega} = \Omega_{\mathrm{N}}, \Omega_g$ or Ω_{B}:

$$\Delta_2^G \ni \{\Xi_{r,p}, \tilde{\Omega}, \Lambda_1\} \in \Sigma_2^A, \qquad \Delta_3^G \ni \{\Xi_{cap}, \tilde{\Omega}, \Lambda_1\} \in \Pi_3^A$$
$$\Delta_1^G \ni \{\Xi_{r,p}, \tilde{\Omega}, \Lambda_2\} \in \Sigma_1^A, \qquad \Delta_2^G \ni \{\Xi_{cap}, \tilde{\Omega}, \Lambda_2\} \in \Pi_2^A.$$

Remark 7.3.5. *Note here that we do not use the assumption $g(x) \leq (1-\delta)x$. We also fix the polynomial p for the strongest possible negative results. However, the existence of towers of algorithms also holds when considering the polynomial p itself as an input. Finally, the proof shows the same classifications for the class of bounded self-adjoint operators as Ω_{N} for these problems.*

Remark 7.3.6. *A natural way of computing the spectral radius is through Gelfand's formula and $\Xi_{r,p}$. The extra limit incurred compared to the computation of $\Xi_{r,p}$ for Ω_f and Ω_{B} is due to the need to consider the family of polynomials x^n with $n \to \infty$.*

Remark 7.3.7. *Somewhat surprising is the result that the computation of $\|p(A)\|$ requires two limits for normal operators. The proof makes clear that one reason for this is spectral pollution associated with finite section methods. This also shows that computing the capacity from first approximating the spectrum via finite sections, computing the nth diameters of those approximations and letting $n \to \infty$ will not converge in general.*

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!

Denote the problem function $W_e(A)$ by Ξ_{we}. 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 then add pointwise evaluations of

[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. [MÖ92]) it follows that generically the asymptotic speeds are the same.

the relevant sequences $\{(a_m, b_m)\}$ or $\{(r_m, z_m)\}$ to Λ_i. Let Ω_{SA} denote the class of self-adjoint operators in Ω_{B}.

Theorem 7.3.8. *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 7.3.9. *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 $\tilde{\Gamma}_{n_k,\ldots,n_1}$ 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 $n_1, \ldots, n_k$. It is then clear that

$$\Gamma_{n_k,\ldots,n_1}(A) = \sup_{z \in \tilde{\Gamma}_{n_k,\ldots,n_1}(A)} |z| + \frac{1}{2^{n_k}}$$

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 $\tilde{\Gamma}_{n_k,\ldots,n_1}$ provides a Σ_k^A tower of algorithms for the spectrum (output a finite set for each $n_1, \ldots, n_k$),

$$\Gamma_{n_k,\ldots,n_1}(A) = \sup_{z \in \tilde{\Gamma}_{n_k,\ldots,n_1}(A)} |z| - \frac{1}{2^{n_k}}$$

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+19]. 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 $P_n A P_n \in \mathbb{C}^{n \times n}$. It is straightforward to show that

$$\|P_n A P_n\| \uparrow \|A\| \quad \text{as } n \to \infty.$$

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 $1/n$ using Corollary 3.2.9 in Chapter 3 which then yields a Σ_1^A algorithm for $\{\Xi_r, \Omega_{\rm N}\}$.

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

$$B_k = \left(C_{l_1} \oplus A_{l_1}\right) \oplus ... \oplus \left(C_{l_k} \oplus A_{l_k}\right) \oplus C, \quad C = \text{diag}\{0, 0, ...\}.$$

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 (note in particular this is not discrete so we use Remark 2.4.8), 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 2.4.7 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. $\square$

Remark 7.4.1. *The algorithm in step 1 of the above proof will work for all operators whose operator norm is equal to the spectral radius. If, instead, the operator is spectraloid, meaning the spectral radius is equal to the numerical radius*

$$w(A) := \sup\{|\langle Ax, x\rangle| : \|x\| = 1\},$$

then a similar argument will hold by estimating $w(P_nAP_n)$. To do this, we need a way of computing $w(A)$ to a given accuracy using finitely many arithmetic operations and comparisons on its matrix values. This is given by Lemma 7.5.1 below.

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 can be proven using exactly the same example and a similar argument to step 3 of the proof of Theorem 7.3.1 (hence omitted). □

Proof of Theorem 7.3.3. 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$^+$19]. 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$^+$19]. Finally, we can use Λ_2 to collapse the first limit of the algorithm for the essential spectrum in [BACH$^+$19], 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 7.3.8 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 7.3.4. 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

$$\Gamma_{n_{k+1},\dots,n_1}(A) = \min_{1\leq m\leq n_{k+1}} \tilde{\Gamma}_{n_k,\dots,n_1}(A, p_m).$$

The fact that this is a convergent Π^A_{k+1} 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 $\{\Xi_{r,p}, \Omega_f, \Lambda_1\} \in \Sigma^A_1$, $\{\Xi_{r,p}, \Omega_{\mathrm{B}}, \Lambda_1\} \in \Sigma^A_2$ and $\{\Xi_{r,p}, \Omega_{\mathrm{B}}, \Lambda_2\} \in \Sigma^A_1$. Likewise, for the negative results we only need to prove $\{\Xi_{cap}, \Omega_{\mathrm{D}}, \Lambda_2\} \notin \Delta^G_2$ (the fact that $\{\Xi_{r,p}, \Omega_{\mathrm{D}}, \Lambda_2\} \notin \Delta^G_1$ is obvious), $\{\Xi_{cap}, \Omega_{\mathrm{N}}, \Lambda_1\} \notin \Delta^G_3$ and $\{\Xi_{r,p}, \Omega_{\mathrm{N}}, \Lambda_2\} \notin \Delta^G_2$. We shall prove these results with Ω_{N} replaced by the class of self-adjoint bounded operators denoted by Ω_{SA}.

Remark 7.4.2 (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 8.1.9 (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^A_2 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: $\{\Xi_{r,p}, \Omega_f, \Lambda_1\} \in \Sigma^A_1$. 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: $\{\Xi_{r,p}, \Omega_{\mathrm{B}}, \Lambda_1\} \in \Sigma^A_2$ and $\{\Xi_{r,p}, \Omega_{\mathrm{B}}, \Lambda_2\} \in \Sigma^A_1$. 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 $\{\Xi_{r,p}, \Omega_{\mathrm{B}}, \Lambda_2\} \in \Sigma^A_1$, 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: $\{\Xi_{cap}, \Omega_{\mathrm{SA}}, \Lambda_1\} \notin \Delta^G_3$. 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 2.4.7 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 7.4.3. *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*

$$B(z_1,...,z_k) = \left(\begin{array}{cccc|cccc} z_1 & 0 & \cdots & & a_1 & 0 & \cdots & \\ 0 & z_2 & 0 & \cdots & 0 & a_2 & 0 & \cdots \\ \vdots & 0 & \ddots & & \vdots & 0 & \ddots & \\ & \vdots & & & & \vdots & & \\ & & & z_k & & & & a_k \\ \hline a_1 & 0 & \cdots & & -z_1 & 0 & \cdots & \\ 0 & a_2 & 0 & \cdots & 0 & -z_2 & 0 & \cdots \\ \vdots & 0 & \ddots & & \vdots & 0 & \ddots & \\ & \vdots & & & & \vdots & & \\ & & & a_k & & & & -z_k \end{array}\right) \in \mathbb{C}^{2k\times 2k}$$

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 7.4.3. 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 7.3.8

Proof of Theorem 7.3.8 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 7.3.3. 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 7.3.8 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

$$\gamma_{n_2,n_1}(z;A)=\min\{\sigma_1(P_{n_1}(A-zI)|_{P_{n_2}\mathcal{H}}),\sigma_1(P_{n_1}(A^*-\bar{z}I)|_{P_{n_2}\mathcal{H}})\}$$

and note that this can be approximated to any given accuracy in finitely many arithmetic operations and comparisons (see §3.2, in particular Corollary 3.2.9). 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

$$\operatorname{dist}(z, \operatorname{cl}(U)) \leq \operatorname{dist}(z, V_{n_1}) < 1/n_2 - 1/n_1,$$

it follows that $\operatorname{dist}(\hat{z}, \operatorname{cl}(U)) < 1/n_2$ 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 $\operatorname{dist}(z, \operatorname{cl}(U)) < 1/n_2$, then there exists $\epsilon > 0$ and $j > 0$ such that $\operatorname{dist}(z, U_j) < 1/n_2 - \epsilon$. There also exists $z_{n_1} \in \tilde{\Gamma}_{n_2,n_1}(A)$ with $z_{n_1} \to z$. It must hold for $n_1 > j$ that

$$\begin{aligned}\operatorname{dist}(z_{n_1}, V_{n_1}) \leq \operatorname{dist}(z_{n_1}, V_j) &\leq |z_{n_1} - z| + \operatorname{dist}(z, U_j) \\ &< |z_{n_1} - z| + \frac{1}{n_2} - \epsilon.\end{aligned}$$

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 $\gamma_{n_2}(z;A) = \min\{\sigma_1((A-zI)|_{P_{n_2}\mathcal{H}}), \sigma_1((A^*-\bar{z}I)|_{P_{n_2}\mathcal{H}})\}$. 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 $W_e(A) \cap \operatorname{cl}(U)$ 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

$$Q(A,U) = \sup_{z \in W_e(A) \cap \operatorname{cl}(U)} \|R(z,A)\|^{-1}.$$

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

$$\lim_{n_1 \to \infty} \Gamma_{n_3,n_2,n_1}(A,U) = 1 - \chi_{[0,1/n_3]}(Q_{n_2}(A,U)) =: \Gamma_{n_3,n_2}(A,U).$$

Since $\chi_{[0,1/n_3]}$ has right limits and $Q_{n_2}(A,U)$ are non-increasing,

$$\lim_{n_2\to\infty} \Gamma_{n_3,n_2}(A,U) = 1 - \chi_{[0,1/n_3]}(Q(A,U)\pm) := \Gamma_{n_3}(A,U),$$

where $\pm$ denotes one of the right or left limits (it is possible to have either). Now if $\Xi^{\mathbb{C}}_{poll}(A,U) = 0$, then $\Gamma_{n_3}(A,U) = 0$ for all n_3. But if $\Xi^{\mathbb{C}}_{poll}(A,U) = 1$, 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^A_3 tower.

Step 2: $\{\Xi^{\mathbb{R}}_{poll}, \Omega_{\mathrm{D}}, \Lambda_2\} \notin \Delta^G_3$. 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 $\{\Xi^{\mathbb{R}}_{poll}, \Omega_{\mathrm{D}}, \Lambda_2\}$. 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 $\tilde{\Xi}(\{a_{i,j}\}) = \Xi^{\mathbb{R}}_{poll}(A,\mathbb{R})$ and hence we get a contradiction. □

7.6 Numerical Examples

In this section, we demonstrate that 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

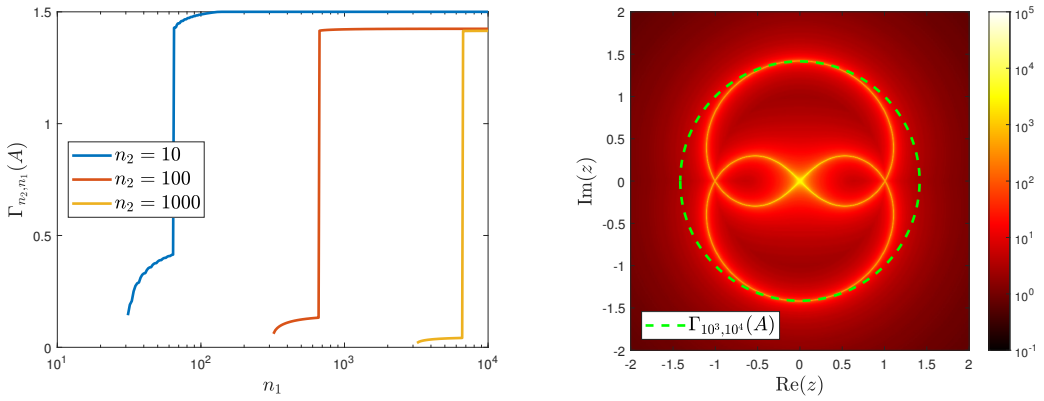


Figure 7.1: Left: Output of the algorithm for computing the spectral radius. Right: Pseudospectrum computed using the method of Chapter 3 (the colour scale corresponds to the resolvent norm $\|(A - zI)^{-1}\|$) which provides error control. We have show the output of $\Gamma_{10^3,10^4}(A)$ via the green dashed circle.

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. In the examples that follow, we have reminded the reader what each parameter n_k intuitively does in the relevant algorithm. Finally, we remind the reader of the comments in §2.3 - all of the algorithms can be implemented rigorously using arithmetic operations over the rationals or with methods such as interval arithmetic.

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.

[3]For this example and others on $l^2(\mathbb{Z})$, we reorder the basis so that the operator A acts on $l^2(\mathbb{N})$.

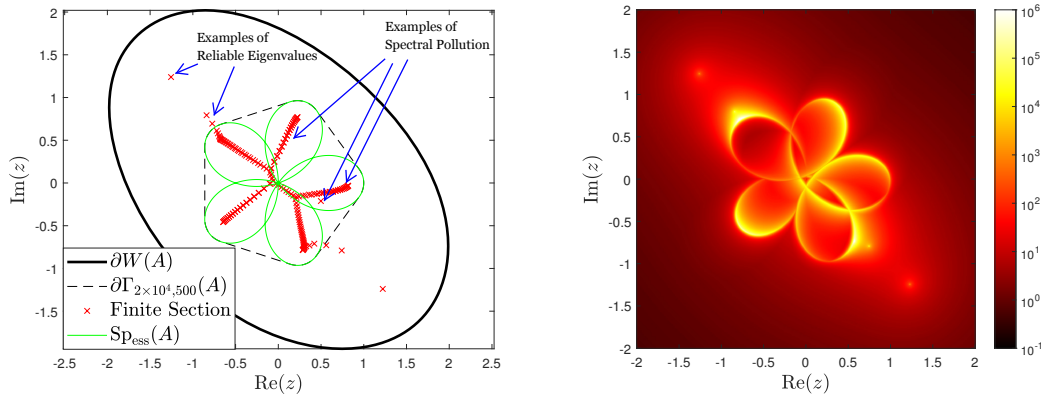


Figure 7.2: Left: The boundaries of $\partial W(A)$ and $\partial\Gamma_{2\times10^4,500}(A)$. We have also shown the essential spectrum of A (whose convex hull, in this example, corresponds to $W_e(A)$) and the output of finite section for a 200×200 truncation. Right: Pseudospectrum computed using the 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 $\partial\Gamma_{2\times10^4,500}(A)$ are accurate and, in this example, indicates that the other eigenvalues correspond to spectral pollution.

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

$$Ee_j = -\frac{3i}{1 + |j|}e_{j-1}.$$

Recall that the SCI classification for computing the essential numerical range is Π_2^A (see Theorem 7.3.8). 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 7.1.3).

The algorithm can also be extended to unbounded operators, as outlined in [Col19b].[4] For example, we

[4]The essential numerical range for unbounded operators was defined and studied in [BMT20].

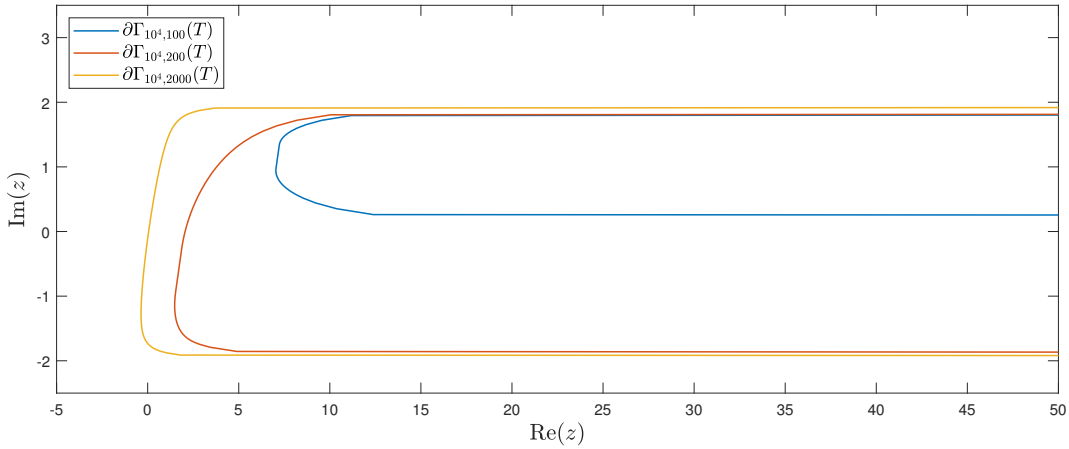


Figure 7.3: The output of the algorithm for computing the essential numerical range of closed operators, applied to the complex Schrödinger operator T in (7.6.1).

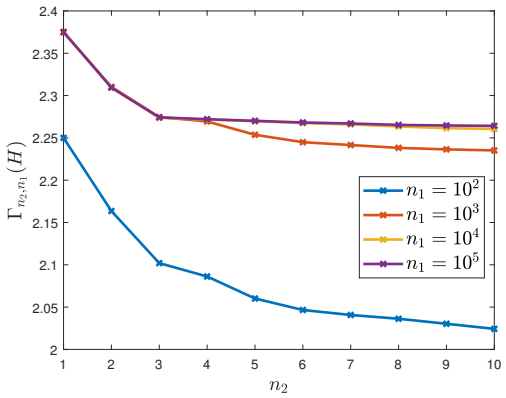


Figure 7.4: Output of the algorithm for computing the capacity of $\mathrm{Sp}(H)$, where H is the operator in (7.6.2).

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.

7.6.3 Numerical example for capacity

We now consider the transport Hamiltonian on a Penrose tile (shown in the left of Figure 4.9) discussed in §3.6.1 of Chapter 3. Let G be the graph consisting of the vertices, $V(G)$, of the Penrose tiling and $E(G)$ the set of edges. If there is an edge connecting two vertices x and y, we write $x \sim y$. Recall that the free

Hamiltonian H (Laplacian) acts on $\psi \in l^2(V(G)) \cong l^2(\mathbb{N})$ by

$$(H\psi)_i = \sum_{i \sim j} (\psi_j - \psi_i)\,, \tag{7.6.2}$$

with summation over nearest neighbour sites (vertices). Recall that by choosing a suitable ordering of the vertices, we can represent H 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 7.3.4 for further classifications). The first parameter, n_1, controls the size of the truncation used to test if intervals intersect the spectrum via Lemma 8.1.9, 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 7.4.2 for a discussion of computational efficiency.

Figure 7.4 shows the output of $\Gamma_{n_2,n_1}(H)$ 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 $\mathrm{Sp}(H)$ ($\mathrm{Sp}(H)$ is shown in Figure 3.2).

Chapter 8

Lebesgue Measure and Fractal Dimensions of Spectra

In this chapter, 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 great progress recently made in the field of Schrödinger operators with random or almost periodic potentials [Avi09, Avi08, AJ09, AK06, AV07, Pui04, Süt89]. Perhaps surprisingly, Cantor-like spectra occur in many families of one-dimensional operators. Whilst results are known for specific one-dimensional examples such as the almost Mathieu operator [AK06] (see §1.3 for a discussion regarding this problem, which was open for many years following numerical evidence [AA80, Tho83, Tho90, TT91]) or the Fibonacci Hamiltonian [Süt89], the problems of computing the Lebesgue measure and fractal dimensions of spectra remain open in the general case (see remarks in [DGS15] and references therein). 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.

Zero Lebesgue measure implies the absence of absolutely continuous spectrum (whose SCI is discussed in Chapter 5), which is related to transport properties if the operator represents a Hamiltonian of a quantum system. 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$^+$13, GG13, HSYY$^+$13, PGY$^+$13], strained materials [NBLOLT17, RTN14] or quasicrystals [BRS16, TGB$^+$14, KST87, LRF$^+$11].

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.

Recall that by a choice of basis we can, as in previous chapters, deal with arbitrary separable Hilbert spaces. As discussed in the first footnote of §4.1, some bases may be preferable to others, and one can view a different choice of basis as changing the evaluations functions Λ_i. The classes of operators we discuss are basis independent, apart from Ω_f and Ω_{D}.

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 $\mathrm{Leb}(\mathrm{Sp}(A)) = 0$?

A few comments are in order. First, we do not consider the final question for the pseudospectrum since $\mathrm{Leb}(\widehat{\mathrm{Sp}}_\epsilon(A)) > 0$. Second, 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 $\Omega = \Omega_{\mathrm{B}}, \Omega_{\mathrm{SA}}, \Omega_{\mathrm{N}}$ *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 $\mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}(A) \cap \mathbb{R})$.

Remark 8.1.2. *Although we consider* Ω_{D} *with* $\mathrm{Leb}_{\mathbb{R}}$ *throughout, all the proven lower bounds hold when considering bounded diagonal operators (dropping the restriction of self-adjointness) and using* Leb *instead of* $\mathrm{Leb}_{\mathbb{R}}$*. The proofs trivially generalise to the two-dimensional Lebesgue measure without altering the SCI classification.*

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 8.1.3. *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 $\Omega = \Omega_{\mathrm{B}}, \Omega_{\mathrm{SA}}, \Omega_{\mathrm{N}}$ *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. For simplicity, we shall consider these maps restricted to Ω_{D} and equip these diagonal operators with the operator norm topology.

Proposition 8.1.4. *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* $\epsilon > 0$.

It follows that Ξ_2^L is more stable than Ξ_1^L, explaining why it is easier to approximate. Again, this is the same state of affairs to comparing $\mathrm{Sp}(A)$ and $\mathrm{Sp}_\epsilon(A)$ as sets.

When is $\mathrm{Leb}(\mathrm{Sp}(A)) = 0$**?**

In this section we let $(\mathcal{M}, d)$ be the set $\{0, 1\}$ endowed 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 `LebSpec`, the algorithm constructed in Theorem 8.1.1. This relies on monotonicity of `LebSpec`. 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 8.1.5. *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,

$$\Delta_4^G \not\ni \{\Xi_3^L, \Omega, \Lambda_1\} \in \Pi_4^A, \quad \Delta_3^G \not\ni \{\Xi_3^L, \Omega, \Lambda_2\} \in \Pi_3^A.$$

Remark 8.1.6. *These are the first examples of computational spectral problems that require four limits to compute in the SCI hierarchy. To prove this, we need the tools from descriptive set theory/classical computational theory in Chapter 2 §2.4. Note that we prove the lower bounds for general algorithms, so regardless of the model of computation.*

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 (see references at the start of the chapter). 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]

[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.

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 [Col19b], 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.

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 8.1.7. *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 8.1.8. *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

$$\mathcal{H}^d(F) = \lim_{\delta \downarrow 0} \mathcal{H}_\delta^d(F).$$

[2]That is, the set of covers $\{U_i\}_{i \in I}$ with I at most countable and with $\mathrm{diam}(U_i) \leq \delta$.

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

$$\dim_H(F) \leq \underline{\dim}_B(F) \leq \overline{\dim}_B(F).$$

One of the properties that makes the Hausdorff dimension harder to compute than the box-counting dimension 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 8.1.9. *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 8.1.10. *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

$$\Delta_4^G \not\ni \{\Xi_H, \Omega_{\mathrm{SA}}, \Lambda_1\} \in \Sigma_4^A, \quad \Delta_3^G \not\ni \{\Xi_H, \Omega_{\mathrm{SA}}, \Lambda_2\} \in \Sigma_3^A.$$

Remark 8.1.11. *The results in this section and §8.1.1 can be interpreted in terms of real bounded sequences. Given such a sequence $\{a_i\}_{i\in\mathbb{N}}$ we can ask the same questions about $\mathrm{cl}(\{a_1, a_2, ...\})$ as we have asked about the spectrum. We can embed these problems as spectral problems for the class Ω_{D} of bounded self-adjoint diagonal operators, by simply considering diagonal operators with entries $\{a_1, a_2, ...\}$. Theorems 8.1.1, 8.1.5, 8.1.7 and 8.1.10 immediately then give the classifications. With regards to fractal dimensions, the key problem is to try and relate the amount of data that has been seen to the resolution obtained from the data (as highlighted in the numerical examples below). This has also been approached statistically [Hun90], assuming that the samples are taken from a probability distribution. Once we have the framework of the SCI, we can immediately see why the problem is so difficult - the computational problem requires three limits for the Hausdorff dimension.*

8.2 Proofs of Theorems on Lebesgue Measure

We will use the function `DistSpec` discussed in §3.5.1 of Chapter 3. 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.*

Proof. Without loss of generality we assume that the rectangle is $\{x + iy : x, y \in [0,1]\}$. Consider dividing the rectangle into n^2 subrectangles using the division of $[0,1]$ into n equal intervals. Given such a subrectangle, we can easily test whether the centre is in the union of the circles via a finite number of arithmetic operations and comparisons. Let $r(n)$ denote the number of subrectangles whose centre lies in the union. Then, since the boundary of the union of the circles has measure zero, it is easy to see that $r(n)/n^2$ converges to the desired Lebesgue measure. What is more, we can bound the number of subrectangles that intersect the boundary of any of the circles, and this can be used to obtain known precision. □

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

$$U(n_1, n_2, A) = [-n_1, n_1] \times [-n_1, n_1] \cap (\cup_{z \in \texttt{Grid}(n_1,n_2)} B(z, F_{n_1}(z))).$$

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 using the fact that `DistSpec` is and the above proposition. The only non-trivial part is convergence. The algorithm is summarised in the routine `LebSpec` in §8.4.1.

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 $n_1 \geq C$

$$U(n_1, n_2, A) = ([-C, C]^2 \cap (\cup_{z \in \texttt{Grid}(n_1,n_2)} B(z, F_{n_1}(z)))) \cup ([-n_1, n_1]^2 \backslash [-C, C]^2)$$

since $F_n(z) \geq \|R(z,A)\|^{-1}$. It follows that for large n_1

$$\Gamma_{n_2,n_1}(A) = 4C^2 - \mathrm{Leb}([-C, C]^2 \cap (\cup_{z \in \texttt{Grid}(n_1,n_2)} B(z, F_{n_1}(z)))).$$

As $n_1 \to \infty$, $[-C, C]^2 \cap (\cup_{z \in \texttt{Grid}(n_1,n_2)} B(z, F_{n_1}(z)))$ converges to the closed set

$$X(n_2, A) = [-C, C]^2 \cap (\cup_{z \in \texttt{Grid}(C,n_2)} B(z, \|R(z,A)\|^{-1}))$$

[3] We set $D(x, 0) = \emptyset$.

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 $(i,j)^{\mathrm{th}}$ 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

$$\gamma_{n,m}(z; A) = \min\{\sigma_1(P_m(A - zI)|_{P_n\mathcal{H}}), \sigma_1(P_m(A^* - \bar{z}I)|_{P_n\mathcal{H}})\},$$

where σ_1 denotes the injection modulus/smallest singular value. One can show that $\gamma_{n,m}$ converges uniformly on compact subsets to

$$\gamma_n(z; A) = \min\{\sigma_1((A - zI)|_{P_n\mathcal{H}}), \sigma_1((A^* - \bar{z}I)|_{P_n\mathcal{H}})\},$$

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 $\texttt{DistSpec}(A, n, m, z)$ (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

$$U(n_1, n_2, n_3, A) = [-n_2, n_2]^2 \cap (\cup_{z\in\texttt{Grid}(n_2,n_3)} B(z, \gamma_{n_2,n_1}(z; A))),$$
$$\Gamma_{n_3,n_2,n_1}(A) = (2n_2)^2 - \mathrm{Leb}(U(n_1, n_2, n_3, A))$$

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 7.3.4. 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 8.1.3. 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 8.1.3. 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

$$E_n = \frac{1}{n}\left(\mathbb{Z} + i\mathbb{Z}\right) \cap \{z \in \mathbb{C} : F_n(z) \leq \epsilon\} \cap [-n, n]^2.$$

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. The algorithm is presented in the `LebPseudoSpec` routine in §8.4.1 and the following shows that this algorithm is sharp in the SCI hierarchy.

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:

$$\chi_{\cup_{z\in E_n} D(z, \max\{0, \epsilon - F_n(z)\}} \to \chi_{\widehat{\mathrm{Sp}}_\epsilon(A)},$$

where χ_E denotes the indicator function of a set E. It follows by the dominated convergence theorem that $\Gamma_n(A) \to \mathrm{Leb}(\widehat{\mathrm{Sp}}_\epsilon(A))$. 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. □

Proof of Proposition 8.1.4. We begin with the proof of 1. Suppose $A \in \Omega_{\mathrm{D}}$ has $\mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}(A)) = 0$ and let $A_n \in \Omega_{\mathrm{D}}$ be such that $\|A - A_n\| \to 0$ as $n \to \infty$. This implies that $\mathrm{Sp}(A_n) \to \mathrm{Sp}(A)$ since all our operators are normal. To prove that $\mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}(A_n)) \to 0$, it is enough to prove that

$$\mathrm{Leb}(F_n) \downarrow 0, \tag{8.2.1}$$

where $F_n = \mathrm{Sp}(A) \cup (\cup_{m\geq n}\mathrm{Sp}(A_m))$. But F_n decreases to $\mathrm{Sp}(A)$ and is bounded in measure so (8.2.1) holds. For the converse, let $\mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}(A)) > 0$. Without loss of generality, assume that all of A's entries

lie in $[0,1]$. Let $\mathbb{D}_n$ denote the set $\{j/2^n\}_{j=1}^n$ and let us consider the map $\phi_n : x \hookrightarrow 2^{-n}\lceil x2^n \rceil$ on $[0,1]$. Let A_n be the diagonal operator obtained by applying ϕ_n to each of A's entries. We clearly have that $\|A - A_n\| \to 0$ as $n \to \infty$ but note that $\mathrm{Sp}(A_n)$ is finite so has Lebesgue measure 0. Hence Ξ_1^L is discontinuous at A.

To prove 2, note that for $A \in \Omega_{\mathrm{D}}$, $\mathrm{Leb}_{\mathbb{R}}(S_\epsilon(A)) = 0$. Let $A_n \in \Omega_{\mathrm{D}}$ have $\|A - A_n\| \to 0$. Then given some $0 < \delta < \epsilon$ it holds for large n that $\mathrm{Sp}_{\epsilon-\delta}(A) \subset \mathrm{Sp}_\epsilon(A_n) \subset \mathrm{Sp}_{\epsilon+\delta}(A)$ and hence that

$$\limsup_{n\to\infty} \mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}_\epsilon(A_n)) \leq \mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}_{\epsilon+\delta}(A))$$
$$\liminf_{n\to\infty} \mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}_\epsilon(A_n)) \geq \mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}_{\epsilon-\delta}(A)).$$

Now let $\delta \downarrow 0$ and use the fact that Ξ_2^L is continuous in ϵ. □

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 8.1.5. 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 $\Omega = \Omega_{\mathrm{B}}, \Omega_{\mathrm{SA}}$, Ω_{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 $\Omega = \Omega_{\mathrm{SA}}$, Ω_{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 $n_1 \to \infty$.
- $\widetilde{\Gamma}_{n_2}(A)$ converges to $\mathrm{Leb}(\mathrm{Sp}(A))$ monotonically from above as $n_2 \to \infty$.

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 $n_1 \to \infty$. 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

$$\lim_{n_1\to\infty} \Gamma_{n_3,n_2,n_1}(A) = \chi_{[0,1/n_3]}(\widetilde{\Gamma}_{n_2}(A)).$$

It follows that

$$\lim_{n_2\to\infty} \lim_{n_1\to\infty} \Gamma_{n_3,n_2,n_1}(A) = \chi_{[0,1/n_3]}(\mathrm{Leb}(\mathrm{Sp}(A))\pm),$$

where $\pm$ 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: $\{\Xi_3^L,\Omega_{\rm D},\Lambda_1\}\notin\Delta_3^G$. Assume for a contradiction that this is false and $\widehat{\Gamma}_{n_2,n_1}$ is a general height two tower for $\{\Xi_3^L,\Omega_{\rm D}\}$. 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 2.4.7 in Chapter 2 §2.4 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 $\{A_{i,j}:i,j\le N\}$ 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: $\{\Xi_3^L,\Omega_{\rm SA},\Lambda_1\}\notin\Delta_4^G$. Suppose for a contradiction that Γ_{n_3,n_2,n_1} is a height three tower of general algorithms for the problem $\{\Xi_3^L,\Omega_{\rm SA},\Lambda_1\}$. 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

$$\tilde{\Xi}_4(\{a_{m,i,j}\}):\ \text{For every } m, \text{ does } \{a_{m,i,j}\}_{i,j} \text{ have (only) finitely many columns with (only) finitely many 1's?}$$

Recall that it was shown in Theorem 2.4.7 in Chapter 2 §2.4 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 7.3.4 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 7.3.4 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 8.1.7. 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+19]. 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 8.1.9 on the problem of determining whether $\mathrm{Sp}(A)\cap(a,b)\neq\emptyset$.

Proof of Lemma 8.1.9. 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 $\mathtt{DistSpec}(A,n,f(n),c)+c_n<\delta$. 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 $\mathtt{DistSpec}(A,n,f(n),c)+c_n$ 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 $\mathtt{DistSpec}(A,n,f(n),c)+c_n$ 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 8.1.9.

Proof of Theorem 8.1.10. 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 8.1.5. 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 2.4.7.

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

$$\tilde{\Xi}_4(\{a_{m,i,j}\}) : \text{ For every } m, \text{ does } \{a_{m,i,j}\}_{i,j} \text{ have (only) finitely many columns with (only) finitely many 1's?}$$

Recall that it was shown in Theorem 2.4.7 in Chapter 2 §2.4 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 8.1.5. 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

$$\mathcal{A}_{n_3,n_2,n_1}(A) = \{\{U_i\}_{i\in I} : I \text{ is finite }, S_{n_1,n_2}(A) \subset \cup_{i\in I} U_i, U_i \in \cup_{n_3\le l\le n_2} \rho_l\}$$

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 8.1.9 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

$$\mathcal{C}_{n_3,n_2}(A) = \left\{\{U_i\}_{i\in I} : I \text{ is finite }, \mathrm{Sp}(A) \cap \mathbb{D}_{n_2}^c \subset \cup_{i\in I} U_i, U_i \in \cup_{n_3\le l\le n_2} \rho_l\right\}$$

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

$$\Gamma_{n_3,n_2,n_1}(A) = \sup_{j=1,\ldots,2^{n_3}} \left\{ \frac{j}{2^{n_3}} : h_{n_3,n_2,n_1}(A, k/2^{n_3}) + \frac{1}{n_2} > \frac{1}{2} \text{ for } k = 1, ..., j \right\},$$

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

$$\lim_{n_1 \to \infty} \Gamma_{n_3,n_2,n_1}(A) = \sup_{j=1,\ldots,2^{n_3}} \left\{ \frac{j}{2^{n_3}} : h_{n_3,n_2}(A, k/2^{n_3}) + \frac{1}{n_2} > \frac{1}{2} \text{ for } k = 1, ..., j \right\} =: \Gamma_{n_3,n_2}(A).$$

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

$$\lim_{n_2 \to \infty} \Gamma_{n_3,n_2}(A) = \sup_{j=1,\ldots,2^{n_3}} \left\{ \frac{j}{2^{n_3}} : h_{n_3}(A, k/2^{n_3}) \geq \frac{1}{2} \text{ for } k = 1, ..., j \right\} =: \Gamma_{n_3}(A).$$

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 $\{1/2^{n_3}, 2/2^{n_3}, ..., 1\}$ 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 8.1.9 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 8.1.9. □

8.4 Numerical Examples

In this section, 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. Finally, we remind the reader of the comments in §2.3 - all of the algorithms can be implemented rigorously using arithmetic operations over the rationals or with methods such as interval arithmetic.

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. The routines for these examples are shown in pseudocode below. Recall that $F_n(z) := \texttt{DistSpec}(A, n, f(n), z) + c_n$ converges uniformly to $\|R(z, A)\|^{-1}$ from above on compact subsets of $\mathbb{C}$.

Function `LebSpec` $(n_1, n_2, f(n_1), c_{n_1}, A)$
  Input : $n_1, n_2, f(n_1) \in \mathbb{N}$, $c_{n_1} \in \mathbb{R}_+$, $A \in \Omega_f$
  Output: $\Gamma_{n_2,n_1}(A)$, a Π_2^A approximation of $\mathrm{Leb}(\mathrm{Sp}(A))$
  $G = \frac{1}{2^{n_2}}(\mathbb{Z} + i\mathbb{Z}) \cap [-n_1, n_1]^2 = \{z_1, ..., z_m\}$
  for $z \in G$ **do**
    $F_{n_1}(z) = \texttt{DistSpec}(A, n_1, f(n_1), z) + c_{n_1}$
  end
  NB: WLOG we adapt F_{n_1} to be non-increasing in n_1.
  $U = [-n_1, n_1]^2 \cap (\cup_{j=1}^m B(z_j, F_{n_1}(z_j)))$
  $\Gamma_{n_2,n_1}(A) = 4n_1^2 - \mathrm{Leb}(U(n_2, n_1, A))$
end

Function `LebPseudoSpec`(n, A, ϵ)

Input : $n \in \mathbb{N}$, $A \in \Omega_\epsilon^L$, $\epsilon > 0$

Output: $\Gamma_n(A)$, a Σ_1^A approximation of $\mathrm{Leb}(\mathrm{Sp}_\epsilon(A))$

$G = \frac{1}{n}(\mathbb{Z} + i\mathbb{Z}) \cap [-n, n]^2 = \{z_1, ..., z_m\}$

for $z \in G$ **do**

  $F_n(z) =$ `DistSpec`$(A, n, f(n), z) + c_n$

end

NB: WLOG we adapt F_n to be non-increasing in n.

$S = \{z \in G : F_n(z) \leq \epsilon\}$

$\Gamma_n(A) = \mathrm{Leb}(\cup_{z \in S} D(z, \max\{0, \epsilon - F_n(z)\})$

end

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 the true value of $\mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}(H_\alpha))$, which is well-estimated by `LebSpec` (Method 1).

We also compare Method 1 with the naive estimate provided by finite section estimates $\mathrm{Sp}(P_n H_\alpha P_n)$, where P_n is the orthogonal projection onto $\mathrm{span}\{e_k : |k| \leq n\}$. In general, suppose there is some algorithm $\tilde{\Gamma}_n(H_\alpha)$ convergent to $\mathrm{Sp}(H_\alpha)$ in the Hausdorff metric (this is not true in general for $\mathrm{Sp}(P_n H_\alpha P_n)$). Let $\mathcal{I}_m$ be the collection of open intervals $\{(j/2^m, (j+1)/2^m) : j \in \mathbb{Z}\}$ and set

$$\hat{\Gamma}_{n_2,n_1}(H_\alpha) = \frac{1}{2^{n_2}} \left| \{I \in \mathcal{I}_{n_2} : I \cap \tilde{\Gamma}_{n_1}(H_\alpha) \neq \emptyset\} \right|,$$

where we denote the cardinality of set using the absolute value notation. Since the intervals are open,

$$\liminf_{n_1 \to \infty} \hat{\Gamma}_{n_2,n_1}(H_\alpha) \geq \frac{1}{2^{n_2}} \left| \{I \in \mathcal{I}_{n_2} : I \cap \mathrm{Sp}(H_\alpha) \neq \emptyset\} \right| := \hat{\Gamma}_{n_2}(H_\alpha)$$

and

$$\hat{\Gamma}_{n_2}(H_\alpha) \downarrow \mathrm{Leb}_{\mathbb{R}}(\mathrm{Sp}(H_\alpha)) \text{ as } n_2 \to \infty.$$

Furthermore, for any $\epsilon > 0$,

$$\limsup_{n_1 \to \infty} \hat{\Gamma}_{n_2,n_1}(H_\alpha) \leq \frac{1}{2^{n_2}} \left| \{I \in \mathcal{I}_{n_2} : I \cap \mathrm{Sp}_\epsilon(H_\alpha) \neq \emptyset\} \right| := \hat{\Upsilon}_{n_2}(H_\alpha, \epsilon)$$

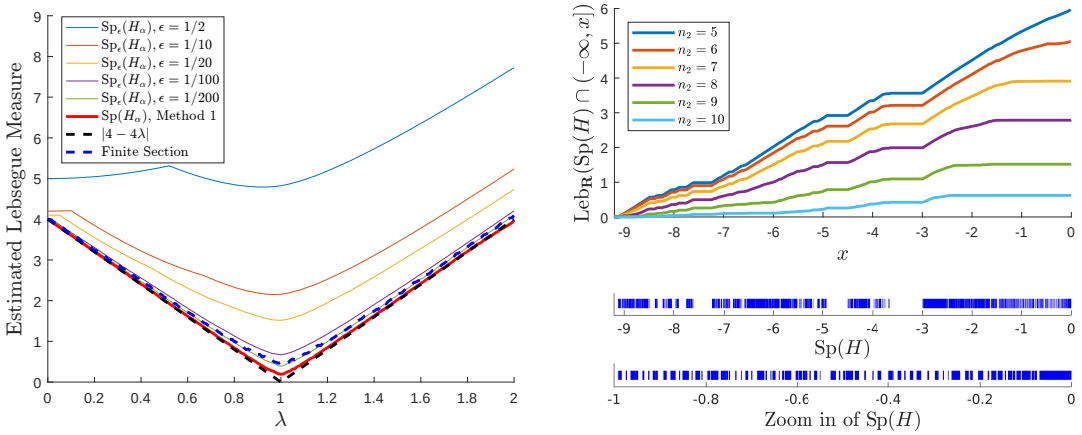


Figure 8.1: Left: Output of algorithm to compute $\mathrm{Leb}_\mathbb{R}(\mathrm{Sp}_\epsilon(H_\alpha))$ as well as the direct algorithm for $\mathrm{Leb}_\mathbb{R}(\mathrm{Sp}(H_\alpha))$ from §8.1.1 (Method 1). Note that we gain convergence to the true value as $\epsilon \downarrow 0$. Right: Estimates 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.

and

$$\hat{\Upsilon}_{n_2}(H_\alpha, \epsilon) \downarrow \mathrm{Leb}_\mathbb{R}(\mathrm{Sp}_\epsilon(H_\alpha)) \text{ as } n_2 \to \infty.$$

Since $\epsilon > 0$ was arbitrary, if $\lim_{n_1 \to \infty} \hat{\Gamma}_{n_2,n_1}(H_\alpha)$ exists, then

$$\lim_{n_2 \to \infty} \lim_{n_1 \to \infty} \hat{\Gamma}_{n_2,n_1}(H_\alpha) = \mathrm{Leb}_\mathbb{R}(\mathrm{Sp}(H_\alpha)).$$

For comparison, $\hat{\Gamma}_{7,5000}(H_\alpha)$ is shown for $\tilde{\Gamma}_n(H_\alpha) = \mathrm{Sp}(P_n H_\alpha P_n)$. 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 and speed/parallelisation. Furthermore, the finite section method does not always converge.

Graphical Laplacian on Penrose tile

We now consider the transport Hamiltonian on a Penrose tile (shown in the left of Figure 4.9) discussed in §3.6.1 of Chapter 3. Recall that the free Hamiltonian H (Laplacian) is given by

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

with summation over nearest neighbour sites (vertices). 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

via `DistSpec`. 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

Now we turn to demonstrating the algorithms for fractal dimensions. For $A \in \Omega_f^{BD}$, the routine `BoxDim` computes the box-counting dimension of $\mathrm{Sp}(A)$ (see §3.5.1 for the routine `CompSpec`). The routine `HausDim` computes $\dim_H(\mathrm{Sp}(A))$ for $A \in \Omega_f \cap \Omega_{\mathrm{SA}}$.

Function `BoxDim`$(n_1, n_2, f(n_1), c_{n_1}, A)$

Input : $n_1, n_2, f(n_1) \in \mathbb{N}$, $c_{n_1} \in \mathbb{R}_+$, $A \in \Omega_f^{BD}$

Output: $\Gamma_{n_2,n_1}(A)$, a Π_2^A approximation of $\dim_B(\mathrm{Sp}(A))$

if $n_1 \geq n_2$ **then**

for $l = 1, ..., n_1$ **do**

$S_l = \mathtt{CompSpec}(A, l, g : x \to x, f(l), c_l) = \{z_{1,l}, ..., z_{k_l,l}\}$

NB: WLOG we assume that $\mathrm{dist}(z_{j,l}, \mathrm{Sp}(A)) \leq 2^{-l}$.

for $j = 1, ..., k_l$ **do**

$I_{j,l} = [z_{j,l} - 2^{-l}, z_{j,l} + 2^{-l}]$

end

end

for $k \in \{n_2, n_2 + 1, ..., n_1\}$, $j \in \{1, 2, ..., n_1\}$ **do**

Let $\Upsilon_{k,j}$ be any union of $2^{-k}-$mesh intervals of minimal length $|\Upsilon_{k,j}|$ (where length is number of mesh intervals that make up the union) such that

$$\Upsilon_{k,j} \cap I_{p,q} \neq \emptyset, \quad 1 \leq q \leq j, \quad 1 \leq p \leq k_q.$$

$a_{k,j} = \frac{\log(|\Upsilon_{k,j}(A)|)}{k \log(2)}$

end

$\Gamma_{n_2,n_1}(A) = \max\{a_{k,j} : n_2 \leq k \leq n_1, 1 \leq j \leq n_1\}$ (max over empty set is zero).

else

$\Gamma_{n_2,n_1}(A) = 0$

end

end

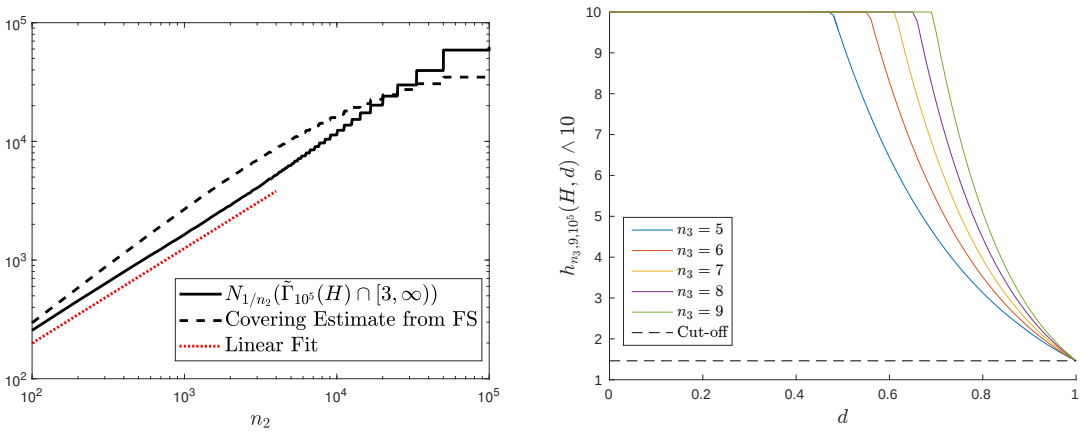


Figure 8.2: Left: A plot of $N_{1/n_2}(\tilde{\Gamma}_{10^5}(H)\cap[-3,\infty))$ against n_2. We found a scaling region with estimated box-counting dimension ≈ 0.8. Note that for large $n_2 \gtrsim 5000$, scalings are not resolved by $\tilde{\Gamma}_{10^5}$ (we can predict when this happens using the 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 $\mathrm{Sp}(H)\cap[-3,\infty)$.

Function `HausDimSpec` (n_1, n_2, n_3, A)

Input : $n_1, n_2, n_3 \in \mathbb{N}$, $A \in \Omega_f \cap \Omega_{\mathrm{SA}}$

Output: $\Gamma_{n_3,n_2,n_1}(A)$, a Σ_3^A approximation of $\dim_H(\mathrm{Sp}(A))$

Notation: ρ_k denotes set of all closed intervals of form $[2^{-k}m, 2^{-k}(m+1)]$, $m \in \mathbb{Z}$

S_{n_1,n_2} = union of all $S \in \rho_{n_2}$ with $S \subset [-n_1, n_1]$ and such that the algorithm discussed in Lemma 8.1.9 outputs 'Yes' for the interior of S and input parameter n_1.

$\mathcal{A}_{n_3,n_2,n_1} = \{\{U_i\}_{i\in I} : I \text{ is finite }, S_{n_1,n_2} \subset \cup_{i\in I} U_i, U_i \in \cup_{n_3\le l\le n_2}\rho_l\}$

for $m \in \{1, ..., 2^{n_3}\}$ **do**

  $b_m = \inf\left\{\sum_i \mathrm{diam}(U_i)^{m/2^{n_3}} : \{U_i\} \in \mathcal{A}_{n_3,n_2,n_1}\right\} + n_2^{-1}$

end

$\Gamma_{n_3,n_2,n_1}(A) = \max\{m/2^{n_3} : b_j > 1/2 \text{ for } j = 1, ..., m\}$ (max over empty set is zero).

end

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 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

$$N_\delta \sim 1/\delta^d, \quad \delta \to 0. \tag{8.4.2}$$

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. This corresponds to a rectangular truncation with $n_1 = 10^5$ columns. Recall that $\tilde{\Gamma}_n$ denotes the algorithm that converges to the spectrum with error control, in particular avoiding spectral pollution (see Chapter 3). 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 (this is at least as large as the region shown in the plot). 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.

We have also shown, in Figure 8.2, what happens when one performs the same experiment but with finite section replacing $\tilde{\Gamma}_n$ (now using a square $10^5 \times 10^5$ truncation). There are two noticeable features. 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[4] for approximating fractal dimensions and why the new algorithms of this chapter (which are proven to converge) are needed.

Finally, we investigate the Hausdorff dimension. An efficient way to compute a minimal covering is to use binary trees. In general, there is no way of dealing with a height three tower without extra information describing which subsequences $n_2(n_3)$ and $n_1(n_3)$ to pick. 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 $\mathcal{H}^d_\delta$ (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. We found this region to be largely consistent as we varied n_2 and similar results were found for other regions of the spectrum above -3. There is a cut-off point bounding the estimates below 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 $\mathrm{Sp}(H) \cap [-3, \infty)$. Both algorithms support the possibility that the spectrum above -3 is fractal and hence has Lebesgue measure zero.

[4]There do exist examples of operators, typically with a lot of structure, where one can use periodic versions of finite section [PEF15].

Part III
Extensions of Classical Finite-Dimensional Algorithms

Chapter 9

The Infinite-Dimensional QR Algorithm

In this chapter, we examine the QR algorithm, the most celebrated of eigenvalue algorithms for finite-dimensional problems and regarded as one of the ten most important algorithms of the 20th century [DS00]. We consider the algorithm in infinite dimensions, dubbed the IQR algorithm. The work here is based on [CH19b] and many of the results from finite dimensions carry over. However, there are two important caveats. First, it is possible for part of the spectrum not to be recovered by the IQR algorithm (see Example 9.2.11). Second, the IQR algorithm cannot, in general, be accelerated with the use of shift strategies, which considerably accelerate the finite-dimensional algorithm [Par98]. This is because there is no final column of an infinite matrix, and hence the usual link with inverse iteration cannot be made.

Despite these drawbacks of the infinite-dimensional setting, we show how the IQR algorithm can be used to gain new classification results and convergent algorithms in the SCI hierarchy. The outcome is more elaborate than the finite-dimensional case, as the infinite-dimensional setting includes more intricate instances. In particular, it is proven that, for normal operators, the algorithm converges to the discrete spectrum outside the convex hull of the essential spectrum, with the rate of convergence generalising the well-known result in finite dimensions. This is extended to dominant invariant subspaces for general (possibly non-normal) operators as well as the spectrum of a large class of operators that includes compact normal operators with eigenvalues of distinct magnitude. For example, we demonstrate Δ_1 classification for the extremal part of the spectrum, and dominant invariant subspaces, as well as Σ_1 results for spectra of certain classes of compact operators. Note that the general spectral problem for compact operators is not in Σ_1.

Historically, Deift, Li and Tomei [DLT85] provided the first results on the IQR algorithm in connection with Toda flows with infinitely many variables (covering self-adjoint infinite matrices with real entries). Their results are purely functional analytic and do not take implementation and computability issues into account. However, these results provide the fundamentals of the IQR algorithm. In [Han08b], the results of [DLT85] were expanded with a convergence result for eigenvectors corresponding to eigenvalues outside the essential numerical range for normal operators. However, this paper did not consider convergence rates, actual numerical calculation, or any classification results. A discussion of implementation for banded operators was given in [Han08a]. We will extend the analysis to more general operators and instances, and also answer the crucial question: can one actually implement the IQR algorithm? We show that for matrices with finitely many entries in each column, the computation collapses to a finite one. In general, for an invertible operator, we can compute the iterates to any given accuracy in finite time.

The IQR algorithm and convergence theorems are demonstrated on a large collection of difficult com-

putational spectral problems from the sciences in §9.5, with comparisons to the finite section method. Moreover, the examples demonstrate that the IQR algorithm performs much better than predicted by our theory, working on much larger classes of operators. Hence, we are left with many open problems on the theoretical understanding of the potential of this algorithm (see the concluding remarks of this thesis).

9.1 Background

As well as the notation defined in §1.4, we will need the following. We will consider the canonical separable Hilbert space $\mathcal{H} = l^2(\mathbb{N})$ (the set of square summable sequences). We denote the canonical orthonormal basis of $l^2(\mathbb{N})$ by $\{e_j\}_{j\in\mathbb{N}}$, and if $\xi \in \mathcal{H}$ we write $\xi(j) = \langle \xi, e_j\rangle$. Note that $A \in \mathcal{B}(l^2(\mathbb{N}))$ is uniquely determined by its matrix elements $a_{ij} = \langle Ae_j, e_i\rangle$. Hence we will use the words bounded operator and infinite matrix interchangeably in this chapter. Given a sequence of operators $\{A_n\}$, we will use the notation

$$A_n \xrightarrow{\text{SOT}} A, \qquad A_n \xrightarrow{\text{WOT}} A$$

to mean convergence in the strong and weak operator topology, respectively.

We also need a notion of convergence of subspaces. We follow the notation in [Kat95]. Let $M \subset \mathcal{B}$ and $N \subset \mathcal{B}$ be two non-trivial closed subspaces of a Banach space $\mathcal{B}$. The distance between them is defined by

$$\delta(M,N) = \sup_{\substack{x\in M\\ \|x\|=1}} \inf_{y\in N} \|x-y\| \in [0,1], \qquad \hat{\delta}(M,N) = \max[\delta(M,N), \delta(N,M)].$$

Given subspaces M and $\{M_k\}$ such that $\hat{\delta}(M_k, M) \to 0$ as $k \to \infty$, we will use the notation $M_k \to M$. If we replace $\mathcal{B}$ with a Hilbert space $\mathcal{H}$, we can express δ and $\hat{\delta}$ conveniently in terms of projections and operator norms. In particular, if E and F are the orthogonal projections onto subspaces $M \subset \mathcal{H}$ and $N \subset \mathcal{H}$ respectively, then

$$\delta(M,N) = \sup_{\substack{x\in M\\ \|x\|=1}} \inf_{y\in N} \|x-y\| = \sup_{\substack{x\in M\\ \|x\|=1}} \|F^{\perp}x\| = \|F^{\perp}E\|.$$

Since the operator $E - F = F^{\perp}E - FE^{\perp}$ is essentially the direct sum of operators $F^{\perp}E \oplus (-FE^{\perp})$, its norm is $\hat{\delta}(M,N)$, i.e.

$$\hat{\delta}(M,N) = \max(\|F^{\perp}E\|, \|E^{\perp}F\|) = \max(\|F^{\perp}E\|, \|FE^{\perp}\|) = \|E-F\|.$$

This allows us to extend the definition to allow the trivial subspace $\{0\}$ and gives rise to a metric on the set of all closed subspaces of $\mathcal{H}$ (first introduced by Krein and Krasnoselski in [KK47]). Since $0 \le \hat{\delta}(M,N) \le 1$, we also define the (maximal) subspace angle, $\phi(M,N) \in [0, \pi/2]$, between M and N by

$$\sin\big(\phi(M,N)\big) = \hat{\delta}(M,N). \tag{9.1.1}$$

Finally, we will use two further well-known properties in the Hilbert space setting. First, if M and N are both finite l-dimensional subspaces, then

$$\delta(M,N) \le l^{\frac{1}{2}}\delta(N,M), \tag{9.1.2}$$

showing that to prove convergence of finite-dimensional subspaces, it is enough to prove δ-convergence. Second, suppose we have

$$M = \bigoplus_{j=1}^{n} M_j, \quad N^{(k)} = N_1^{(k)} + \ldots + N_n^{(k)},$$

where the $N_j^{(k)}$ need not be orthogonal. Then a simple application of Hölder's inequality yields

$$\delta(M, N^{(k)}) \leq \left(\sum_{j=1}^{n} \delta\left(M_j, N_j^{(k)}\right)^2 \right)^{\frac{1}{2}}, \tag{9.1.3}$$

which shows that if the dimensions of M_j and $N_j^{(k)}$ are finite and equal, then to prove convergence $N^{(k)} \to M$ we only need to prove that $\delta\left(M_j, N_j^{(k)}\right) \to 0$ as $k \to \infty$. For further properties (including other notions of distances between subspaces) and a discussion on two projections theory, we refer the reader to the excellent article of Böttcher and Spitkovsky [BS10].

9.1.1 The QR decomposition

The QR decomposition forms the basic building block of the QR algorithm. Given a finite matrix $A \in \mathbb{C}^{n\times n}$, one may apply (a stable version of) the Gram–Schmidt procedure to the columns of A and store these columns together in a matrix Q. This gives the QR decomposition

$$A = QR,$$

where Q is a unitary matrix and R is upper triangular. A QR decomposition also exists in the infinite-dimensional case. One key ingredient in the QR algorithm is the Householder transformation used for computational reasons (they are backwards stable). Our goal is to extend the construction of the QR decomposition, via Householder transformations, to infinite matrices, allowing implementation on a finite machine.

Definition 9.1.1. *A Householder reflection is an operator $S \in \mathcal{B}(\mathcal{H})$ of the form*

$$S = I - \frac{2}{\|\xi\|^2}\xi \otimes \bar{\xi}, \qquad \xi \in \mathcal{H},$$

where $\bar{\xi}$ denotes the associated functional in $\mathcal{H}^*$ given by $x \to \langle x, \xi\rangle$. In the case where $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ and I_i is the identity on $\mathcal{H}_i$ then

$$U = I_1 \oplus \left(I_2 - \frac{2}{\|\xi\|^2}\xi \otimes \bar{\xi} \right) \qquad \xi \in \mathcal{H}_2,$$

will be called a Householder transformation.

A straightforward calculation shows that $S^* = S^{-1} = S$ and thus also $U^* = U^{-1} = U$. An important property of the operator S is that if $\{e_j\}$ is an orthonormal basis for $\mathcal{H}$ and $\eta \in \mathcal{H}$, then one can choose $\xi \in \mathcal{H}$ such that

$$\langle S\eta, e_j\rangle = \left\langle \left(I - \frac{2}{\|\xi\|^2}\xi \otimes \bar{\xi} \right)\eta, e_j \right\rangle = 0, \qquad \forall j \neq 1.$$

In other words, one can introduce zeros in the column below the diagonal entry. Indeed, if $\eta_1 = \langle \eta, e_1\rangle \neq 0$ one may choose $\xi = \eta \pm \|\eta\|\zeta$, where $\zeta = \eta_1/|\eta_1|e_1$ and if $\eta_1 = 0$ choose $\xi = \eta \pm \|\eta\|e_1$. The following theorem gives the existence of a QR decomposition, even in the case where the operator is not invertible.

Theorem 9.1.2 ([Han08b])**.** *Let A be a bounded operator on a separable Hilbert space $\mathcal{H}$ and let $\{e_j\}_{j\in\mathbb{N}}$ be an orthonormal basis for $\mathcal{H} \cong l^2(\mathbb{N})$. Then there exists an isometry Q such that $A = QR$, where R is upper triangular with respect to $\{e_j\}$. Moreover,*

$$Q = \underset{n\to\infty}{\text{SOT-lim}}\ V_n$$

where $V_n = U_1 \cdots U_n$ are unitary and each U_j is a Householder transformation.

9.1.2 The IQR algorithm

Suppose now that $A \in \mathcal{B}(\mathcal{H})$ is invertible and let $\{e_j\}$ be an orthonormal basis for $\mathcal{H}$. Theorem 9.1.2 gives the decomposition $A = QR$, where Q is an isometry and R is upper triangular with respect to $\{e_j\}$. Since we assumed that A is invertible, Q is in fact unitary. To define the IQR algorithm, we proceed as in the finite-dimensional case via unitary operators $\{\hat{Q}_k\}$ and upper triangular (with respect to $\{e_j\}$) operators $\{\hat{R}_k\}$ as follows. Let $A = Q_1R_1$ be a QR decomposition of A and define $A_1 = R_1Q_1$. Then QR factorise $A_1 = Q_2R_2$ and define $A_2 = R_2Q_2$. The recursive procedure becomes

$$A_{m-1} = Q_mR_m, \quad A_m = R_mQ_m. \tag{9.1.4}$$

Now define

$$\hat{Q}_m = Q_1Q_2 \dots Q_m, \quad \hat{R}_m = R_mR_{m-1} \dots R_1. \tag{9.1.5}$$

This is known as the QR algorithm and is completely analogous to the finite-dimensional case. Note also that we have $A_n = \hat{Q}_n^* A \hat{Q}_n$. In the finite-dimensional case under favourable conditions, $\hat{Q}_n^* A \hat{Q}_n$ converges to a diagonal operator and the columns of $\hat{Q}_n$ converge to the corresponding eigenvectors as $n \to \infty$ (see Theorem 9.2.1 below). We will see that the IQR algorithm behaves similarly for the extreme parts of the spectrum.

Definition 9.1.3. *Let $A \in \mathcal{B}(\mathcal{H})$ be invertible and let $\{e_j\}$ be an orthonormal basis for $\mathcal{H}$. The sequences $\{\hat{Q}_j\}$ and $\{\hat{R}_j\}$ constructed as in (9.1.4) and (9.1.5) will be called a Q-sequence and an R-sequence of A with respect to $\{e_j\}$.*

Remark 9.1.4. *Note that since the Householder transformations used in the proof of Theorem 9.1.2 are unique up to a $\pm$ sign we will with some abuse of language refer to the QR decomposition constructed as the QR decomposition. In general for an invertible operator, the IQR algorithm is uniquely defined up to phase - see §9.3.2. This will not be a problem for our theorems or numerical examples.*

Slightly more can be immediately said about the above construction. We have

$$A = Q_1R_1 = \hat{Q}_1\hat{R}_1,$$

$$A^2 = Q_1R_1Q_1R_1 = Q_1Q_2R_2R_1 = \hat{Q}_2\hat{R}_2,$$

$$A^3 = Q_1R_1Q_1R_1Q_1R_1 = Q_1Q_2R_2Q_2R_2R_1 = Q_1Q_2Q_3R_3R_2R_1 = \hat{Q}_3\hat{R}_3.$$

An easy induction gives us that

$$A^m = \hat{Q}_m\hat{R}_m. \tag{9.1.6}$$

Note that $\hat{R}_m$ must be upper triangular with respect to $\{e_j\}_{j\in\mathbb{N}}$ since R_j, $j \leq m$ is upper triangular with respect to $\{e_j\}_{j\in\mathbb{N}}$. Also, if A is invertible then $\langle Re_i, e_i\rangle \neq 0$. Hence we obtain the useful result that

$$\operatorname{span}\{A^m e_j\}_{j=1}^J = \operatorname{span}\{\hat{Q}_m e_j\}_{j=1}^J, \quad J \in \mathbb{N}. \tag{9.1.7}$$

9.2 Convergence Theorems

In finite dimensions we have the following well-known theorem:

Theorem 9.2.1 (Finite dimensions). *Let $A \in \mathbb{C}^{N\times N}$ be a normal matrix with eigenvalues satisfying $|\lambda_1| > \ldots > |\lambda_N|$. Let $\{Q_m\}$ be a Q-sequence of unitary operators. Then (up to reordering of the basis)*

$$Q_m^* A Q_m \longrightarrow \bigoplus_{j=1}^{N} \lambda_j e_j \otimes e_j, \qquad \text{as } m \to \infty.$$

In this section, we will address the convergence of the IQR algorithm for normal operators under similar assumptions and prove an analogue of Theorem 9.2.1 in infinite dimensions (Theorem 9.2.9). As well as this, and for more general operators A that are not necessarily normal, we address block convergence (Theorem 9.2.13), relevant when the eigenvalues do not have distinct moduli, and convergence to dominant invariant subspaces (Theorem 9.2.15).

9.2.1 Preliminary definitions and results

To state and prove our theorems, we need some preliminary results. The reader only interested in the main results themselves is referred to §9.2.2. If A is a normal operator, we will use $\chi_S(A)$ to denote the indicator function of the set S defined via the functional calculus. Without loss of generality, we deal with the Hilbert space $\mathcal{H} = l^2(\mathbb{N})$ and the canonical orthonormal basis $\{e_j\}_{j\in\mathbb{N}}$. Our first set of results concerns the convergence of spanning sets under power iterations and is analogous to the finite-dimensional case. The following proposition can be found in [Han08b] and together with Lemma 9.2.6 below, these are the only results we will use from [Han08b].

Proposition 9.2.2 ([Han08b]). *Suppose that $A \in \mathcal{B}(\mathcal{H})$ is normal, is invertible and that $\mathrm{Sp}(A) = \omega \cup \Psi$ is a disjoint union such that $\omega = \{\lambda_i\}_{i=1}^N$ consists of finitely many isolated eigenvalues of A with $|\lambda_1| > |\lambda_2| > ... > |\lambda_N|$. Suppose further that $\sup\{|z| : z \in \Psi\} < |\lambda_N|$. Let $l \in \mathbb{N}$ and suppose that $\{\xi_i\}_{i=1}^l$ are linearly independent vectors in $\mathcal{H}$ such that $\{\chi_\omega(A)\xi_i\}_{i=1}^l$ are also linearly independent. Then*

(i) The vectors $\{A^k\chi_\omega(A)\xi_i\}_{i=1}^l$ are linearly independent and there exists an l-dimensional subspace $B \subset \mathrm{ran}\chi_\omega(A)$ such that

$$\mathrm{span}\{A^k\xi_i\}_{i=1}^l \to B, \quad \text{as } k \to \infty.$$

(ii) If

$$\mathrm{span}\{A^k\xi_i\}_{i=1}^{l-1} \to D \subset \mathcal{H}, \quad \text{as } k \to \infty,$$

where D is an $(l-1)$-dimensional subspace, then

$$\mathrm{span}\{A^k\xi_i\}_{i=1}^{l} \to D \oplus \mathrm{span}\{\xi\}, \quad \text{as } k \to \infty,$$

where $\xi \in \mathrm{ran}\chi_\omega(A)$ is an eigenvector of A.

In order to extend this proposition to describe rates of convergence and prove our main theorems, we need to describe the space B in more detail. This is done inductively as follows. The first step is to choose $\nu_{1,1} \in \{\lambda_i\}_{i=1}^N$ of maximum modulus such that

$$\mathrm{span}\{\chi_{\nu_{1,1}}(A)\xi_1\} \neq \{0\}.$$

We then let $\xi_{1,1}$ be a linear multiple of ξ_1 such that $\chi_{\nu_{1,1}}(A)\xi_{1,1}$ has norm one. Now suppose that at the m-th stage we have constructed vectors $\{\xi_{m,i}\}_{i=1}^m$ with the same linear span as $\{\xi_i\}_{i=1}^m$ and such that there exist $\{\nu_{m,j}\}_{j=1}^{s_m} \subset \{\lambda_i\}_{i=1}^N$ with the following properties. After reordering the vectors $\{\xi_{m,i}\}_{i=1}^m$ if necessary, there exist integers $0 = k_{m,0} < k_{m,1} < k_{m,2} < ... < k_{m,s_m} = m$ such that

(1) $|\nu_{m,s_m}| < |\nu_{m,s_m-1}| < ... < |\nu_{m,1}|$.

(2) $\chi_\lambda(A)\xi_{m,i} = 0$ if $i > k_{m,j}$ and $\lambda \in \{\lambda_i\}_{i=1}^N$ has $|\lambda| > |\nu_{m,j+1}|$.

(3) $\{\chi_{\nu_{m,j}}(A)\xi_{m,i}\}_{i=k_{m,j-1}+1}^{k_{m,j}}$ are orthonormal.

We seek to add the space spanned by the vector ξ_{m+1} whilst preserving these properties.

First we deal with (2). Let $\eta_{m+1} \in \{\lambda_i\}_{i=1}^N$ be of maximal modulus such that $\chi_{\{\lambda_1,...,\eta_{m+1}\}}(A)\xi_{m+1} \notin \operatorname{span}\{\chi_{\{\lambda_1,...,\eta_{m+1}\}}(A)\xi_j\}_{j=1}^m$. If $|\eta_{m+1}| < |\nu_{m,1}|$ then let $t(m+1)$ be maximal such that $|\eta_{m+1}| < \left|\nu_{m,t(m+1)}\right|$. We then choose complex numbers $\{a_{m,j}\}_{j=1}^{k_{m,t(m+1)}}$ such that writing

$$\tilde{\xi}_{m+1,m+1} = \xi_{m+1} + \sum_{j=1}^{k_{m,t(m+1)}} a_{m,j}\xi_{m,j}$$

we have that $\chi_\lambda(A)\tilde{\xi}_{m+1,m+1} = 0$ if $\lambda \in \{\lambda_i\}_{i=1}^N$ has $|\lambda| > |\eta_{m+1}|$. Note that by (2), (3) and the definition of η_{m+1}, the coefficients $a_{m,j}$ are determined uniquely in terms of $\{\xi_{m,i}\}_{i=1}^{k_{m,t(m+1)}}$. If $|\eta_{m+1}| \geq |\nu_{m,1}|$ then let $t(m+1) = 0$ and we set $\tilde{\xi}_{m+1,m+1} = \xi_{m+1}$. In this case we still have that $\chi_\lambda(A)\tilde{\xi}_{m+1,m+1} = 0$ if $\lambda \in \{\lambda_i\}_{i=1}^N$ has $|\lambda| > |\eta_{m+1}|$.

We then define $\xi_{m+1,j} = \xi_{m,j}$ for $1 \leq j \leq m$ and now deal with (3). If $\eta_{m+1} \notin \{\nu_{m,j}\}_{j=1}^{s_m}$ then let $\xi_{m+1,m+1}$ be a linear multiple of $\tilde{\xi}_{m+1,m+1}$ such that $\chi_{\eta_{m+1}}(A)\xi_{m+1,m+1}$ has norm 1 and we let $\{\nu_{m+1,j}\}_{j=1}^{s_m+1}$ be a reordering of $\{\nu_{m,j}\}_{j=1}^{s_m} \cup \{\eta_{m+1}\}$. Otherwise, we have $\eta_{m+1} = \nu_{m,t(m+1)+1}$ and we apply Gram–Schmidt to

$$\{\chi_{\nu_{m,t(m+1)+1}}(A)\xi_{m+1,i}\}_{i=k_{m,t(m+1)}+1}^{k_{m,t(m+1)+1}} \cup \{\chi_{\nu_{m,t(m+1)+1}}(A)\tilde{\xi}_{m+1,m+1}\}$$

(without changing $\{\xi_{m+1,i}\}_{i=k_{m,t(m+1)}+1}^{k_{m,t(m+1)+1}}$). Note that by (2) and the definition of η_{m+1} these vectors are linearly independent. This gives $\xi_{m+1,m+1}$ such that

$$\{\chi_{\nu_{m,t(m+1)+1}}(A)\xi_{m+1,i}\}_{i=k_{m,t(m+1)}+1}^{k_{m,t(m+1)+1}} \cup \{\chi_{\nu_{m,t(m+1)+1}}(A)\xi_{m+1,m+1}\}$$

are orthonormal and $\chi_\lambda(A)\xi_{m+1,m+1} = 0$ if $\lambda \in \{\lambda_i\}_{i=1}^N$ has $|\lambda| > \left|\nu_{m,t(m+1)+1}\right|$. After reordering indices if necessary, we see that (1)–(3) now hold for $m+1$.

After l steps the above process terminates giving a new basis $\{\tilde{\xi}_i\}_{i=1}^l = \{\xi_{l,i}\}_{i=1}^l$ for $\operatorname{span}\{\xi_i\}_{i=1}^l$ along with $\{\nu_j\}_{j=1}^n = \{\nu_{l,j}\}_{j=1}^n \subset \{\lambda_i\}_{i=1}^N$ and $0 = k_0 < k_1 < k_2 < ... < k_n = l$ such that

(i) $|\nu_n| < |\nu_{n-1}| < ... < |\nu_1|$.

(ii) $\chi_\lambda(A)\tilde{\xi}_i = 0$ if $i > k_j$ and $\lambda \in \{\lambda_i\}_{i=1}^N$ has $|\lambda| > |\nu_{j+1}|$.

(iii) $\{\chi_{\nu_j}(A)\tilde{\xi}_i\}_{i=k_{j-1}+1}^{k_j}$ are orthonormal (and hence $\|\tilde{\xi}_i\|^2 \geq 1$).

The subspace B can then be described as

$$B = \bigoplus_{j=1}^n \operatorname{span}\{\chi_{\nu_j}(A)\tilde{\xi}_i\}_{i=k_{j-1}+1}^{k_j}.$$

Definition 9.2.3. *With respect to the above construction we define the following:*

$$E_j := \operatorname{span}\{\chi_{\nu_j}(A)\tilde{\xi}_i\}_{i=k_{j-1}+1}^{k_j}, \quad Z(A, \{\xi_j\}_{j=1}^l) := \left(\sum_{i=1}^l (\|\tilde{\xi}_i\|^2 - 1)\right)^{\frac{1}{2}}. \tag{9.2.1}$$

Since the Gram–Schmidt process is defined uniquely up to phases, we see that $Z(A, \{\xi_j\}_{j=1}^l)$ is well-defined. The above construction also shows that if $\{\chi_\omega(A)\xi_i\}_{i=1}^{l+1}$ are linearly independent then

$$Z(A, \{\xi_j\}_{j=1}^{l+1}) \geq Z(A, \{\xi_j\}_{j=1}^l).$$

We can now prove the following refinement of Proposition 9.2.2:

Proposition 9.2.4. *Suppose the assumptions of Proposition 9.2.2 hold. Let $J \leq N$ be minimal such that $\{\chi_{\{\lambda_1,...,\lambda_J\}}(A)\xi_i\}_{i=1}^l$ are linearly independent. Set*

$$\rho = \sup\{|z| : z \in \Psi \cup \{\lambda_{J+1}, ..., \lambda_N\}\},$$
$$r = \max\{|\lambda_2/\lambda_1|, ..., |\lambda_J/\lambda_{J-1}|, \rho/|\lambda_J|\}.$$

Then $r < 1$ and $\delta(B, \operatorname{span}\{A^k\xi_i\}_{i=1}^l) \leq Z(A, \{\xi_j\}_{j=1}^l)r^k$. Since the spaces are l-dimensional, it follows from (9.1.2) that we have the convergence rate

$$\hat{\delta}(B, \operatorname{span}\{A^k\xi_i\}_{i=1}^l) \leq Z(A, \{\xi_j\}_{j=1}^l)l^{\frac{1}{2}}r^k.$$

Proof. Consider the subspaces

$$E_j^k = \operatorname{span}\{A^k\tilde{\xi}_i\}_{i=k_{j-1}+1}^{k_j}.$$

Let $\zeta = \sum_{i=k_{j-1}+1}^{k_j} \alpha_i\chi_{\nu_j}(A)\tilde{\xi}_i \in E_j$ be a unit vector (hence $\sum_{i=k_{j-1}+1}^{k_j} |\alpha_i|^2 = 1$) and consider

$$\eta_k = \sum_{i=k_{j-1}+1}^{k_j} \alpha_i A^k\tilde{\xi}_i/\nu_j^k \in E_j^k.$$

By construction, we have for any such $\tilde{\xi}_i$ in the above sum that

$$\tilde{\xi}_i = (\chi_{\nu_j}(A) + \chi_{\theta_j}(A))\tilde{\xi}_i, \qquad \theta_j = \{\lambda \in \operatorname{Sp}(A) : |\lambda| < |\nu_j|\}.$$

This gives $A^k\tilde{\xi}_i = \nu_j^k\chi_{\nu_j}(A)\tilde{\xi}_{j,i} + A^k\chi_{\theta_j}(A)\tilde{\xi}_i$. Now, by the assumption on $\operatorname{Sp}(A)$, we have

$$\rho_j = \sup\{|z| : z \in \theta_j\} < |\nu_j|.$$

Thus, since

$$\|A^k\chi_{\theta_j}(A)\tilde{\xi}_i\|/|\nu_j^k| < |\rho_j/\nu_j|^k\|\chi_{\theta_j}(A)\tilde{\xi}_i\|,$$

we have

$$\|\zeta - \eta_k\| \leq |\rho_j/\nu_j|^k \sum_{i=k_{j-1}+1}^{k_j} |\alpha_i| \, \|\chi_{\theta_j}(A)\tilde{\xi}_i\| \leq \left(\sum_{i=k_{j-1}+1}^{k_j} (\|\tilde{\xi}_i\|^2 - 1)\right)^{\frac{1}{2}} r^k.$$

Here we have used Hölder's inequality together with the fact that $\|\chi_{\theta_j}(A)\tilde{\xi}_i\|^2 = \|\tilde{\xi}_i\|^2 - 1$ by orthonormality of $\{\chi_{\nu_j}(A)\tilde{\xi}_i\}_{i=k_{j-1}+1}^{k_j}$. The right-hand side gives an upper bound for $\delta(E_j, E_j^k)$. Analogous rates of convergence hold for the other subspaces and from (9.1.3) we have

$$\delta(B, \operatorname{span}\{A^k\tilde{\xi}_i\}_{i=1}^l) \leq Z(A, \{\xi_j\}_{j=1}^l)r^k,$$

since the spaces E_j are orthogonal. □

For the rest of this section, we shall assume the following:

(A1) $A \in \mathcal{B}(\mathcal{H})$ is an invertible normal operator and $\{e_j\}_{j\in\mathbb{N}}$ an orthonormal basis for $\mathcal{H}$. $\{Q_k\}$ and $\{R_k\}$ are Q- and R-sequences of A with respect to the basis $\{e_j\}_{j\in\mathbb{N}}$.

(A2) $\mathrm{Sp}(A) = \omega \cup \Psi$ such that $\omega \cap \Psi = \emptyset$ and $\omega = \{\lambda_i\}_{i=1}^N$, where the λ_is are isolated eigenvalues with (possibly infinite) multiplicity m_i. Let $M = m_1 + ... + m_N = \dim(\mathrm{ran}\chi_\omega(A))$ and suppose that $|\lambda_1| > \ldots > |\lambda_N|$. Suppose further that $\sup\{|\theta| : \theta \in \Psi\} < |\lambda_N|$.

To apply Propositions 9.2.2 and 9.2.4 to prove the main result Theorem 9.2.9, we need to take care of the case that some of the e_j may have $\chi_\omega(A)e_j = 0$.

Definition 9.2.5. *Suppose that (A1) and (A2) hold and let $K \in \mathbb{N} \cup \{\infty\}$ be minimal with the property that $\dim(\mathrm{span}\{\chi_\omega(A)e_j\}_{j=1}^K) = M$. Define*

$$\begin{aligned}\Lambda_\omega &= \{e_j : \chi_\omega(A)e_j \neq 0, j \leq K\},\\ \Lambda_\Psi &= \{e_j : \chi_\omega(A)e_j = 0, j \leq K\},\\ \tilde{\Lambda}_\omega &= \{e_j \in \Lambda_\omega : \chi_\omega(A)e_j \in \mathrm{span}\{\chi_\omega(A)e_i\}_{i=1}^{j-1}\}.\end{aligned}$$

Define also the corresponding subset $\{\hat{e}_j\}_{j=1}^M \subset \{e_j\}_{j=1}^K$ such that $\{\hat{e}_j\}_{j=1}^M = \Lambda_\omega \setminus \tilde{\Lambda}_\omega$ and such that upon writing $\hat{e}_j = e_{p_j}$, the p_j are increasing.

Note that we have the following decomposition of A into

$$A = \left(\sum_{j=1}^M \lambda_{c_j}\, \xi_j \otimes \bar{\xi}_j\right) \oplus \chi_\Psi(A)A, \quad \lambda_{c_j} \in \omega,$$

where $\{\xi_j\}_{j=1}^M$ is an orthonormal set of eigenvectors of A. The following simple lemma extends Lemma 39 in [Han08b] to infinite M but the proof is verbatim so omitted.

Lemma 9.2.6. *If $e_m \in \Lambda_\Psi \cup \tilde{\Lambda}_\omega$, then*

$$\mathrm{span}\{\chi_\omega(A)q_{k,j}\}_{j=1}^m = \mathrm{span}\{\chi_\omega(A)\hat{q}_{k,j}\}_{j=1}^{s(m)}, \quad q_{k,j} = Q_k e_j, \quad \hat{q}_{k,j} = Q_k\hat{e}_j,$$

where $s(m)$ is the largest integer such that $\{\hat{e}_j\}_{j=1}^{s(m)} \subset \{e_j\}_{j=1}^m$.

The following theorem is the key step of the proof of Theorem 9.2.9 and concerns convergence to the eigenvectors of A.

Theorem 9.2.7. *Assume (A1) and (A2) and define*

$$\rho = \sup\{|z| : z \in \Psi\}, \quad r = \max\{|\lambda_2/\lambda_1|, ..., |\lambda_N/\lambda_{N-1}|, \rho/|\lambda_N|\}.$$

Then there exists a collection of orthonormal eigenvectors $\{\hat{q}_j\}_{j=1}^M \subset \mathrm{ran}\chi_\omega(A)$ of A and collections of constants $\mathcal{A}(m)$, $\mathcal{B}(j)$ and $\mathcal{C}(\mu)$ such that

(a) If $e_m \in \Lambda_\Psi \cup \tilde{\Lambda}_\omega$ and μ is maximal with $p_\mu < m$ (recall that $\hat{e}_j = e_{p_j}$), then we have

$$\|\chi_\omega(A)q_{k,m}\| \leq \mathcal{A}(m)Z(A, \{\hat{e}_j\}_{j=1}^\mu)r^k. \tag{9.2.2}$$

In the case that $m < p_1$, we interpret this as $\|\chi_\omega(A)q_{k,m}\| = 0$ which holds from Lemma 9.2.6.

(b) For any $j < M+1$,

$$\hat{\delta}(\text{span}\{\hat{q}_j\}, \text{span}\{\hat{q}_{k,j}\}) \leq \mathcal{B}(j) Z(A, \{\hat{e}_i\}_{i=1}^{j}) r^k.$$

(c) For any $\mu < M+1$,

$$\delta(\text{span}\{\hat{q}_{j,k}\}_{j=1}^{\mu}, \text{span}\{\hat{q}_j\}_{j=1}^{\mu}) \leq \mathcal{C}(\mu) Z(A, \{\hat{e}_j\}_{j=1}^{\mu}) r^k$$

and hence

$$\hat{\delta}(\text{span}\{\hat{q}_{j,k}\}_{j=1}^{\mu}, \text{span}\{\hat{q}_j\}_{j=1}^{\mu}) \leq \mu^{\frac{1}{2}} \mathcal{C}(\mu) Z(A, \{\hat{e}_j\}_{j=1}^{\mu}) r^k.$$

Here, as in Lemma 9.2.6, $q_{k,j} = Q_k e_j$ and $\hat{q}_{k,j} = Q_k \hat{e}_j$. Finally, if M is finite then we must have $\text{span}\{\hat{q}_j\}_{j=1}^{M} = \text{ran}\chi_\omega(A)$.

We will provide an inductive proof of Theorem 9.2.7 which requires the following for the inductive step of part (a).

Lemma 9.2.8. *Assume the conditions in the statement of Theorem 9.2.7. Suppose also that (b) in Theorem 9.2.7 holds for $j = 1, ..., \mu$ and that (c) holds for a given $\mu < M$. Let $e_{p_{\mu+1}} = \hat{e}_{\mu+1}$, then if $e_m \in \Lambda_\Psi \cup \tilde{\Lambda}_\omega$, where $m < p_{\mu+1}$, (9.2.2) also holds with*

$$\mathcal{A}(m) = \left\{ \sum_{j=1}^{\mu} \left[\mathcal{C}(\mu) + \mathcal{B}(j) \right]^2 \right\}^{\frac{1}{2}} + \mathcal{C}(\mu).$$

Proof. First note that from (9.1.7), invertibility of A and the fact that $\{\chi_\omega(A)\hat{e}_j\}_{j=1}^{\mu}$ are linearly independent, it must hold that $\{\chi_\omega(A)\hat{q}_{k,j}\}_{j=1}^{\mu}$ are linearly independent also. Then by using the assumptions stated and the fact that $\chi_\omega(A)\hat{q}_j = \hat{q}_j$ we have

$$\begin{aligned} \delta(\text{span}\{\chi_\omega(A)\hat{q}_{k,j}\}_{j=1}^{\mu}, \text{span}\{\hat{q}_j\}_{j=1}^{\mu}) &\leq \delta(\text{span}\{\hat{q}_{k,j}\}_{j=1}^{\mu}, \text{span}\{\hat{q}_j\}_{j=1}^{\mu}) \\ &\leq \mathcal{C}(\mu) Z(A, \{\hat{e}_j\}_{j=1}^{\mu}) r^k. \end{aligned}$$

Also, we have that $s(m) \leq \mu$ and Lemma 9.2.6 implies

$$\text{span}\{\chi_\omega(A)q_{k,j}\}_{j=1}^{m} = \text{span}\{\chi_\omega(A)\hat{q}_{k,j}\}_{j=1}^{s(m)} \subset \text{span}\{\chi_\omega(A)\hat{q}_{k,j}\}_{j=1}^{\mu}.$$

Using the fact that $\|\chi_\omega(A)q_{k,m}\| \leq 1$ and the definition of δ (along with the fact that $\text{span}\{\hat{q}_j\}_{j=1}^{\mu}$ is finite-dimensional), it follows that there exists some $v_k = \sum_{j=1}^{\mu} \beta_{j,k}\hat{q}_j \in \text{span}\{\hat{q}_j\}_{j=1}^{\mu}$ with $\|v_k\| \leq 1$ and

$$\|\chi_\omega(A)q_{k,m} - v_k\| \leq \mathcal{C}(\mu) Z(A, \{\hat{e}_j\}_{j=1}^{\mu}) r^k. \tag{9.2.3}$$

We also have from assumption (b) that

$$|\langle \chi_\omega(A)q_{k,m}, \hat{q}_j \rangle| = |\langle q_{k,m}, \hat{q}_j \rangle| \leq \mathcal{B}(j) Z(A, \{\hat{e}_i\}_{i=1}^{j}) r^k + |\langle q_{k,m}, \hat{q}_{k,j} \rangle| = \mathcal{B}(j) Z(A, \{\hat{e}_i\}_{i=1}^{j}) r^k,$$

since $q_{k,m}$ is orthogonal to $\hat{q}_{k,j}$. This together with (9.2.3) gives $|\beta_{j,k}| \leq \left[\mathcal{C}(\mu) + \mathcal{B}(j) \right] Z(A, \{\hat{e}_j\}_{j=1}^{\mu}) r^k$. Hence we must have

$$\|v_k\| \leq \left\{ \sum_{j=1}^{\mu} \left[\mathcal{C}(\mu) + \mathcal{B}(j) \right]^2 \right\}^{\frac{1}{2}} Z(A, \{\hat{e}_j\}_{j=1}^{\mu}) r^k.$$

Using (9.2.3) again then gives the result. Note that we have used orthonormality of $\{\hat{q}_j\}_{j=1}^{\mu}$ which will be proven as part of the induction. □

Proof of Theorem 9.2.7: We begin with the initial step of the induction for (b) and (c). Note that (a) trivially holds by construction with $\mathcal{A}(m) = 0$ for any $m < p_1$ where $e_{p_1} = \hat{e}_1$ and this provides the initial step for (a).

By Propositions 9.2.2 and 9.2.4, there exists a unit eigenvector $\hat{q}_1 \in \mathrm{ran}\chi_\omega(A)$ such that

$$\delta(\mathrm{span}\{\hat{q}_1\}, \mathrm{span}\{A^k\hat{e}_1\}) \leq Z(A, \{\hat{e}_1\})r^k.$$

Since $\mathrm{span}\{A^k\hat{e}_1\} \subset \mathrm{span}\{A^k e_i\}_{i=1}^{p_1}$, this implies that

$$\delta(\mathrm{span}\{\hat{q}_1\}, \mathrm{span}\{A^k e_i\}_{i=1}^{p_1}) \leq Z(A, \{\hat{e}_1\})r^k.$$

Thus, from (9.1.7) it follows that

$$\delta(\mathrm{span}\{\hat{q}_1\}, \mathrm{span}\{q_{k,i}\}_{i=1}^{p_1}) = \delta(\mathrm{span}\{\hat{q}_1\}, \mathrm{span}\{A^k e_i\}_{i=1}^{p_1}) \leq Z(A, \{\hat{e}_1\})r^k. \tag{9.2.4}$$

Note that $\{q_{k,i}\}_{i=1}^{p_1}$ are orthonormal (recall that Q_k is unitary) and hence by (9.2.4) there exists some coefficients $\alpha_{k,i}$ with $\sum_{i=1}^{p_1} |\alpha_{k,i}|^2 \leq 1$ such that defining $\tilde{\eta}_k = \sum_{i=1}^{p_1} \alpha_{k,i} q_{k,i}$ we have

$$\|\hat{q}_1 - \tilde{\eta}_k\| \leq Z(A, \{\hat{e}_1\})r^k.$$

If $e_m \in \Lambda_\Psi \cup \tilde{\Lambda}_\omega$, where $m < p_1$ then by Lemma 9.2.6 $\langle q_{k,m}, \hat{q}_1 \rangle = 0$. It follows that we must have

$$\delta(\mathrm{span}\{\hat{q}_1\}, \mathrm{span}\{\hat{q}_{k,1}\}) \leq \|\hat{q}_1 - \alpha_{k,p_1}\hat{q}_{k,1}\| \leq Z(A, \{\hat{e}_1\})r^k.$$

Hence we can take $\mathcal{B}(1) = 1$ and $\mathcal{C}(1) = 1$ in (b) and (c) respectively, which completes the initial step.

For the induction step we will argue simultaneously for (a), (b) and (c) using induction on μ. Suppose that (a) holds for $m < p_\mu$ with $e_{p_\mu} = \hat{e}_\mu$ together with (b) and (c) for $j \leq \mu$ and some $\mu < M$. Let $e_{p_{\mu+1}} = \hat{e}_{\mu+1}$ then we can use Lemma 9.2.8 to extend (a) to all $m < p_{\mu+1}$ and this provides the step for (a). For (b), we note that Propositions 9.2.2 and 9.2.4 imply that

$$\delta(\mathrm{span}\{\hat{q}_i\}_{i=1}^{\mu} \oplus \mathrm{span}\{\xi\}, \mathrm{span}\{A^k\hat{e}_i\}_{i=1}^{\mu+1},) \leq Z(A, \{\hat{e}_j\}_{j=1}^{\mu+1})r^k, \quad \xi \in \mathrm{ran}\chi_\omega(A),$$

where ξ is a unit eigenvector of A. We may also assume without loss of generality that ξ is orthogonal to $\hat{q}_j$ for $j = 1, ..., \mu$. As before, since $\mathrm{span}\{A^k\hat{e}_i\}_{i=1}^{\mu+1} \subset \mathrm{span}\{A^k e_i\}_{i=1}^{p_{\mu+1}}$ we have

$$\delta(\mathrm{span}\{\hat{q}_i\}_{i=1}^{\mu} \oplus \mathrm{span}\{\xi\}, \mathrm{span}\{A^k e_i\}_{i=1}^{p_{\mu+1}}) \leq Z(A, \{\hat{e}_j\}_{j=1}^{\mu+1})r^k,$$

and hence by invertibility of A

$$\begin{aligned}\delta(\mathrm{span}\{\hat{q}_i\}_{i=1}^{\mu} \oplus \mathrm{span}\{\xi\}, \mathrm{span}\{q_{k,i}\}_{i=1}^{p_{\mu+1}}) &= \delta(\mathrm{span}\{\hat{q}_i\}_{i=1}^{\mu} \oplus \mathrm{span}\{\xi\}, \mathrm{span}\{A^k e_i\}_{i=1}^{p_{\mu+1}}) \\ &\leq Z(A, \{\hat{e}_j\}_{j=1}^{\mu+1})r^k.\end{aligned}$$

Again, using orthonormality of $\{q_{k,i}\}_{i=1}^{p_{\mu+1}}$, there exists some coefficients $\alpha_{k,i}$ with $\sum_{i=1}^{p_{\mu+1}} |\alpha_{k,i}|^2 \leq 1$ such that, defining $\tilde{\eta}_k = \sum_{i=1}^{p_{\mu+1}} \alpha_{k,i} q_{k,i}$, we have

$$\|\xi - \tilde{\eta}_k\| \leq Z(A, \{\hat{e}_j\}_{j=1}^{\mu+1})r^k. \tag{9.2.5}$$

If $e_m \in \Lambda_\Psi \cup \tilde{\Lambda}_\omega$, where $m < p_{\mu+1}$ then as shown above we have

$$|\langle q_{k,m}, \xi \rangle| = |\langle \chi_\omega(A) q_{k,m}, \xi \rangle| \leq \mathcal{A}(m) Z(A, \{\hat{e}_j\}_{j=1}^{\mu})r^k \leq \mathcal{A}(m) Z(A, \{\hat{e}_j\}_{j=1}^{\mu+1})r^k.$$

Taking the inner product of $\xi - \tilde{\eta}_k$ with $q_{k,m}$ and using (9.2.5) together with the orthonormality of the $q_{k,j}$s, it follows that $|\alpha_{k,m}| \leq \big(\mathcal{A}(m)+1\big)Z(A, \{\hat{e}_j\}_{j=1}^{\mu+1})r^k$. Similarly, if $j \leq \mu$ then for any $c \in \mathbb{C}$

$$|\langle \hat{q}_{k,j}, \xi\rangle| \leq |\langle c\hat{q}_j, \xi\rangle| + |c\hat{q}_j - \hat{q}_{k,j}| = |c\hat{q}_j - \hat{q}_{k,j}|\,,$$

since ξ is orthogonal to $\hat{q}_j$. Minimising over c, we can bound this by $\mathcal{B}(j)Z(A, \{\hat{e}_j\}_{j=1}^{\mu})r^k$. In the same way, it then follows that $|\alpha_{k,p_j}| \leq \big(\mathcal{B}(j)+1\big)Z(A, \{\hat{e}_j\}_{j=1}^{\mu+1})r^k$ where $\hat{e}_j = e_{p_j}$. Together, these imply that

$$\big\|\xi - \alpha_{k,p_{\mu+1}}\hat{q}_{k,\mu+1}\big\| \leq \left[1 + \left\{\sum_{m=1, e_m \in \Lambda_\Psi \cup \tilde{\Lambda}_\omega}^{p_{\mu+1}} \big[\mathcal{A}(m)+1\big]^2 + \sum_{j=1}^{\mu}\big[\mathcal{B}(j)+1\big]^2\right\}^{\frac{1}{2}}\right] Z(A, \{\hat{e}_j\}_{j=1}^{\mu+1})r^k.$$

To finish the inductive step, we define $\hat{q}_{\mu+1} = \xi$. Recall that ξ is orthogonal to any $\hat{q}_l$ with $l \leq \mu$. Hence it follows that $\{\hat{q}_i\}_{i=1}^{\mu+1}$ are orthonormal and we can take

$$\mathcal{B}(\mu+1) = 1 + \left\{\sum_{m=1, e_m \in \Lambda_\Psi \cup \tilde{\Lambda}_\omega}^{p_{\mu+1}} \big[\mathcal{A}(m)+1\big]^2 + \sum_{j=1}^{\mu}\big[\mathcal{B}(j)+1\big]^2\right\}^{\frac{1}{2}}$$

in (b). For the induction step for (c), the fact that $\{\hat{q}_{k,i}\}_{i=1}^{\mu+1}$ are orthonormal and (9.1.3) imply we can take

$$\mathcal{C}(\mu+1) = \left(\sum_{j=1}^{\mu+1}\mathcal{B}(j)^2\right)^{\frac{1}{2}}.$$

Finally, if M is finite we demonstrate that $\operatorname{span}\{\hat{q}_j\}_{j=1}^{M} = \operatorname{span}\{\xi_j\}_{j=1}^{M}$. Since the $\{\hat{q}_i\}_{i=1}^{M}$ are orthogonal and are eigenvectors of $\sum_{j=1}^{M}\lambda_{c_j}\,\xi_j \otimes \bar{\xi}_j$, it follows that $\operatorname{span}\{\hat{q}_j\}_{j=1}^{M} = \operatorname{span}\{\xi_j\}_{j=1}^{M} = \operatorname{ran}\chi_\omega(A)$. □

9.2.2 Main results

Our first result generalises Theorem 9.2.1 to infinite dimensions and relies on Theorem 9.2.7 (which concerns convergence to eigenvectors).

Theorem 9.2.9 (Convergence theorem for normal operators in infinite dimensions)**.** *Let $A \in \mathcal{B}(l^2(\mathbb{N}))$ be an invertible normal operator with $\operatorname{Sp}(A) = \omega \cup \Psi$ and $\omega = \{\lambda_i\}_{i=1}^{N}$, where the λ_i's are isolated eigenvalues with (possibly infinite) multiplicity m_i satisfying $|\lambda_1| > \ldots > |\lambda_N|$. Suppose further that $\sup\{|\theta| : \theta \in \Psi\} < |\lambda_N|$, and let $\{e_j\}_{j\in\mathbb{N}}$ be the canonical orthonormal basis. Let $\{Q_n\}_{n\in\mathbb{N}}$ and $\{R_n\}_{n\in\mathbb{N}}$ be Q- and R-sequences of A with respect to $\{e_j\}_{j\in\mathbb{N}}$. Let $\{\hat{e}_j\}_{j=1}^{M} \subset \{e_j\}_{j\in\mathbb{N}}$, where $M = m_1 + \ldots + m_N$, be the subset described in Definition 9.2.5 and Theorem 9.2.7, i.e. $\operatorname{span}\{Q_k\hat{e}_j\} \to \operatorname{span}\{\hat{q}_j\}$ where $\{\hat{q}_j\}_{j=1}^{M} \subset \operatorname{ran}\chi_\omega(A)$ is a collection of orthonormal eigenvectors of A and if $e_j \notin \{\hat{e}_j\}_{j=1}^{M}$, then $\chi_\omega(A)Q_k e_j \to 0$. Then:*

(i) Every subsequence of $\{Q_n^* A Q_n\}_{n\in\mathbb{N}}$ has a convergent subsequence $\{Q_{n_k}^* A Q_{n_k}\}_{k\in\mathbb{N}}$ such that

$$Q_{n_k}^* A Q_{n_k} \xrightarrow{WOT} \left(\bigoplus_{j=1}^{M}\langle A\hat{q}_j, \hat{q}_j\rangle \hat{e}_j \otimes \hat{e}_j\right) \bigoplus \sum_{j\in\Theta} \xi_j \otimes e_j,$$

as $k \to \infty$, where

$$\Theta = \{j : e_j \notin \{\hat{e}_l\}_{l=1}^{M}\}, \quad \xi_j \in \operatorname{cl}\left(\operatorname{span}\{e_i\}_{i\in\Theta}\right)$$

and only $\sum_{j\in\Theta}\xi_j \otimes e_j$ depends on the choice of subsequence. Furthermore, if A has only finitely many non-zero entries in each column then we can replace WOT convergence by SOT convergence.

(ii) We have the following convergence of sections:

$$\widehat{P}_M Q_n^* A Q_n \widehat{P}_M \xrightarrow{\text{SOT}} \bigoplus_{j=1}^{M} \langle A\hat{q}_j, \hat{q}_j\rangle \hat{e}_j \otimes \hat{e}_j, \qquad \text{as } n \to \infty,$$

where $\widehat{P}_M$ denotes the orthogonal projection onto $\operatorname{cl}\left(\operatorname{span}\{\hat{e}_j\}_{j=1}^{M}\right)$. *Furthermore, if we define*

$$\rho = \sup\{|z| : z \in \Psi\}, \quad r = \max\{|\lambda_2/\lambda_1|, ..., |\lambda_N/\lambda_{N-1}|, \rho/|\lambda_N|\}$$

then $r < 1$ and for any fixed $x \in \operatorname{span}\{\hat{e}_j\}_{j=1}^{M}$ we have the following rate of convergence

$$\left\| \widehat{P}_M Q_n^* A Q_n \widehat{P}_M x - \left(\bigoplus_{j=1}^{M} \langle A\hat{q}_j, \hat{q}_j\rangle \hat{e}_j \otimes \hat{e}_j \right) x \right\| = \mathcal{O}(r^n), \quad \text{as } n \to \infty. \tag{9.2.6}$$

If M is finite then we can write (after possibly reordering)

$$\bigoplus_{j=1}^{M} \langle A\hat{q}_j, \hat{q}_j\rangle \hat{e}_j \otimes \hat{e}_j = \bigoplus_{k=1}^{N} \left(\lambda_k \bigoplus_{j=1+\sum_{l<k} m_l}^{\sum_{l\le k} m_l} \hat{e}_j \otimes \hat{e}_j \right), \tag{9.2.7}$$

and in part (ii) we have the rate of convergence

$$\left\| \widehat{P}_M Q_n^* A Q_n \widehat{P}_M - \bigoplus_{j=1}^{M} \langle A\hat{q}_j, \hat{q}_j\rangle \hat{e}_j \otimes \hat{e}_j \right\| = \mathcal{O}(r^n), \quad \text{as } n \to \infty. \tag{9.2.8}$$

If $\{\chi_\omega(A)e_l\}_{l=1}^{M}$ are linearly independent, then we can take $\hat{e}_j = e_j$.

Remark 9.2.10. *What Theorem 9.2.9 essentially says is that if we take the n-th iteration of the IQR algorithm and truncate to an $m \times m$ matrix (i.e. $P_m Q_n^* A Q_n P_m$) then, as n grows, the eigenvalues of this matrix will converge to the extremal parts of the spectrum of A. In particular, the theorem suggests that the IQR algorithm can locate the extremal parts of the spectrum.*

Proof of Theorem 9.2.9: To prove (i), since a closed ball in $\mathcal{B}(l^2(\mathbb{N}))$ is weakly sequentially compact, it follows that any subsequence of $\{Q_n^* A Q_n\}_{n\in\mathbb{N}}$ must have a weakly convergent subsequence which we denote by $\{Q_{n_k}^* A Q_{n_k}\}_{k\in\mathbb{N}}$. In particular, there exists a $W \in \mathcal{B}(l^2(\mathbb{N}))$ such that

$$Q_{n_k}^* A Q_{n_k} \xrightarrow{\text{WOT}} W, \qquad k \to \infty.$$

Let $\widehat{P}_M$ denote the projection onto $\operatorname{cl}\left(\operatorname{span}\{\hat{e}_j\}_{j=1}^{M}\right)$. Note that part (i) of the theorem will follow if we can show that

$$\widehat{P}_M W \widehat{P}_M = \bigoplus_{j=1}^{M} \langle A\hat{q}_j, \hat{q}_j\rangle \hat{e}_j \otimes \hat{e}_j, \tag{9.2.9}$$

and

$$\widehat{P}_M^{\perp} W \widehat{P}_M = 0, \quad \widehat{P}_M W \widehat{P}_M^{\perp} = 0.$$

We will indeed show this, and we start by observing that, due to the weak convergence and the standard functional calculus, we have

$$\langle W\hat{e}_j, e_i\rangle = \lim_{k\to\infty} \langle A Q_{n_k} \hat{e}_j, \chi_\omega(A) Q_{n_k} e_i\rangle + \lim_{k\to\infty} \langle A Q_{n_k} \hat{e}_j, \chi_\Psi(A) Q_{n_k} e_i\rangle, \tag{9.2.10}$$

$$\langle W e_i, \hat{e}_j\rangle = \lim_{k\to\infty} \langle \chi_\omega(A) Q_{n_k} e_i, A^* Q_{n_k} \hat{e}_j\rangle + \lim_{k\to\infty} \langle A Q_{n_k} e_i, \chi_\Psi(A) Q_{n_k} \hat{e}_j\rangle. \tag{9.2.11}$$

We then have the following,

$$\chi_\omega(A)Q_n e_i \to 0, \quad n \to \infty, \quad i \in \Theta$$
$$\implies \begin{cases} \lim_{k\to\infty}\langle AQ_{n_k}\hat{e}_j, \chi_\omega(A)Q_{n_k}e_i\rangle = 0, & i \in \Theta, \\ \lim_{k\to\infty}\langle \chi_\omega(A)Q_{n_k}e_i, A^*Q_{n_k}\hat{e}_j\rangle = 0, & i \in \Theta, \end{cases} \tag{9.2.12}$$

$$\operatorname{span}\{Q_n\hat{e}_j\} \to \operatorname{span}\{\hat{q}_j\}, \quad n \to \infty, \quad A\hat{q}_j = \lambda\hat{q}_j,\ \lambda \in \omega,$$
$$\implies \begin{cases} \lim_{k\to\infty}\langle AQ_{n_k}\hat{e}_j, \chi_\Psi(A)Q_{n_k}e_i\rangle = 0, & i \in \mathbb{N}, \\ \lim_{k\to\infty}\langle AQ_{n_k}e_i, \chi_\Psi(A)Q_{n_k}\hat{e}_j\rangle = 0, & i \in \mathbb{N}, \\ \lim_{k\to\infty}\langle AQ_{n_k}\hat{e}_j, \chi_\omega(A)Q_{n_k}\hat{e}_l\rangle = \delta_{j,l}\lambda. \end{cases} \tag{9.2.13}$$

Thus, by (9.2.10), (9.2.12), (9.2.13) and Theorem 9.2.7 we get (9.2.9) and also that $\widehat{P}_M^{\perp} W \widehat{P}_M = 0$. Also, by (9.2.11), (9.2.12), (9.2.13) and Theorem 9.2.7 we get that $\widehat{P}_M W \widehat{P}_M^{\perp} = 0$. Note that in all of these cases, Theorem 9.2.7 implies that the rate of convergence is such that the difference between $\langle W\hat{e}_j, e_i\rangle$, $\langle We_i, \hat{e}_j\rangle$ and their limiting values is $\mathcal{O}(r^{n_k})$ (however, not necessarily uniformly over the indices). Now suppose that A has finitely many non-zero entries in each column. This can be described by a function $f : \mathbb{N} \to \mathbb{N}$ non-decreasing with $f(n) \geq n$ such that $\langle Ae_j, e_i\rangle = 0$ when $i > f(j)$ as in Definition 9.3.1. Proposition 9.3.2 shows that this is preserved under the iteration in the IQR algorithm, i.e. $Q_{n_k}^* A Q_{n_k}$ also has this property. So let $x \in l^2(\mathbb{N})$ and $\epsilon > 0$. Choose y of finite support such that $\|x - y\| \leq \epsilon$. It is then clear that $\|Q_{n_k}^* A Q_{n_k} y - Wy\| \to 0$ as $n_k \to \infty$ (since we only require convergence in finitely many entries). Hence

$$\limsup_{n_k\to\infty} \|Q_{n_k}^* A Q_{n_k} x - Wx\| \leq (\|A\| + \|W\|)\epsilon.$$

Since $\epsilon > 0$ and x were arbitrary, we have $Q_{n_k}^* A Q_{n_k} \xrightarrow{\text{SOT}} W$.

To prove (ii), suppose that $x \in \operatorname{span}\{\hat{e}_j\}_{j=1}^M$, then x can be written as

$$x = \sum_{j=1}^M x_j \hat{e}_j,$$

with at most finitely many non-zero x_j. We have that $\hat{\delta}(\operatorname{span}\{Q_n\hat{e}_j\}, \operatorname{span}\{\hat{q}_j\}) = \mathcal{O}(r^n)$ and hence there exists some $a_{n,j}$ of unit modulus such that $\|Q_n\hat{e}_j - a_{n,j}\hat{q}_j\| = \mathcal{O}(r^n)$. Since Q_n is unitary, we then have

$$\left\| \widehat{P}_M Q_n^* A Q_n \widehat{P}_M x - \left(\bigoplus_{j=1}^M \langle A\hat{q}_j, \hat{q}_j\rangle \hat{e}_j \otimes \hat{e}_j \right) x \right\| \leq \left\| Q_n^* A Q_n \widehat{P}_M x - \left(\bigoplus_{j=1}^M \langle A\hat{q}_j, \hat{q}_j\rangle \hat{e}_j \otimes \hat{e}_j \right) Q_n^* Q_n x \right\|$$
$$= \left\| \sum_{j=1}^M x_j (A - \langle A\hat{q}_j, \hat{q}_j\rangle I) Q_n \hat{e}_j \right\| = \mathcal{O}(r^n),$$

where we use in the last line the fact that A is bounded. We therefore have convergence on $\operatorname{span}\{\hat{e}_j\}_{j=1}^M$, and, since the operators are uniformly bounded, we must have convergence on $\operatorname{cl}\left(\operatorname{span}\{\hat{e}_j\}_{j=1}^M\right)$ which implies that

$$\widehat{P}_M Q_n^* A Q_n \widehat{P}_M \xrightarrow{\text{SOT}} \bigoplus_{j=1}^M \langle A\hat{q}_j, \hat{q}_j\rangle \hat{e}_j \otimes \hat{e}_j, \qquad \text{as } n \to \infty.$$

For the last parts, suppose that M is finite. Theorem 9.2.7 then implies (9.2.7) after a possible reordering. The rate of convergence in (9.2.6) also implies that

$$\left\| \widehat{P}_M Q_n^* A Q_n \widehat{P}_M - \bigoplus_{j=1}^M \langle A\hat{q}_j, \hat{q}_j\rangle \hat{e}_j \otimes \hat{e}_j \right\| = \mathcal{O}(r^n).$$

More generally, let $K \in \mathbb{N} \cup \{\infty\}$ be minimal such that $\dim(\text{span}\{\chi_\omega(A)e_j\}_{j=1}^K) = M$. Recall that we defined

$$\Lambda_\omega = \{e_j : \chi_\omega(A)e_j \neq 0, j \leq K\}, \quad \Lambda_\Psi = \{e_j : \chi_\omega(A)e_j = 0, j \leq K\}$$
$$\text{and } \tilde{\Lambda}_\omega = \{e_j \in \Lambda_\omega : \chi_\omega(A)e_j \in \text{span}\{\chi_\omega(A)e_i\}_{i=1}^{j-1}\}.$$

Recall also from the proof of Theorem 9.2.7 that $\{\hat{e}_j\}_{j=1}^M = \Lambda_\omega \setminus \tilde{\Lambda}_\omega$. If $\{\chi_\omega(A)e_j\}_{j=1}^M$ are linearly independent then $\tilde{\Lambda}_\omega = \emptyset$, and therefore $\{\hat{e}_j\}_{j=1}^M = \{e_j\}_{j=1}^M$, which yields that the projection $\widehat{P}_M$ in (9.2.9) is the projection onto $\text{cl}\left(\text{span}\{e_j\}_{j=1}^M\right)$. □

Theorems 9.2.9 and 9.2.7 also give us convergence to the eigenvectors. With the use of (possibly countably many) shifts and rotations, the above theorem allows us to find all eigenvalues, their multiplicities and eigenspaces outside the convex hull of the essential spectrum, i.e. outside the essential numerical range.

Example 9.2.11. It is possible in the case of infinite M that the $\hat{q}_j$ do not form an orthonormal basis of $\text{ran}\chi_\omega(A)$ and we can even lose part of ω in the convergence of $\widehat{P}_M Q_n^* A Q_n \widehat{P}_M$ to a diagonal operator. This is to be contrasted to the finite-dimensional case. For example, suppose that with respect to an initial orthonormal basis $\{v_j\}_{j\in\mathbb{N}}$, A is given by the diagonal matrix $\text{diag}(1/2, 1, 1, ...)$. Now define $f_j = v_1 + (1/j)v_{j+1}$ and apply Gram–Schmidt to the sequence $\{f_j\}_{j\in\mathbb{N}}$ to generate orthonormal vectors $\{e_j\}_{j\in\mathbb{N}}$. It is easy to see that any v_j can be approximated to arbitrary accuracy using finite linear combinations of e_j and hence $\{e_j\}_{j\in\mathbb{N}}$ is an orthonormal basis of our Hilbert space. We also have that the $\chi_1(A)(f_j) = (1/j)v_{j+1}$ are linearly independent and hence so are $\chi_1(A)(e_j)$. It follows that the IQR iterates converge in the strong operator topology to the identity operator. However, we could equally take $\omega = \{1, 1/2\}$ in Theorem 9.2.9. Hence we have the curious case that $\text{cl}\left(\text{span}\{\hat{q}_j\}_{j\in\mathbb{N}}\right) \subset \text{cl}\left(\text{span}\{\hat{v}_j\}_{j>1}\right)$ and we lose the eigenvalue $1/2$.

The following corollary is entirely analogous to the finite-dimensional case.

Corollary 9.2.12. *Suppose that the conditions of Theorem 9.2.9 hold with M finite. Suppose also that for $j = 1, ..., N$ the vectors $\{\chi_{\{\lambda_1,...,\lambda_j\}}(A)e_i\}_{i=1}^{\sum_{l\leq j} m_l}$ are linearly independent. In the notation of Theorem 9.2.9, let $\rho = \sup\{|z| : z \in \Psi\}$. For $j < N$ define $r_j = \max\{|\lambda_{k+1}/\lambda_k| : k \leq j\}$ and for $j = N$ define $r_N = \max\{|\lambda_{k+1}/\lambda_k|, |\lambda_N/\rho| : k \leq j\}$. We then have the following rates of convergence to the diagonal operator for $i, j \leq M$:*

1. *$|\langle Q_n^* A Q_n e_j, e_i\rangle| = \mathcal{O}(r_k^n)$ as $n \to \infty$ if $i > j$ and k is minimal such that $i \leq \sum_{l\leq k} m_l$,*
2. *$|\langle Q_n^* A Q_n e_i, e_i\rangle - \lambda_k| = \mathcal{O}(r_k^n)$ as $n \to \infty$ if k is minimal such that $i \leq \sum_{l\leq k} m_l$.*

Proof. The result follows from Theorem 9.2.9 applied successively to $\omega_1, \omega_2, ..., \omega_N$ where $\omega_j = \{\lambda_k : k \leq j\}$. In general, analogous results follow from Theorem 9.2.9 when M is infinite and with other linear independence conditions on $\chi_{\omega'}(A)e_i$ with $\omega' \subset \omega$ but the statements become less succinct. □

In the finite-dimensional case and the case of distinct eigenvalues of the same magnitude, the QR algorithm applied to a normal matrix will 'converge' to a block diagonal matrix (without necessarily converging in each block). This can be extended to infinite dimensions by inductively using the following theorem which also extends to non-normal operators.

Theorem 9.2.13 (Block convergence theorem in infinite dimensions)**.** *Let $A \in \mathcal{B}(l^2(\mathbb{N}))$ be an invertible operator (not necessarily normal) and suppose that there exists an orthogonal projection P of rank M*

(possibly infinite) such that both the ranges of P and of $I-P$ are invariant under A. Suppose also that there exists $\alpha > \beta > 0$ such that

- $\|Ax\| \geq \alpha\|x\| \quad \forall x \in \mathrm{ran}(P)$,
- $\|Ax\| \leq \beta\|x\| \quad \forall x \in \mathrm{ran}(I-P)$.

Let $\{Q_n\}_{n\in\mathbb{N}}$ and $\{R_n\}_{n\in\mathbb{N}}$ be Q- and R-sequences of A with respect to $\{e_i\}$. Then there exists a subset $\{\hat{e}_j\}_{j=1}^M \subset \{e_i\}_{i\in\mathbb{N}}$ such that

(i) For any finite $\mu \leq M$ we have $\delta(\mathrm{span}\{Q_n\hat{e}_j\}_{j=1}^{\mu}, \mathrm{ran}(P)) = \mathcal{O}(\beta^n/\alpha^n)$ as $n\to\infty$. If M is finite this implies full convergence $\hat{\delta}(\mathrm{span}\{Q_n\hat{e}_j\}_{j=1}^{M}, \mathrm{ran}(P)) = \mathcal{O}(\beta^n/\alpha^n)$ as $n\to\infty$.

(ii) Every subsequence of $\{Q_n^*AQ_n\}_{n\in\mathbb{N}}$ has a convergent subsequence $\{Q_{n_k}^*AQ_{n_k}\}_{k\in\mathbb{N}}$ such that

$$Q_{n_k}^*AQ_{n_k} \xrightarrow{WOT} \sum_{j=1}^{M} \xi_j \otimes \hat{e}_j \bigoplus \sum_{i\in\Theta} \zeta_i \otimes e_i,$$

as $k\to\infty$, where

$$\Theta = \{j : e_j \notin \{\hat{e}_l\}_{l=1}^M\}, \quad \xi_j \in \mathrm{cl}\left(\mathrm{span}\{\hat{e}_l\}_{l=1}^M\right), \quad \zeta_i \in \mathrm{cl}\left(\mathrm{span}\{e_l\}_{l\in\Theta}\right).$$

If $\{Pe_l\}_{l=1}^M$ are linearly independent then we can take $\hat{e}_j = e_j$. Furthermore, if A has only finitely many non-zero entries in each column then we can replace WOT convergence by SOT convergence.

Remark 9.2.14. *Theorem 9.2.13 essentially says that the IQR algorithm can compute the invariant subspace $\mathrm{ran}(P)$ of such an operator if there is enough separation between A restricted to $\mathrm{ran}(P)$ and $\mathrm{ran}(I-P)$. In other words, provided the existence of a dominant invariant subspace.*

Proof of Theorem 9.2.13: The main ideas of the proof of Theorem 9.2.13 have already been presented so we sketch the proof. We first define the vectors $\{\hat{e}_j\}_{j=1}^M$ in a similar way to Definition 9.2.5 inductively by $\hat{e}_j = e_{p_j}$ where

$$p_j = \min\{i : Pe_i \notin \mathrm{span}\{P\hat{e}_k\}_{k=1}^{j-1}\}.$$

Let $r = \beta/\alpha < 1$. We will prove inductively that

(a) $\hat{\delta}(\mathrm{span}\{Q_n\hat{e}_j\}_{j=1}^{\mu}, \mathrm{span}\{PQ_n\hat{e}_j\}_{j=1}^{\mu}) \leq C_1(\mu)r^n$ for any finite $\mu \leq M$,

(b) $\|PQ_ne_j\| \leq C_2(j)r^n$ for any $j \in \Theta$,

for some constants $C_1(\mu)$ and $C_2(j)$. Suppose that this has been done. Part (i) of Theorem 9.2.13 now follows since $\mathrm{span}\{PQ_n\hat{e}_j\}_{j=1}^{\mu} \subset \mathrm{ran}(P)$. We then argue as in the proof of Theorem 9.2.9 to gain

$$Q_{n_k}^*AQ_{n_k} \xrightarrow{\mathrm{WOT}} W, \qquad k\to\infty.$$

Then by studying the inner products $\langle AQ_{n_k}e_j, Q_{n_k}e_i\rangle$ using the invariance of $\mathrm{ran}(P)$, $\mathrm{ran}(I-P)$ under A and from (b), part (ii) of Theorem 9.2.13 easily follows (note that (a) implies that $\|(I-P)Q_n\hat{e}_j\| \leq C_1(j)r^n$). The final part of the theorem then follows from the same arguments in the proof of Theorem 9.2.9. Hence we only need to prove (a) and (b).

We first claim that

$$\delta(\mathrm{span}\{PA^n\hat{e}_j\}_{j=1}^{\mu}, \mathrm{span}\{A^n\hat{e}_j\}_{j=1}^{\mu}) \leq C_3(\mu)r^n. \tag{9.2.14}$$

P commutes with A which is invertible and hence both of these spaces have dimension μ by the construction of the $\hat{e}_j$. It follows that (9.2.14) implies

$$\hat{\delta}(\text{span}\{PA^n\hat{e}_j\}_{j=1}^{\mu}, \text{span}\{A^n\hat{e}_j\}_{j=1}^{\mu}) \leq \mu^{\frac{1}{2}} C_3(\mu) r^n = C_4(\mu) r^n. \tag{9.2.15}$$

To show (9.2.14), let $x_1^n, ..., x_\mu^n$ be an orthonormal basis for $\text{span}\{PA^n\hat{e}_j\}_{j=1}^{\mu}$ and let $\xi = \sum_{j=1}^{\mu} \alpha_j x_j^n$ have norm at most 1. Now, we may choose coefficients $\beta_{j,n}$ such that $A^n \sum_{j=1}^{\mu} \beta_{j,n} x_j^n = \xi$ since $A|_{\text{ran}(P)}$ is invertible when viewed as an operator acting on $\text{ran}(P)$. By the assumptions on A we must have

$$\left(\sum_{j=1}^{m} |\beta_{j,n}|^2 \right)^{1/2} \leq \frac{1}{\alpha^n}.$$

We may change basis from $\{\hat{e}_j\}_{j=1}^{\mu}$ to $\{\tilde{e}_j\}_{j=1}^{\mu}$ such that $P\tilde{e}_j = x_j^n$. Form the vector

$$\eta_n = A^n \left(\sum_{j=1}^{\mu} \beta_{j,n} \tilde{e}_j \right) \in \text{span}\{A^n\hat{e}_j\}_{j=1}^{\mu}.$$

Then clearly by Hölder's inequality

$$\|\xi - \eta_n\| \leq \frac{\left(\sum_{j=1}^{\mu} \|A^n(I-P)\tilde{e}_j\|^2 \right)^{1/2}}{\alpha^n} \leq C_3(\mu) \frac{\beta^n}{\alpha^n},$$

proving (9.2.14) and hence (9.2.15).

Note that the proof of Lemma 9.2.6 carries over (replacing the projection $\chi_\omega(A)$ by P) to prove that

$$\text{span}\{PQ_n e_j\}_{j=1}^{m} = \text{span}\{PQ_n\hat{e}_j\}_{j=1}^{s(m)} \tag{9.2.16}$$

where $s(m)$ is maximal with $\{\hat{e}_j\}_{j=1}^{s(m)} \subset \{e_j\}_{j=1}^{m}$. It follows that

$$\begin{aligned} \delta(\text{span}\{A^n\hat{e}_j\}_{j=1}^{\mu}, \text{span}\{PQ_n\hat{e}_j\}_{j=1}^{\mu}) &= \delta(\text{span}\{A^n\hat{e}_j\}_{j=1}^{\mu}, \text{span}\{PQ_n e_j\}_{j=1}^{p_\mu}) \\ &= \delta(\text{span}\{A^n\hat{e}_j\}_{j=1}^{\mu}, \text{span}\{PA^n e_j\}_{j=1}^{p_\mu}) \\ &\leq \delta(\text{span}\{A^n\hat{e}_j\}_{j=1}^{\mu}, \text{span}\{PA^n\hat{e}_j\}_{j=1}^{\mu}) \leq C_4(\mu) r^n, \end{aligned}$$

where we have used (9.1.7) to reach the second line and $\text{span}\{PA^n\hat{e}_j\}_{j=1}^{\mu} \subset \text{span}\{PA^n e_j\}_{j=1}^{p_\mu}$ to reach the third line. Again, both spaces have dimension μ so we have

$$\begin{aligned} \delta(\text{span}\{PQ_n\hat{e}_j\}_{j=1}^{\mu}, \text{span}\{Q_n e_j\}_{j=1}^{p_\mu}) &= \delta(\text{span}\{PQ_n\hat{e}_j\}_{j=1}^{\mu}, \text{span}\{A^n e_j\}_{j=1}^{p_\mu}) \\ &\leq \delta(\text{span}\{PQ_n\hat{e}_j\}_{j=1}^{\mu}, \text{span}\{A^n\hat{e}_j\}_{j=1}^{\mu}) \leq C_5(\mu) r^n. \end{aligned} \tag{9.2.17}$$

With these arguments out of the way (these are the analogue of Proposition 9.2.4) we can now form our inductive argument, similar to the proof of Theorem 9.2.7. Suppose first that (a) holds for μ (allowing $\mu = 0$ for the initial step) and let $j \in \Theta$ have $j < p_{\mu+1}$ (where $p_{\mu+1} = \infty$ if $\mu = M$). From (a) for μ and (9.2.16) we have that

$$PQ_n e_j = v_n + \sum_{i=1}^{\mu} a_{n,i} Q_n \hat{e}_i$$

for some v_n with $\|v_n\| \leq C_1(\mu) r^n$. Then we must have

$$a_{n,i} + \langle v_n, Q_n\hat{e}_i \rangle = \langle PQ_n e_j, Q_n\hat{e}_i \rangle = \langle Q_n e_j, PQ_n\hat{e}_i \rangle.$$

Using (a) again, along with the fact that $Q_n e_j$ is orthogonal to $\{Q_n \hat{e}_i\}_{i=1}^{\mu}$, we must have $|a_{n,i}| \leq 2C_1(\mu)r^n$. It follows that we can take $C_2(j) = (2\sqrt{\mu}+1)C_1(\mu)$ for $j \in [p_\mu + 1, ..., p_{\mu+1})$ in (b). Now we use (9.2.17). Let $\xi \in \operatorname{span}\{PQ_n\hat{e}_j\}_{j=1}^{\mu+1}$ have unit norm and assume that $p_{\mu+1} < \infty$ (or else there is nothing to prove since then $\mu = M$). Then there exists $b_{n,j}$ and w_n such that

$$\xi = \sum_{j=1}^{p_{\mu+1}} b_{n,j} Q_n e_j + w_n$$

and $\|w_n\| \leq C_5(\mu+1)r^n$. Now let $j \in \Theta$ with $j < p_{\mu+1}$ then we must have

$$\langle \xi, PQ_n e_j \rangle = \langle \xi, Q_n e_j \rangle = b_{n,j} + \langle w_n, Q_n e_j \rangle.$$

We have proven (b) for such j and hence we have $|b_{n,j}| \leq \big(C_2(j) + C_5(\mu+1)\big)r^n$. It follows that we can take

$$C_1(\mu+1) = \mu^{\frac{1}{2}} \left[C_5(\mu+1) + \left\{ \sum_{j=1, j\in\Theta}^{p_{\mu+1}} [C_2(j) + C_5(\mu+1)]^2 \right\}^{\frac{1}{2}} \right],$$

where the square root factor appears since the relevant spaces are μ-dimensional. This completes the inductive step (the initial step is identical) and hence the proof of the theorem. □

Theorem 9.2.13 can be made sharper (under a slightly stricter assumption on the linear independence of $\{e_j\}_{j=1}^{M}$) with the following theorem which includes the case of $\operatorname{ran}(I-P)$ not being $A-$invariant.

Theorem 9.2.15 (Convergence to invariant subspace in infinite dimensions)**.** *Let $A \in \mathcal{B}(l^2(\mathbb{N}))$ be an invertible operator (not necessarily normal) and suppose that there exists an orthogonal projection P of finite rank M such that the range of P is invariant under A. Suppose also that there exists $\alpha > \beta > 0$ such that*

- $\|Ax\| \geq \alpha\|x\| \quad \forall x \in \operatorname{ran}(P)$,
- $\|(I-P)A(I-P)\| \leq \beta$.

Under these conditions, there exists a canonical M-dimensional A^*-invariant subspace S and we let $\tilde{P}$ denote the orthogonal projection onto S (in the special case that $\operatorname{ran}(I-P)$ is also A-invariant such as in Theorems 9.2.9 and 9.2.13, then $S = \operatorname{ran}(P)$). Suppose also that $\{\tilde{P}e_j\}_{j=1}^{M}$ are linearly independent. Let $\{Q_n\}_{n\in\mathbb{N}}$ and $\{R_n\}_{n\in\mathbb{N}}$ be Q- and R-sequences of A with respect to $\{e_i\}$. Then

(i) The subspace angle $\phi(\operatorname{span}\{e_j\}_{j=1}^{M}, S) < \pi/2$ and we have

$$\hat{\delta}(\operatorname{span}\{Q_n e_j\}_{j=1}^{M}, \operatorname{ran}(P)) \leq \frac{\sin\left(\phi(\operatorname{span}\{e_j\}_{j=1}^{M}, \operatorname{ran}(P))\right)}{\cos\left(\phi(\operatorname{span}\{e_j\}_{j=1}^{M}, S)\right)} \frac{\beta^n}{\alpha^n} \left(1 + \frac{\|PA(I-P)\|}{\alpha - \beta}\right), \tag{9.2.18}$$

(ii) Every subsequence of $\{Q_n^* A Q_n\}_{n\in\mathbb{N}}$ has a convergent subsequence $\{Q_{n_k}^* A Q_{n_k}\}_{k\in\mathbb{N}}$ such that

$$Q_{n_k}^* A Q_{n_k} \xrightarrow{WOT} \sum_{j=1}^{M} \xi_j \otimes e_j \bigoplus \sum_{i=M+1}^{\infty} \zeta_i \otimes e_i,$$

as $k \to \infty$, where

$$\xi_j \in \operatorname{cl}\left(\operatorname{span}\{e_l\}_{l=1}^{M}\right), \quad \zeta_i \in \mathcal{H}.$$

Furthermore, if A has only finitely many non-zero entries in each column then we can replace WOT convergence by SOT convergence.

Remark 9.2.16. *Theorem 9.2.15 says that the IQR algorithm can be used to approximate dominant invariant subspaces. In particular, we shall use the bound (9.2.18) to build a Δ_1 algorithm in §9.4. Note in the normal case that Theorem 9.2.9 is more precise, both in giving convergence of individual vectors to eigenvectors and in the less restrictive assumptions on spanning sets and M. In the normal case (and that of Theorem 9.2.13), we also have that the limit operator has a block diagonal form.*

9.2.3 Proof of Theorem 9.2.15

In this section, we will prove Theorem 9.2.15. The proof technique is different from those used above, and hence we have given it a separate section. Throughout, we will denote the ratio β/α by r. Note that since M is finite, the bound α implies that $A|_{\mathrm{ran}(P)} : \mathrm{ran}(P) \to \mathrm{ran}(P)$ is invertible with $\|A|_{\mathrm{ran}(P)}^{-1}\| \leq 1/\alpha$. First, let Q denote a unitary change of basis matrix from $\{e_j\}$ to $\{\tilde{e}_j\}$ where $\{\tilde{e}_j\}_{j=1}^M$ is a basis for $\mathrm{ran}(P)$. Then as matrices with respect to the original basis we can write

$$Q = [P_1, P_2], \quad Q^*AQ = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix},$$

where $A_{11} \in \mathbb{C}^{M\times M}$ and A_{12} has M rows. Our assumptions imply that $\|A_{11}^{-1}\| \leq 1/\alpha$ and $\|A_{22}\| \leq \beta$. The next lemma shows that we can change the basis further to eliminate the sub-block A_{12}. This is needed to apply a power iteration type argument.

Lemma 9.2.17. *Define the linear function $F : \mathcal{B}(l^2(\mathbb{N}), \mathbb{C}^M) \to \mathcal{B}(l^2(\mathbb{N}), \mathbb{C}^M)$ by*

$$F(T) = A_{11}^{-1} T A_{22},$$

where we identify elements of $\mathcal{B}(l^2(\mathbb{N}), \mathbb{C}^M)$ as matrices. Then we can define $T \in \mathcal{B}(l^2(\mathbb{N}), \mathbb{C}^M)$ by $T - F(T) = -A_{11}^{-1}A_{12}$. Furthermore, if we define

$$B(T) = \begin{pmatrix} I & T \\ 0 & I \end{pmatrix},$$

then $B(T)$ has inverse $B(-T)$ and

$$B(-T)\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} B(T) = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}. \tag{9.2.19}$$

Proof. Our assumptions on A ensure that F is a contraction with $\|F\| \leq r < 1$. Hence we can define T via the series

$$T = \sum_{k=0}^{\infty} F^k(-A_{11}^{-1}A_{12}).$$

It is then straightforward to check $T - F(T) = -A_{11}^{-1}A_{12}$, $B(T)B(-T) = B(-T)B(T) = I$ and the identity (9.2.19). □

Let

$$Y = Q\begin{pmatrix} I & 0 \\ -T^* & I \end{pmatrix}$$

then we have the matrix identity

$$Y^{-1}A^*Y = \begin{pmatrix} A_{11}^* & 0 \\ 0 & A_{22}^* \end{pmatrix}.$$

The canonical A^*-invariant subspace alluded to in Theorem 9.2.15 is then simply $S = \operatorname{span}\{Ye_j\}_{j=1}^M$. The space is canonical since it is easily seen that it is unchanged if we use a different basis for $\operatorname{ran}(P_1)$ and $\operatorname{ran}(P_2)$ in the definition of Q.

Now let $P_0 = \begin{pmatrix} e_1 & e_2 & \dots & e_M \end{pmatrix} \in \mathcal{B}(\mathbb{C}^M, l^2(\mathbb{N}))$ denote the matrix whose columns are the first M basis elements $\{e_j\}_{j=1}^M$. Since the $\{R_i\}$ are upper triangular, it is easy to see that

$$A^n P_0 = Q_n R_n P_0 = Q_n P_0 P_0^* R_n P_0.$$

We will denote the (invertible) matrix $P_0^* R_n P_0 \in \mathbb{C}^{M\times M}$ by Z_n. Now define

$$V_n^1 = P_1^* Q_n P_0 \in \mathcal{B}(\mathbb{C}^M), \quad V_n^2 = P_2^* Q_n P_0 \in \mathcal{B}(\mathbb{C}^M, l^2(\mathbb{N})),$$

then we have the relation

$$\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}^n \begin{pmatrix} V_0^1 \\ V_0^2 \end{pmatrix} = \begin{pmatrix} V_n^1 \\ V_n^2 \end{pmatrix} Z_n.$$

But by Lemma 9.2.17 we have

$$\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}^n = B(T) \begin{pmatrix} A_{11}^n & 0 \\ 0 & A_{22}^n \end{pmatrix} B(-T).$$

Unwinding the definitions, this implies the matrix identities

$$A_{11}^n (V_0^1 - TV_0^2) = (V_n^1 - TV_n^2) Z_n, \tag{9.2.20}$$

$$A_{22}^n V_0^2 = V_n^2 Z_n. \tag{9.2.21}$$

Lemma 9.2.18. *The following identity holds*

$$\hat{\delta}(\operatorname{span}\{Q_n e_j\}_{j=1}^M, \operatorname{ran}(P)) = \|V_n^2\|. \tag{9.2.22}$$

Proof. Note that $\operatorname{span}\{Q_n e_j\}_{j=1}^M = \operatorname{ran}(Q_n P_0)$ and $\operatorname{ran}(P) = \operatorname{ran}(P_1)$. Since $P_1 P_1^*$ and $Q_n P_0 P_0^* Q_n^*$ are orthogonal projections, it follows that

$$\begin{aligned} \hat{\delta}(\operatorname{span}\{Q_n e_j\}_{j=1}^M, \operatorname{ran}(P)) &= \|Q_n P_0 P_0^* Q_n^* - P_1 P_1^*\| \\ &= \|Q_n^*(Q_n P_0 P_0^* Q_n^* - P_1 P_1^*)Q\| \\ &= \left\| \begin{pmatrix} 0 & P_0^* Q_n^* P_2 \\ -(I-P_0)^* Q_n^* P_1 & 0 \end{pmatrix} \right\|. \end{aligned}$$

But we have that $\|P_0^* Q_n^* P_2\| = \|V_n^2\|$ and hence we are done if we can show $\|P_0^* Q_n^* P_2\| = \|(I - P_0)^* Q_n^* P_1\|$. Consider the unitary matrix

$$U := Q_n^* Q = \begin{pmatrix} P_0^* Q_n^* P_1 & P_0^* Q_n^* P_2 \\ (I-P_0)^* Q_n^* P_1 & (I-P_0)^* Q_n^* P_2 \end{pmatrix} = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}.$$

Now let $x \in \mathbb{C}^M$ be of unit norm, then $\|U_{11}x\|^2 + \|U_{21}x\|^2 = 1$. It follows that $\|U_{21}\|^2 = 1 - \sigma_1(U_{11})^2$, where σ_1 denotes the smallest singular value. Applying the same argument to U^* we see that $\|U_{12}\|^2 = 1 - \sigma_1(U_{11})^2 = \|U_{21}\|^2$, completing the proof. □

Lemma 9.2.19. *The matrix $(V_0^1 - TV_0^2)$ is invertible with*

$$\|(V_0^1 - TV_0^2)^{-1}\| \leq \frac{1}{\cos\left(\phi(\text{span}\{e_j\}_{j=1}^M, S)\right)}. \tag{9.2.23}$$

Proof. First note that since $\{\tilde{P}e_j\}_{j=1}^M$ are linearly independent, we must have $\phi(\text{span}\{e_j\}_{j=1}^M, S) < \pi/2$ and hence the bound in (9.2.23) is finite. Let $W = (P_1 - P_2T^*)(I + TT^*)^{-1/2} \in \mathcal{B}(\mathbb{C}^M, l^2(\mathbb{N}))$. By considering $W^*W = I \in \mathbb{C}^{M\times M}$, we see that the columns of W are orthonormal. In fact, expanding Y we have

$$Y = [P_1 - P_2T^* \quad P_2]$$

and hence the columns of W are a basis for the subspace S. Arguing as in the proof of Lemma 9.2.18, we have

$$\hat{\delta}(\text{span}\{e_j\}_{j=1}^M, S) = \sqrt{1 - \sigma_1(W^*P_0)^2} < 1.$$

This implies that W^*P_0 is invertible with

$$\sigma_1(W^*P_0) = \cos\left(\phi(\text{span}\{e_j\}_{j=1}^M, S)\right) > 0.$$

We also have the identity

$$V_0^1 - TV_0^2 = (I + TT^*)^{1/2}(W^*P_0).$$

Since $(I + TT^*)^{-1/2}$ has norm at most 1, we see that $(V_0^1 - TV_0^2)$ is invertible and (9.2.23) holds. □

Proof of Theorem 9.2.15: Using Lemma 9.2.19 and the matrix identities (9.2.20) and (9.2.21), we can write

$$V_n^2 = A_{22}^n V_0^2 (V_0^1 - TV_0^2)^{-1} A_{11}^{-n}(V_n^1 - TV_n^2).$$

Using (9.2.22) and (9.2.23), this implies

$$\begin{aligned}\hat{\delta}(\text{span}\{Q_ne_j\}_{j=1}^M, \text{ran}(P)) &\leq \frac{\|V_0^2\|\|V_n^1 - TV_n^2\|r^n}{\cos\left(\phi(\text{span}\{e_j\}_{j=1}^M, S)\right)} \\ &= \frac{\sin\left(\phi(\text{span}\{e_j\}_{j=1}^M, \text{ran}(P))\right)\|V_n^1 - TV_n^2\|r^n}{\cos\left(\phi(\text{span}\{e_j\}_{j=1}^M, S)\right)}.\end{aligned} \tag{9.2.24}$$

It is clear by summing a geometric series that

$$\|T\| \leq \frac{\|A_{12}\|}{\alpha(1-r)} = \frac{\|PA(I-P)\|}{\alpha - \beta}.$$

It follows that $\|V_n^1 - TV_n^2\| \leq 1 + \|PA(I-P)\|/(\alpha - \beta)$. Substituting this into (9.2.24) proves part (i) of the theorem.

Next we argue that if $i > M$ then $\|PQ_ne_i\| \to 0$ as $n \to \infty$. We have that

$$PQ_ne_i = \sum_{j=1}^M \alpha_{j,n}Q_ne_j + v_n$$

with $\|v_n\| \to 0$ by part (i). Note that we then have

$$\alpha_{j,n} = \langle PQ_ne_i, Q_ne_j\rangle + \epsilon_{j,n} = \langle Q_ne_i, PQ_ne_j\rangle + \epsilon_{j,n},$$

where $\{\epsilon_{j,n}\}_{n\in\mathbb{N}}$ converges to 0. But again by (i) we have that PQ_ne_j approaches $\text{span}\{Q_ne_k\}_{k=1}^M$ which is orthogonal to Q_ne_i and hence $\{\alpha_{j,n}\}_{n\in\mathbb{N}}$ converges to zero. The proof of part (ii) now follows the same argument as in the proof of part (i) of Theorem 9.2.9 and of the final part of Theorem 9.2.13. The key property being that if $j \leq M$ and $i > M$ then $\langle Q_n^*AQ_ne_j, e_i\rangle \to 0$ due to the invariance of $\text{ran}(P)$ under A. Note that it does not necessarily follow (as is easily seen by considering upper triangular A) that $\langle Q_n^*AQ_ne_i, e_j\rangle \to 0$ for such i, j. □

9.3 The IQR Algorithm can be computed

The previous section gives a theoretical justification for why the IQR algorithm may work. However, we are faced with the problem of how to compute with infinite data structures on a computer. Fortunately, there is a way to overcome such a problem. The key is to impose some structural requirements on the infinite matrix.

9.3.1 Quasi-banded subdiagonals

Definition 9.3.1. *Let A be an infinite matrix acting as a bounded operator on $l^2(\mathbb{N})$ with basis $\{e_j\}_{j\in\mathbb{N}}$. For $f : \mathbb{N} \to \mathbb{N}$ non-decreasing with $f(n) \geq n$ we say that A has quasi-banded subdiagonals with respect to f if $\langle Ae_j, e_i\rangle = 0$ when $i > f(j)$.*

This is the class of infinite matrices with a finite number of non-zero elements in each column (and not necessarily in each row), which is captured by the function f. It is for this class that the computation of the IQR is feasible on a finite machine. For this class of operators, one can actually compute (without any approximation or any extra discretisation) the matrix elements of the n-th iteration of the IQR algorithm as if it was done on an infinite computer (meaning the computation collapses to a finite one). The following result of independent interest is needed in the proof. It generalises the well-known fact in finite dimensions that the QR algorithm preserves bandwidth (see [Par98] for an excellent discussion of the tridiagonal case).

Proposition 9.3.2. *Let $A \in \mathcal{B}(l^2(\mathbb{N}))$ and let A_n be the n-th element in the IQR iteration, such that $A_n = Q_n^* \cdots Q_1^* A Q_1 \cdots Q_n$, where*

$$Q_j = \underset{l\to\infty}{\text{SOT-lim}}\ U_1^j \cdots U_l^j$$

and U_l^j is a Householder transformation. If A has quasi-banded subdiagonals with respect to f then so does A_n.

Proof. By induction, it is enough to prove the result for $n = 1$. From the construction of the Householder reflections $U_m^1 = P_{m-1} \oplus S_m$, the chosen η_m (see Theorem 9.1.2) have

$$\langle \eta_m, e_j\rangle = 0, \quad j > f(m).$$

Using the fact that f is increasing, it follows that each U_m^1 has quasi-banded subdiagonals with respect to f, as does the product $U_1^1 \cdots U_m^1$. It follows that Q_1 must have quasi-banded subdiagonals with respect to f and hence so does $A_1 = R_1 Q_1$ since R_1 is upper triangular. □

Theorem 9.3.3. *Let $A \in \mathcal{B}(l^2(\mathbb{N}))$ have quasi-banded subdiagonals with respect to f and let A_n be the n-th element in the IQR iteration, i.e. $A_n = Q_n^* \cdots Q_1^* A Q_1 \cdots Q_n$, where*

$$Q_j = \underset{l\to\infty}{\text{SOT-lim}}\ U_1^j \cdots U_l^j$$

and U_l^j is a Householder transformation (the superscript is not a power, but an index). Let P_m be the usual projection onto $\operatorname{span}\{e_j\}_{j=1}^m$ and denote the a-fold iteration of f by $\underbrace{f \circ f \circ \ldots \circ f}_{a \text{ times}} = f_a$. Then

$$\begin{aligned} P_m A_n P_m = {} & P_m U_m^n \cdots U_1^n U_{f_1(m)}^{n-1} \cdots U_1^{n-1} \cdots U_{f_{(n-2)}(m)}^2 \cdots U_1^2 U_{f_{(n-1)}(m)}^1 \cdots U_1^1 \\ & \cdot P_{f_n(m)} A P_{f_n(m)} \\ & \cdot U_1^1 \cdots U_{f_{(n-1)}(m)}^1 U_1^2 \cdots U_{f_{(n-2)}(m)}^2 \cdots U_1^{n-1} \cdots U_{f_1(m)}^{n-1} U_1^n \cdots U_m^n P_m. \end{aligned}$$

Remark 9.3.4. *What Theorem 9.3.3 says is that to compute the finite section of size m of the n-th iteration of the IQR algorithm (i.e. $P_m A_n P_m$), one only needs information from the finite section of size $f_n(m)$ (i.e. $P_{f_n(m)} A P_{f_n(m)}$) since the relevant Householder reflections can be computed from this information. In other words, the IQR algorithm can be computed. A version of this theorem for banded operators appeared first in [Han08a].*

Proof of Theorem 9.3.3: By induction, it is enough to prove that

$$P_m A_n P_m = P_m U_m^n ... U_1^n P_{f(m)} A_{n-1} P_{f(m)} U_1^n ... U_1^m P_m$$

To see why this is true, note that by the assumption that A has quasi-banded subdiagonals with respect to f, Proposition 9.3.2 shows that A_n has quasi-banded subdiagonals with respect to f for all $n \in \mathbb{N}$. Thus, it follows from the construction in the proof of Theorem 9.1.2 that each U_l^j is of the form

$$U_l^j = I_{l,j,1} \oplus \left(I_{l,j,2} - \frac{2}{\|\xi_{l,j}\|^2} \xi_{l,j} \otimes \bar{\xi}_{l,j} \right) \oplus I_{l,j,3},$$

where $I_{l,j,1}$ denotes the identity on $P_{l-1}\mathcal{H}$, $I_{l,j,2}$ denotes the identity on $\operatorname{span}\{e_k : l \leq k \leq f(l)\}$, $I_{l,j,3}$ denotes the identity on $P_{f(l)}^{\perp}\mathcal{H}$ and $\xi_{l,j} \in \operatorname{span}\{e_k : l \leq k \leq f(l)\}$. Since P_m is compact, it then follows that

$$\begin{aligned} P_m A_n P_m &= (\operatorname*{SOT-lim}_{l\to\infty} P_m U_l^n ... U_1^n) P_{f(m)} A_{n-1} P_{f(m)} (\operatorname*{SOT-lim}_{l\to\infty} U_1^n ... U_l^n P_m) \\ &= P_m U_m^n ... U_1^n P_{f(m)} A_{n-1} P_{f(m)} U_1^n ... U_1^m P_m. \end{aligned}$$

□

Remark 9.3.5. *This result allows us to implement the IQR algorithm because each U_l^j only affects finitely many columns or rows of A if multiplied either on the left or the right. In computer science, it is often referred to as 'lazy evaluation' when one computes with infinite data structures, but defers the use of the information until needed.*

The next question is: how restrictive is the assumption in Definition 9.3.1? In particular, suppose that $A \in \mathcal{B}(\mathcal{H})$ and that $\xi \in \mathcal{H}$ is a cyclic vector for A (i.e. $\operatorname{span}\{\xi, A\xi, A^2\xi, \ldots\}$ is dense in $\mathcal{H}$). Then by applying the Gram–Schmidt procedure to $\{\xi, A\xi, A^2\xi, \ldots\}$ we obtain an orthonormal basis $\{\eta_1, \eta_2, \eta_3, \ldots\}$ for $\mathcal{H}$ such that the matrix representation of A with respect to $\{\eta_1, \eta_2, \eta_3, \ldots\}$ is upper Hessenberg, and thus the matrix representation has only one subdiagonal. The question is therefore about the existence of a cyclic vector. Note that if A does not have invariant subspaces then every vector $\xi \in \mathcal{B}(\mathcal{H})$ is a cyclic vector. Now what happens if ξ is not cyclic for A? We may still form $\{\eta_1, \eta_2, \eta_3, \ldots\}$ as above, however, $\mathcal{H}_1 = \operatorname{cl}(\operatorname{span}\{\eta_1, \eta_2, \eta_3, \ldots\})$ is now an invariant subspace for A and $\mathcal{H}_1 \neq \mathcal{H}$. We may still form a matrix representation of A with respect to $\{\eta_1, \eta_2, \eta_3, \ldots\}$, but this will now be a matrix representation of $A|_{\mathcal{H}_1}$. Obviously, we can have that $\operatorname{Sp}(A|_{\mathcal{H}_1}) \subsetneq \operatorname{Sp}(A)$.

The following example shows that the class of matrices for which we can compute the IQR algorithm covers a wide number of applications. In particular, it includes all finite interaction Hamiltonians on graphs. Such operators play a prominent role in solid-state physics [Mat86, Mog91] describing propagation of waves and spin waves, as well as encompassing Jacobi operators studied in many physical models and integrable lattices [Tes00].

Example 9.3.6. Consider a connected, undirected graph G, such that each vertex degree is finite and the set of vertices $V(G)$ is countably infinite. Consider the set of all bounded operators A on $l^2(V(G)) \cong l^2(\mathbb{N})$ such that the set $S(v) := \{w \in V : \langle w, Av\rangle \neq 0\}$ is finite for any $v \in V$. Suppose our enumeration $\{e_1, e_2, ...\}$ of the vertices obeys the following. All of e_1's neighbours (including itself) are $S_1 = \{e_1, e_2, ..., e_{q_1}\}$ for some finite q_1. The set of neighbours of these vertices is $S_2 = \{e_1, ..., e_{q_2}\}$ for some finite q_2 where we continue the enumeration of S_1 and this process continues inductively enumerating S_m. If we know $S(v)$ for all $v \in V$ then we can find an $f : \mathbb{N} \to \mathbb{N}$ such that $A_{j,m} = 0$ if $|j| > f(m)$. We simply choose $f(n) = q_{r_n}$ where r_n is minimal such that $\cup_{j \leq n} S(e_j) \subset S_{r_n}$. These types of examples were met in §3.1.1 and §3.5.2 of Chapter 3.

9.3.2 Invertible operators

More generally, given an invertible operator A with information on how its columns decay at infinity, we can compute finite sections of the IQR iterates with error control. For computing spectral properties, we can assume, by shifting $A \to A + \lambda I$ then translating by $-\lambda$ back, that the operator we are interested in is invertible. Hence, the invertibility criterion is not that restrictive. Throughout, we will use the following lemma, which says that for invertible operators, the QR decomposition is essentially unique.

Lemma 9.3.7. *Let A be an invertible operator (viewed as a matrix acting on $l^2(\mathbb{N})$), then there exists a unique decomposition $A = QR$ with Q unitary and R invertible, upper triangular such that $R_{ii} \in \mathbb{R}_{>0}$. Furthermore, any other QR decomposition $A = Q'R'$ has a diagonal matrix $D = \mathrm{diag}(t_1, t_2, ...)$ such that $|t_i| = 1$ and $Q = Q'D$. In other words, the QR decomposition is unique up to phase choices.*

Proof. Consider the QR decomposition already discussed in this chapter, $A = Q''R''$. A is invertible, and hence Q'' is a surjective isometry so is unitary. Hence $R'' = Q''^*A$ is invertible. Being upper triangular, it follows that $R''_{ii} \neq 0$ for all i. Choose $t_i \in \mathbb{T}$ such that $t_i R''_{ii} \in \mathbb{R}_{>0}$ and set $D = \mathrm{diag}(t_1, t_2, ...)$. Letting $Q = Q''D^*$ and $R = DR''$ we clearly have the decomposition as claimed.

Now suppose that $A = Q'R'$ then we can write $Q = Q'R'R^{-1}$. It follows that $R'R^{-1}$ is a unitary upper triangular matrix and hence must be of the form $D = \mathrm{diag}(t_1, t_2, ...)$ with $|t_i| = 1$. □

Another way to see this result is to note that the columns of Q are obtained by applying the Gram–Schmidt procedure to the columns of A. The restriction that $R_{ii} \in \mathbb{R}_{>0}$ can also be incorporated into Theorem 9.3.3. Theorem 9.3.3 (in this subcase of invertibility) is then a consequence of the relations (9.1.6) and the fact that if A has quasi-banded subdiagonals with respect to f then

$$P_m A^n P_m = P_m (P_{f_n(m)} A P_{f_n(m)})^n P_m.$$

Assume that given $A \in \mathcal{B}(l^2(\mathbb{N}))$ invertible (not necessarily with quasi-banded subdiagonals), we can evaluate an increasing family of increasing functions $g^j : \mathbb{N} \to \mathbb{N}$ such that defining the matrix $A_{(j)}$ with columns $\{P_{g^j(n)} A e_n\}$ we have that $A_{(j)}$ is invertible and

$$\left\| (P_{g^j(n)} - I) A e_n \right\| \leq \frac{1}{j}. \tag{9.3.1}$$

It is easy to see that such a sequence of functions must exist since any S with $\|S - A\| < \left\|A^{-1}\right\|^{-1}$ is invertible. Given this information, without loss of generality by increasing the g^js pointwise if necessary, applying Hölder's inequality and taking subsequences, we may assume that $\left\|A_{(j)} - A\right\| \leq 1/j$. In

other words, given a sequence of functions satisfying (9.3.1) we can evaluate a sequence of functions with this stronger condition. The following says that given such a sequence of functions, we can compute the truncations $P_m A_n P_m$ to a given precision.

Theorem 9.3.8. *Suppose $A \in \mathcal{B}(l^2(\mathbb{N}))$ is invertible and the family of functions $\{g^l\}_{l\in\mathbb{N}}$ are as above. Suppose also that we are given a bound C such that $\|A\| \leq C$. Let $\epsilon > 0$ and $m, n \in \mathbb{N}$, then we can choose j such that applying Theorem 9.3.3 (with the diagonal operators to ensure $R_{ii} > 0$) to $A_{(j)}$ using the function g^j instead of f, we have the guaranteed bound*

$$\left\|P_m A_n P_m - P_m A_{(j),n} P_m\right\| \leq \epsilon,$$

where $A_{(j),n}$ denotes the n-th IQR iterate of $A_{(j)}$.

Proof of Theorem 9.3.8: First consider the error when applying Theorem 9.3.3 to $A_{(j)}$ with g^j for any fixed j. We will show that we can compute an error bound which converges to zero as $j \to \infty$ and from this, the theorem easily follows by successively computing the bound and halting when this bound is less than ϵ.

Write the QR decompositions

$$A^n = \hat{Q}_n \hat{R}_n, \quad (A_{(j)})^n = \hat{Q}_{(j),n} \hat{R}_{(j),n}.$$

We have $\left\|A - A_{(j)}\right\| \leq 1/j$ and hence, by writing $A_{(j)} = A + (A_{(j)} - A)$, that

$$\left\|A^n - (A_{(j)})^n\right\| \leq \sum_{k=1}^{n} \binom{n}{k} \frac{1}{j^k} C^{n-k} \leq \frac{(C+1)^n}{j} = \frac{\tilde{C}}{j},$$

where $\tilde{C} = (C+1)^n$. The columns of $\hat{Q}_n$ and $\hat{Q}_{(j),n}$ are simply the columns of the matrices A^n and $(A_{(j)})^n$ after the application of Gram–Schmidt. Let the first m columns of A^n and $(A_{(j)})^n$ be denoted by $\{a_k\}_{k=1}^m$ and $\{\tilde{a}_k^j\}_{k=1}^m$ respectively and let $\{q_k\}_{k=1}^m$ and $\{\tilde{q}_k^j\}_{k=1}^m$ be the vectors obtained after applying Gram–Schmidt to these sequences of vectors. We then have

$$\begin{aligned} \|q_1 - \tilde{q}_1^j\| &= \left\| \frac{a_1}{\|a_1\|} - \frac{\tilde{a}_1^j}{\|\tilde{a}_1^j\|} \right\| \\ &= \left\| \frac{a_1(\|\tilde{a}_1^j\| - \|a_1\|)}{\|a_1\|\|\tilde{a}_1^j\|} - \frac{(\tilde{a}_1^j - a_1)\|a_1\|}{\|a_1\|\|\tilde{a}_1^j\|} \right\| \leq \frac{2\|a_1 - \tilde{a}_1^j\|}{\|\tilde{a}_1^j\|} \leq \frac{2\tilde{C}}{j\|\tilde{a}_1^j\|}. \end{aligned} \tag{9.3.2}$$

For a vector v of unit norm, let $P_{\perp v}$ denote the orthogonal projection onto the space of vectors perpendicular to v. Note that for two such vectors v, w, we have $\|P_{\perp v} - P_{\perp w}\| \leq \|v - w\|$. Let

$$v_k = P_{\perp q_{k-1}} \cdots P_{\perp q_1} a_k, \quad \tilde{v}_k^j = P_{\perp \tilde{q}_{k-1}^j} \cdots P_{\perp \tilde{q}_1^j} \tilde{a}_k^j, \tag{9.3.3}$$

then q_k are just the normalised version of v_k and likewise $\tilde{q}_k^j$ are just the normalised version of $\tilde{v}_k^j$. Suppose that for $\mu < k$ we have $\|q_\mu - \tilde{q}_\mu^j\| \leq \delta$ for some $\delta > 0$. Then applying the above products of projections we have

$$\begin{aligned} \|v_k - \tilde{v}_k^j\| &\leq \|P_{\perp q_{k-1}} \cdots P_{\perp q_1}(a_k - \tilde{a}_k^j)\| + \|P_{\perp q_{k-1}} \cdots P_{\perp q_1} \tilde{a}_k^j - \tilde{v}_k^j\| \\ &\leq \|a_k - \tilde{a}_k^j\| + \|P_{\perp q_{k-1}} \cdots P_{\perp q_1} - P_{\perp \tilde{q}_{k-1}^j} \cdots P_{\perp \tilde{q}_1^j}\| \|\tilde{a}_k^j\| \\ &\leq \|a_k - \tilde{a}_k^j\| + (k-1)\delta \|\tilde{a}_k^j\|. \end{aligned}$$

In the last line we have used the fact that if the operators $\{A_l\}_{l=1}^m$ and $\{B_l\}_{l=1}^m$ have norm bounded by 1, then

$$\left\| \prod_{l=1}^{m} A_l - \prod_{l=1}^{m} B_l \right\| \leq \sum_{l=1}^{m} \|A_l - B_l\|.$$

Applying the same argument as in the inequalities (9.3.2) we see that

$$\|q_k - \tilde{q}_k^j\| \leq \frac{2(\|a_k - \tilde{a}_k^j\| + (k-1)\delta\|\tilde{a}_k^j\|)}{\|\tilde{v}_k^j\|} \leq \frac{2(\tilde{C}/j + 2(k-1)\delta\tilde{C})}{\|\tilde{v}_k^j\|}, \tag{9.3.4}$$

since $\|\tilde{a}_k^j\| \leq C + \tilde{C}/j \leq 2\tilde{C}$. Now note that we can compute the $\|\tilde{v}_k^j\|$ from the proof of Theorem 9.3.3. Set $\delta_1(j) = \frac{2\tilde{C}}{j\|\tilde{a}_1^j\|}$ and for $1 < k \leq m$ define iteratively

$$\delta_k(j) = \max\left\{ \delta_{k-1}(j), \frac{2(\tilde{C}/j + 2(k-1)\delta_{k-1}(j)\tilde{C})}{\|\tilde{v}_k^j\|} \right\}.$$

We must have $\|q_k - \tilde{q}_k^j\| \leq \delta_m(j)$ for $1 \leq k \leq m$ where we have now shown the j dependence as an argument.

It follows that $\|(\hat{Q}_n - \hat{Q}_{(j),n})P_m\| \leq \sqrt{m}\delta_m(j)$ and hence that

$$\begin{aligned}
\|P_m A_n P_m - P_m A_{(j),n} P_m\| &\leq \|P_m(\hat{Q}_n - \hat{Q}_{(j),n})^* A \hat{Q}_n P_m\| + \|P_m \hat{Q}_{(j),n}^* (A\hat{Q}_n - A_{(j)}\hat{Q}_{(j),n})P_m\| \\
&\leq \sqrt{m}\delta_m(j)C + \|(A - A_{(j)})\hat{Q}_{(j),n}P_m\| + \|A(\hat{Q}_n - \hat{Q}_{(j),n})P_m\| \\
&\leq 2\sqrt{m}\delta_m(j)C + \frac{1}{j}.
\end{aligned}$$

So we need only show that $\delta_m(j) \to 0$ as $j \to \infty$. Note that as $j \to \infty$, the columns of $(A_{(j)})^n$ converge to that of A^n. It follows that $\tilde{a}_k^j$ converge to a_k and $\tilde{q}_1^j$ converges to q_1. An easy inductive argument using (9.3.3) and (9.3.4) shows that the vectors $\tilde{q}_k^j$ converge to q_k and $\|\tilde{v}_k^j\|$ are bounded below. $\delta_m(j) \to 0$ now follows. □

9.4 SCI Classification Theorems

In this section, we will apply the above results to prove three classification theorems in the SCI hierarchy.

Remark 9.4.1. *For simplicity, we will assume our algorithms can extract radicals but note that it is straightforward to adapt the algorithms to arithmetic algorithms by approximating square roots.*

First, assume that $A \in \mathcal{B}(l^2(\mathbb{N}))$ is an invertible normal operator with $\mathrm{Sp}(A) = \omega \cup \Psi$, where $\omega \cap \Psi = \emptyset$, $\omega = \{\lambda_i\}_{i=1}^N$, and the λ_i's are isolated eigenvalues with multiplicity m_i satisfying $|\lambda_1| > \ldots > |\lambda_N|$. As usual, we also assume that $\sup\{|\theta| : \theta \in \Psi\} < |\lambda_N|$ and set

$$M := m_1 + \ldots + m_N \in \mathbb{N} \cup \{\infty\}. \tag{9.4.1}$$

In this section, we will assume for simplicity that all the m_i except possibly m_N are finite. To be able to obtain the classification results we need two key assumptions.

(I) (*Column decay*): We assume a much weaker condition than bandedness of the infinite matrix. Indeed, we suppose a known decay of the elements in the columns of A that is described through a family of increasing functions $\{g^j\}_{j\in\mathbb{N}}$. In particular, $g^j : \mathbb{N} \to \mathbb{N}$ is such that defining the infinite matrix $A_{(j)}$ with columns $\{P_{g^j(n)}Ae_n\}_{n\in\mathbb{N}}$ we have that $A_{(j)}$ is invertible and

$$\left\|(P_{g^j(n)} - I)Ae_n\right\| \leq \frac{1}{j}, \qquad n \in \mathbb{N}. \tag{9.4.2}$$

(II) (*Distance to span of eigenvectors*): In order to obtain error control (Δ_1 classification), one needs to control the hidden constant in the $\mathcal{O}(r^n)$ estimate in (9.2.8). This is done as follows, where $\{Q_n\}_{n\in\mathbb{N}}$ is a Q-sequence of A with respect to $\{e_j\}_{j\in\mathbb{N}}$. Given finite $k < M+1$ with $m_1 + ... + m_{N-1} < k$, we will assume that if $l < N$ then $\{\chi_{\{\lambda_1,...,\lambda_l\}}(A)e_j\}_{j=1}^{m_1+...+m_l}$ are linearly independent. We also assume that $\{\chi_{\{\lambda_1,...,\lambda_N\}}(A)e_j\}_{j=1}^{k}$ are linearly independent. This simply ensures that the IQR converges with the expected ordering (largest eigenvalue in the first diagonal entry then in descending order). It follows from Theorems 9.2.9 and 9.2.7, that there exist eigenspaces $E_1, ..., E_N$ (with the last space depending on k and the vectors $\{e_j\}$) corresponding to the eigenvalues $\lambda_1, ..., \lambda_N$ such that

- $E_i = \ker(A - \lambda_i I)$ is the full eigenspace if $i < N$
- $\hat{\delta}\Big(\bigoplus_{i=1}^{l} E_i, \operatorname{span}\{Q_n e_j\}_{j=1}^{\min\{m_1+...+m_l,k\}}\Big) \to 0$ as $n \to \infty$ for $l = 1, ..., N$.

We then define the initial supremum subspace angle by

$$\Phi(A, \{e_j\}_{j=1}^{k}) := \sup_{l=1,...,N} \phi\left(\bigoplus_{i=1}^{l} E_i, \operatorname{span}\{e_j\}_{j=1}^{\min\{m_1+...+m_l,k\}}\right),$$

where ϕ, defined by (9.1.1), denotes the subspace angle. Our assumptions and the proofs in §9.2 show that $\Phi(A, \{e_j\}_{j=1}^{k}) < \pi/2$ and hence the key quantity $\tan\big(\Phi(A, \{e_j\}_{j=1}^{k})\big)$ is finite.

Remark 9.4.2. *The quantity* $\tan\big(\Phi(A, \{e_j\}_{j=1}^{k})\big)$ *can be viewed as a measure of how far* $\{e_j\}_{j=1}^{k}$ *is from* $\{q_j\}_{j=1}^{k}$*, the* k *eigenvectors of* A *corresponding to the first* k *eigenvalues (including multiplicity and preserving order). Hence it gives an estimate of how good the initial approximation* $\{e_j\}_{j=1}^{k}$ *to* $\{q_j\}_{j=1}^{k}$ *is. Indeed, we know from* (9.2.8) *that the convergence rate is* $\mathcal{O}(r^n)$*, and the hidden constant* C *depends exactly on this behaviour. In particular, if* $e_j = q_j$ *for* $j \leq k$ *then* $C = 0$*.*

Define also

$$r(A) = \max\{|\lambda_2/\lambda_1|, ..., |\lambda_N/\lambda_{N-1}|, \rho(A)/|\lambda_N|\}, \qquad \rho(A) = \sup\{|z| : z \in \Psi\}.$$

We can now define the class of operators $\Omega_{t,L}^{k}$ for the classification theorem.

Definition 9.4.3. *Given* $k \in \mathbb{N}$*,* $t \in (0,1)$ *and* $L > 0$*, let* $\Omega_{t,L}^{k}$ *denote the class of invertible normal operators* A *acting on* $l^2(\mathbb{N})$ *with* $\|A\| \leq L$ *such that:*

1. *There exists the decomposition* $\operatorname{Sp}(A) = \omega \cup \Psi$ *as above with* $m_1 + ... + m_{N-1} < k \leq M$*, where* M *is defined in* (9.4.1)*.*
2. *If* $m_1 + ... + m_l < k$ *then* $\{\chi_{\{\lambda_1,...,\lambda_l\}}(A)e_j\}_{j=1}^{m_1+...+m_l}$ *are linearly independent. Also, the vectors* $\{\chi_{\{\lambda_1,...,\lambda_N\}}(A)e_j\}_{j=1}^{k}$ *are linearly independent.*
3. *We have access to functions* $g^j : \mathbb{N} \to \mathbb{N}$ *with (9.4.2).*
4. *It holds that* $r(A) \leq t$ *and* $\tan\big(\Phi(A, \{e_j\}_{j=1}^{k})\big) \leq L$*.*

We can now define the computational problem that we want to classify in the SCI hierarchy. Consider for any $A \in \Omega_{t,L}^{k}$, the problem of computing the[1] k-th largest eigenvalues (including multiplicity) and the corresponding eigenspaces. In other words, we consider the set-valued mapping

$$\Xi_1^{\mathrm{QR}}(A) = \mathcal{S} \subset \mathcal{M} = \mathbb{C}^k \times \big(l^2(\mathbb{N})\big)^k$$

[1] In the case of eigenvalues of equal magnitude, we compute a suitable subset - see below.

where we define

$$\mathcal{S} := \Big\{ (\underbrace{\lambda_1, ..., \lambda_1}_{m_1 \text{ times}}, ..., \underbrace{\lambda_N, ..., \lambda_N}_{k-(m_1+...+m_{N-1}) \text{ times}}) \times (\hat{q}_1, ..., \hat{q}_k) :$$
$$\text{s.t. } \{\hat{q}_j\}_{j=m_1+...+m_l}^{m_1+...+m_l} \text{ is an orthonormal basis of } \mathrm{ran}(\chi_{\lambda_l}(A)) \text{ for } l < N$$
$$\text{and } \{\hat{q}_j\}_{j=m_1+...+m_{N-1}+1}^{k} \text{ is an orthonormal basis for a subspace of } \mathrm{ran}(\chi_{\lambda_N}(A)) \Big\}.$$

As discussed in §2.1, when we speak of convergence of $\Gamma_n(A) \in \mathcal{M}$ to $\Xi_1^{\mathrm{QR}}(A)$, we define, with a slight abuse of notation,

$$\mathrm{dist}\left(\Gamma_n(A), \Xi_1^{\mathrm{QR}}(A)\right) := \inf_{y \in \Xi_1^{\mathrm{QR}}(A)} d_{\mathcal{M}}(\Gamma_n(A), y) \to 0.$$

Having established the basic definition, we can now present the classification theorem.

Theorem 9.4.4 (Δ_1 classification for the extremal part of the spectrum)**.** *Given the above set-up we have $\{\Xi_1^{\mathrm{QR}}, \Omega_{t,L}^k\} \in \Delta_1^R$. In other words, for all $n \in \mathbb{N}$, there exists a general tower using radicals, $\Gamma_n(A)$, such that for all $A \in \Omega_{t,L}^k$,*

$$\mathrm{dist}\left(\Gamma_n(A), \Xi_1^{\mathrm{QR}}(A)\right) \leq 2^{-n}.$$

Remark 9.4.5. *Note that this means we converge to the k largest magnitude eigenvalues in order with error control, and not just arbitrary points of the spectrum.*

Proof of Theorem 9.4.4: Let $A \in \Omega_{t,L}^k$ then by the definition of $\Omega_{t,L}^k$, we may take $\hat{e}_j = e_j$ for $j = 1, ..., k$ in the arguments in §9.2.1. The first step is to bound $Z(A, \{e_j\}_{j=1}^k)$ in terms of $\Phi(A, \{e_j\}_{j=1}^k)$. Let $\{\tilde{e}_j\}_{j=1}^k$ denote the basis described in §9.2.1. In our case:

- For any $1 \leq i \leq k$, $\mathrm{span}\{\tilde{e}_j\}_{j=1}^i = \mathrm{span}\{e_j\}_{j=1}^i$.
- If $j > m_1 + ... + m_l$ then $\chi_{\lambda_l}(A)\tilde{e}_j = 0$.
- The vectors $\{\chi_{\lambda_l}(A)\tilde{e}_j\}_{j=m_1+...+m_{l-1}+1}^{\min\{m_1+...+m_l,k\}}$ are orthonormal.

Let $\delta_j = \|\tilde{e}_j\|$ then we must have that if $m_1 + ...m_{l-1} < j \leq m_1 + ... + m_l$ then

$$\begin{aligned}
\frac{\delta_j^2 - 1}{\delta_j^2} &\leq \delta \left(\mathrm{span}\{\tilde{e}_j\}, \bigoplus_{i=1}^{l} \mathrm{span}\{\chi_{\{\lambda_i\}}(A)\tilde{e}_j\}_{j=m_1+...m_{i-1}+1}^{\min\{m_1+...+m_i,k\}} \right)^2 \\
&\leq \delta \left(\mathrm{span}\{\tilde{e}_j\}_{j=1}^{\min\{m_1+...+m_l,k\}}, \bigoplus_{i=1}^{l} \mathrm{span}\{\chi_{\{\lambda_i\}}(A)\tilde{e}_j\}_{j=m_1+...m_{i-1}+1}^{\min\{m_1+...+m_i,k\}} \right)^2 \\
&= \delta \left(\mathrm{span}\{e_j\}_{j=1}^{\min\{m_1+...+m_l,k\}}, \bigoplus_{i=1}^{l} E_i \right)^2 \\
&\leq \sin^2\left(\Phi(A, \{e_j\}_{j=1}^k)\right)
\end{aligned}$$

Where the first line holds since the nearest point to $\tilde{e}_j$ in $\bigoplus_{i=1}^l \mathrm{span}\{\chi_{\{\lambda_i\}}(A)\tilde{e}_j\}_{j=m_1+...m_{i-1}+1}^{\min\{m_1+...+m_i,k\}}$ is simply $\chi_{\lambda_l}(A)\tilde{e}_j$ and the E_i are defined as above and in (9.2.1). Rearranging, this implies that

$$\delta_j^2 \leq \frac{1}{1 - \sin^2\left(\Phi(A, \{e_j\}_{j=1}^k)\right)} = \frac{1}{\cos^2\left(\Phi(A, \{e_j\}_{j=1}^k)\right)}.$$

Hence it follows that

$$Z(A, \{e_j\}_{j=1}^k) = \left(\sum_{j=1}^{k} \delta_j^2 - 1 \right)^{\frac{1}{2}} \leq \left(\sum_{j=1}^{k} \tan^2\left(\Phi(A, \{e_j\}_{j=1}^k)\right) \right)^{\frac{1}{2}} \leq \sqrt{k}L.$$

In particular, Theorem 9.2.7 and its proof now imply that

$$\hat{\delta}(\mathrm{span}\{\hat{q}_j\}, \mathrm{span}\{Q_m e_j\}) \leq B(j)\sqrt{k}Lt^m,$$

where $\{\hat{q}_j\}_{j=1}^k$ are orthonormal eigenvectors of A and Q_m is a $Q-$sequence of A. In particular, $\{B(j)\}_{j=1}^k$ can be computed in finitely many arithmetic operations from the induction proof of Theorem 9.2.7. It follows that there exists $z_{j,m} \in \mathbb{C}$ of unit modulus such that defining $\beta = \max\{B(1), ..., B(k)\}\sqrt{k}L$, we have

$$\|Q_m e_j - z_{j,m}\hat{q}_j\| \leq \beta t^m.$$

Note that we do not need to assume knowledge of N for this bound (trivially $N \leq k$). Using that Q_m is an isometry, this implies that

$$\left|\langle Q_m^* A Q_m e_j, e_j\rangle - \lambda_{a_j}\right| \leq 2\|A\|\beta t^m \leq 2L\beta t^m,$$

where $A\hat{q}_j = \lambda_{a_j}$. Note that we must have $\{\lambda_{a_j}\}_{j=m_1+...+m_{l-1}+1}^{m_1+...+m_l} = \lambda_l$ and $\{\lambda_{a_j}\}_{j=m_1+...+m_{N-1}+1}^k = \lambda_N$ by 3. in the definition of $\Omega_{t,L}^k$.

Given any $\epsilon > 0$, choose m large enough so that $2L\beta t^m \leq \epsilon$ and $\beta t^m \leq \epsilon$. The fact that $\|A\| \leq L$ and (9.4.2) hold implies that we can compute $\langle Q_m^* A Q_m e_j, e_j\rangle$ to accuracy ϵ using finitely many arithmetical and square root operations using Theorem 9.3.8. Call these approximations $\tilde{\lambda}_1, \tilde{\lambda}_2, ..., \tilde{\lambda}_k$. Furthermore, the proof of Theorem 9.3.8 also makes clear that we can compute $Q_m e_j \in l^2(\mathbb{N})$ to accuracy ϵ using finitely many arithmetical and square root operations (the approximations have finite support). Call these approximations $\tilde{q}_1, \tilde{q}_2, ..., \tilde{q}_k$. Then set

$$\Gamma^\epsilon(A) = (\tilde{\lambda}_1, \tilde{\lambda}_2, ..., \tilde{\lambda}_k) \times (\tilde{q}_1, \tilde{q}_2, ..., \tilde{q}_k).$$

The above estimates show that $\mathrm{dist}(\Gamma^\epsilon(A), \Xi_1^{\mathrm{QR}}(A)) \leq 4k\epsilon$. The proof is completed by setting $\Gamma_n(A) = \Gamma^{2^{-(n+2)}/k}(A)$. □

Next, suppose we have a continuous increasing function $g : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ diverging at ∞ such that $g(0) = 0$ and $g(x) \leq x$. Let Ω_{IQR}^g be the set of all operators A acting on $l^2(\mathbb{N})$ (i.e. we fix the representation with respect to the canonical basis) for which the IQR algorithm converges in the weak operator topology to a diagonal matrix with the same spectrum as A and such that

$$\left\|(A - zI)^{-1}\right\|^{-1} \geq g\big(\mathrm{dist}(z, \mathrm{Sp}(A))\big).$$

Note that by Theorem 9.2.9 this includes all normal compact operators, A, such that $\{z \in \mathrm{Sp}(A) : |z| = s\}$ has size at most 1 for all $s > 0$ (where we can take $g(x) = x$).[2] We will allow evaluations of g in our algorithms and also assume that we are given functions that satisfy (9.4.2) and have an upper bound for $\|A\|$. We consider computing $\Xi_2^{\mathrm{QR}}(A) = \mathrm{Sp}(A)$ in the space of compact non-empty subsets of $\mathbb{C}$ with the Hausdorff metric.

Theorem 9.4.6 (Σ_1 classification for spectrum)**.** *Given the above set-up we have* $\{\Xi_2^{\mathrm{QR}}, \Omega_{\mathrm{IQR}}^g\} \in \Sigma_1^R$. *In other words, there is a convergent sequence of general towers using radicals,* $\Gamma_n(A)$*, such that* $\Gamma_n(A) \to \Xi_2^{\mathrm{QR}}(A) = \mathrm{Sp}(A)$ *for any* $A \in \Omega_{\mathrm{IQR}}^g$ *and for all* n *we have*

$$\Gamma_n(A) \subset \mathrm{Sp}(A) + B_{2^{-n}}(0).$$

[2] A simple compactness argument shows that for any bounded operator A there is a corresponding function g that works.

Proof of Theorem 9.4.6: Let $A \in \Omega^g_{\mathrm{IQR}}$ and Q_m be a $Q-$sequence of A. Fix $n \in \mathbb{N}$. Then Theorem 9.3.8 shows that we can compute any finite number of the diagonal entries of $Q_m^* A Q_m$ to any given accuracy using finitely many arithmetical and square root operations. Similarly, the proof shows that we can compute $AQ_m e_j$ and $Q_m e_j$ to any given accuracy in $l^2(\mathbb{N})$ (the approximations have finite support). Now let $\alpha_{j,m}$ be the computed approximations of $\langle Q_m^* A Q_m e_j, e_j \rangle$ to accuracy $1/m$, then since $A \in \Omega^g_{\mathrm{IQR}}$ we have that $\lim_{m\to\infty} \alpha_{j,m} = \alpha_j \in \mathrm{Sp}(A)$. Furthermore, $\{\alpha_j : j \in \mathbb{N}\}$ is dense in $\mathrm{Sp}(A)$. We have that

$$\left\|(A - \alpha_{j,m} I)^{-1}\right\|^{-1} \leq \|AQ_m e_j - \alpha_{j,m} Q_m e_j\|$$

and hence that

$$\mathrm{dist}(\alpha_{j,m}, \mathrm{Sp}(A)) \leq g^{-1}(\|AQ_m e_j - \alpha_{j,m} Q_m e_j\|). \tag{9.4.3}$$

Given m, j, we can compute an upper bound $h_{j,m}$ for the right-hand side of (9.4.3) by approximating the norm $\|AQ_m e_j - \alpha_{j,m} Q_m e_j\|$ from above to accuracy $1/m$ and finitely many evaluations of g. Namely, let $x_{j,m}$ be the approximation of $\|AQ_m e_j - \alpha_{j,m} Q_m e_j\|$ and set

$$h_{j,m} = \frac{\min\{l \in \mathbb{N} : g(l/m) \geq x_{j,m}\}}{m}.$$

It is then clear that $\lim_{m\to\infty} h_{j,m} = 0$ and $h_{j,m} \geq g^{-1}(\|AQ_m e_j - \alpha_{j,m} Q_m e_j\|)$.

We set $\Gamma_n(A) = \{\alpha_{j,m(n,A)} : j = 1, ..., n\}$ where $m(n, A)$ is minimal such that $h_{j,m} \leq 2^{-n}$ for $j = 1, ..., n$. By (9.4.3), we must have that

$$\Gamma_n(A) \subset \mathrm{Sp}(A) + B_{2^{-n}}(0).$$

It is also clear that $\Gamma_n(A) \to \mathrm{Sp}(A)$ in the Hausdorff metric. □

The final result considers dominant invariant subspaces discussed in Theorem 9.2.15. Let $M \in \mathbb{N}$, $t \in (0, 1)$ and $L > 0$. We let $\tilde{\Omega}^M_{t,L}$ denote the class of operators such that the assumptions of Theorem 9.2.15 hold (same M) and such that:

1. $\beta/\alpha < t$,
2. $\max\left\{\|A\|, \frac{\sin\left(\phi(\mathrm{span}\{e_j\}_{j=1}^M, \mathrm{ran}(P))\right)}{\cos\left(\phi(\mathrm{span}\{e_j\}_{j=1}^M, S)\right)}\left(1 + \frac{\|PA(I-P)\|}{\alpha-\beta}\right)\right\} \leq L.$

We also assume that we are given functions that satisfy (9.4.2) and consider computing the dominant invariant subspace $\Xi_3^{\mathrm{QR}}(A) = \mathrm{ran}(P)$ in the space of M-dimensional subspaces of $l^2(\mathbb{N})$ equipped with the metric $\hat{\delta}$.

Theorem 9.4.7 (Δ_1 classification for dominant invariant subspace)**.** *Given the above set-up we have the classification $\{\Xi_3^{\mathrm{QR}}, \tilde{\Omega}^M_{t,L}\} \in \Delta_1^R$. In other words, for all $n \in \mathbb{N}$, there exists a general tower using radicals, $\Gamma_n(A)$, each an M-dimensional subspace of $l^2(\mathbb{N})$, such that for all $A \in \tilde{\Omega}^M_{t,L}$,*

$$\hat{\delta}\left(\Gamma_n(A), \Xi_3^{\mathrm{QR}}(A)\right) \leq 2^{-n}.$$

Proof of Theorem 9.4.7: Let $n \in \mathbb{N}$ and $A \in \tilde{\Omega}^M_{t,L}$. Then from Theorem 9.2.15, we can choose m large so that $t^m L < 2^{-(n+1)}$, and hence

$$\hat{\delta}(\mathrm{span}\{Q_m e_j\}_{j=1}^M, \mathrm{ran}(P)) < 2^{-(n+1)}.$$

Using Theorem 9.3.8 and its proof, given ϵ we can compute in finitely many arithmetical and square root operations, approximations $v_{m,j}(\epsilon)$ (of finite support) such that

$$\|v_{m,j}(\epsilon) - Q_m e_j\| \leq \epsilon.$$

The vectors $\{Q_m e_j\}_{j=1}^M$ are orthonormal, as are the approximations $\{v_{m,j}(\epsilon)\}_{j=1}^M$. A simple application of Hölder's inequality then yields

$$\hat{\delta}\left(\operatorname{span}\{v_{m,j}(\epsilon)\}_{j=1}^M, \operatorname{span}\{Q_m e_j\}_{j=1}^M\right) \leq \sqrt{M}\epsilon.$$

By the triangle inequality, the proof of the theorem is complete by choosing ϵ such that $\sqrt{M}\epsilon \leq 2^{-(n+1)}$ and then setting $\Gamma_n(A) = \operatorname{span}\{v_{m,j}(\epsilon)\}_{j=1}^M$. □

9.5 Numerical Examples

The aim of the is section is threefold:

1. To demonstrate the convergence and implementation results of §§9.2–9.4 on practical examples.

2. To demonstrate that, as well as the proven results, the IQR algorithm performs better than theoretically expected in many cases. In particular, we conjecture that for normal operators whose essential spectrum has exactly one extremal point, the IQR algorithm will also converge to this point. We also demonstrate cases where this seems to hold even if there are multiple extreme points of the essential spectrum and even in non-normal cases.

3. To compare the IQR algorithm to the finite section method and show that in some cases it considerably outperforms it. In general, one can view $\operatorname{Sp}(P_m Q_n^* A Q_n P_m)$ as a generalised version of the finite section method, now with two parameters (m and n) that can be varied with n controlling the number of IQR iterates that are performed in an infinite-dimensional manner. In some cases, we find this avoids spectral pollution whilst still converging to the entire spectrum.

The reader is directed to §7.1 and §7.3.2 of Chapter 7 for a thorough discussion of the finite section method and when it fails (as well as the SCI of detecting this failure). Before embarking with some numerical examples, two remarks are in order. First, extra care has been taken in the case of non-self-adjoint (NSA) operators whose finite truncations can be non-normal, and hence the computation of their spectra can be numerically unstable. Unless stated otherwise, all calculations were performed in double precision (in MATLAB) and have been checked against extended precision [Adv06] to ensure that none of the results are due to numerical artefacts. Second, when dealing with operators acting on $l^2(\mathbb{Z})$ we use $\mathbb{N}$ as an index set by listing the canonical basis as $e_0, e_1, e_{-1}, e_2, e_{-2}, ...$, allowing us to apply the IQR algorithm on $l^2(\mathbb{N})$. Of course, different indexing is possible, and, in general, this would lead to different implementations of the IQR algorithm,[3] but we stick with this ordering throughout.

[3] A discussion of this is beyond the scope of this chapter. In effect, for invertible operators, this corresponds to choosing the order of columns on which to perform a Gram–Schmidt type procedure.

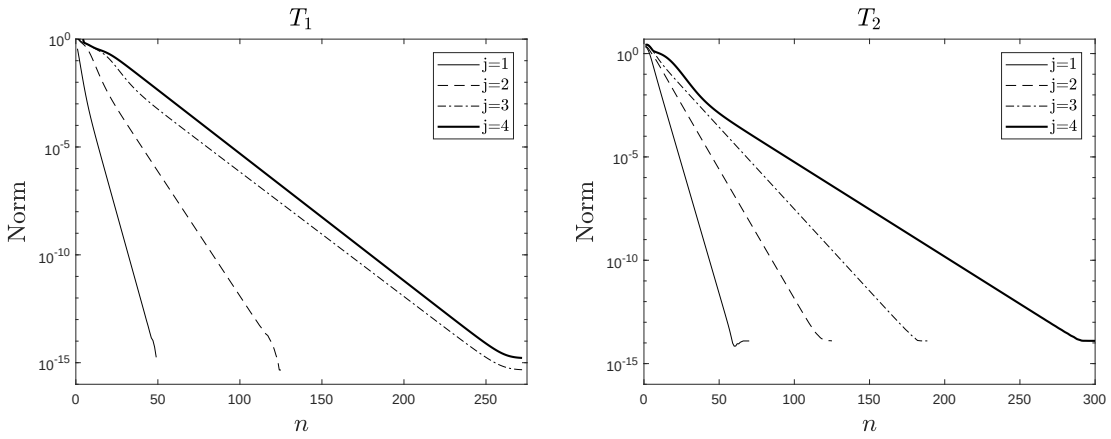


Figure 9.1: Exponential convergence to the diagonal blocks for T_1 and T_2.

9.5.1 Numerical examples I: normal operators

Example 9.5.1 (Convergence of the IQR algorithm)**.** We begin with two simple examples that demonstrate the linear (or exponential) convergence proven in Theorem 9.2.9 and Corollary 9.2.12 (and its generalisations). Consider first the one-dimensional discrete Schrödinger operator given by

$$T_1 = \begin{pmatrix} v_1 & 1 & & & \\ 1 & v_2 & 1 & & \\ & 1 & v_3 & 1 & \\ & & 1 & v_4 & \ddots \\ & & & \ddots & \ddots \end{pmatrix},$$

where $v_j = 5\sin(j)^2/\sqrt{j}$ if $j \leq 10$ and $v_j = 0$ otherwise. As a compact (in fact finite rank) perturbation of the free Laplacian, $\mathrm{Sp}(T_1)$ consists of the interval $[-2, 2]$ together with isolated eigenvalues of finite multiplicity which can be computed [WO17]. The second operator, T_2, consists of taking the operator

$$T_0 = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & \frac{3i}{2} & 0 & 0 \\ 0 & 0 & -\frac{5}{4} & 0 \\ 0 & 0 & 0 & -\frac{9i}{8} \end{pmatrix} \bigoplus U_1,$$

where U_k denotes the bilateral shift $e_j \to e_{j+k}$, writing this as an operator on $l^2(\mathbb{N})$ and then mixing the spaces via a random unitary transformation on the span of the first 9 basis vectors. This ensures T_2 is not written in block form but has known eigenvalues. We have plotted the difference in norm between the first $j \times j$ block of each $Q_n^* T_l Q_n$ and the diagonal operator formed via the largest j eigenvalues for $j = 1, 2, 3$ and 4 in Figure 9.1. The plot clearly shows the exponential convergence.

Example 9.5.2 (Convergence to extremal parts of the spectrum)**.** To see why we may need some condition on the spectrum for convergence of the IQR algorithm to the extreme parts of the spectrum, we consider Laurent and Toeplitz operators with symbol given by a trigonometric polynomial

$$a(t) = \sum_{j=-k}^{k} a_j t^j.$$

Given such a symbol, we define Laurent and Toeplitz operators

$$L(a) = \left(\begin{array}{ccc|cccc} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & a_0 & a_{-1} & a_{-2} & a_{-3} & a_{-4} & \cdots \\ \cdots & a_1 & a_0 & a_{-1} & a_{-2} & a_{-3} & \cdots \\ \hline \cdots & a_2 & a_1 & a_0 & a_{-1} & a_{-2} & \cdots \\ \cdots & a_3 & a_2 & a_1 & a_0 & a_{-1} & \cdots \\ \cdots & a_4 & a_3 & a_2 & a_1 & a_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{array}\right), \quad T(a) = \begin{pmatrix} a_0 & a_{-1} & a_{-2} & \cdots \\ a_1 & a_0 & a_{-1} & \cdots \\ a_2 & a_1 & a_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

acting on $l^2(\mathbb{Z})$ and $l^2(\mathbb{N})$ respectively. Note that $L(a)$ is always normal, whereas $T(a)$ need not be (see for example [BS99]). A simple example is $a(t) = t$ which gives rise to the bilateral and unilateral shifts $L(a) = U_1$ and $T(a) = S$. In this case, both of these operators are invariant under iterations of the IQR algorithm and hence their finite sections $P_m Q_n^* T Q_n P_m$ always have spectrum $\{0\}$. In the case of $L(a)$, this is an example of spectral pollution, whereas in the case of $T(a)$ this does not capture the extremal parts of the spectrum. Regarding pure finite section, the following beautiful result is known:

Theorem 9.5.3 ([SS60]). *If a is a trigonometric polynomial then we have the following convergence in the Hausdorff metric*

$$\lim_{m\to\infty} \mathrm{Sp}(P_m L(a) P_m) = \lim_{m\to\infty} \mathrm{Sp}(P_m T(a) P_m) = \bigcap_{r\in(0,\infty)} \mathrm{Sp}(T(a_r)) =: \Upsilon(a),$$

where $a_r(t) = a(rt)$. Furthermore, this limit set is a connected finite union of analytic arcs, each pair of which has at most endpoints in common.

It is straightforward to construct examples where it appears that both $\lim_{n\to\infty} \mathrm{Sp}(P_m Q_n^* T(a) Q_n P_m)$ and $\lim_{n\to\infty} \mathrm{Sp}(P_m Q_n^* L(a) Q_n P_m)$ exist and are either the extreme parts of $\mathrm{Sp}(L(a))$ or of $\Upsilon(a)$. For example, consider the symbols

$$a(t) = \frac{t^3 + t^{-1}}{2}, \quad \tilde{a}(t) = t + it^{-2}.$$

Figure 9.2 shows the outputs of the IQR algorithm and plain finite section for the corresponding Laurent and Toeplitz operators for $m = 50$ and $n = 1$ and $n = 300$. In the case of a, it appears that both limit sets are the extremal parts of $\mathrm{Sp}(L(a))$ (together with 0 if m is not a multiple of 4). Whereas in the case of $\tilde{a}$ it appears that $\lim_{n\to\infty} \mathrm{Sp}(P_m Q_n^* T(a) Q_n P_m)$ is the extremal parts of $\Upsilon(a)$ and $\lim_{n\to\infty} \mathrm{Sp}(P_m Q_n^* L(a) Q_n P_m)$ is the extremal parts of $\mathrm{Sp}(L(a))$ (again together with a finite collection of points depending on the value of m modulo 3). Curiously, in both cases, we observed convergence in the strong operator topology to block diagonal operators (up to unitary equivalence in each sub-block), whose blocks have spectra corresponding to the limiting sets (hence the dependence on the remainder of m modulo 2 or 3). However, in contrast to convergence to points in the discrete spectrum, convergence to these operators was only algebraic. This is shown in Figure 9.3, where we have plotted the Hausdorff distance between the limiting set and the eigenvalues of the first diagonal block. We also shifted the operators ($+1.1I$ for a and $-1.5iI$ for $\tilde{a}$) so that the extremal points correspond to exactly one point. In this scenario and for all considered operators (Laurent or Toeplitz), the IQR algorithm converges strongly to a diagonal operator whose diagonal entries are the corresponding extremal point of $\mathrm{Sp}(L(a))$. This convergence is also shown in Figure 9.3, and we observed a slower rate of convergence than before. This is possibly due to points from the other tips of the

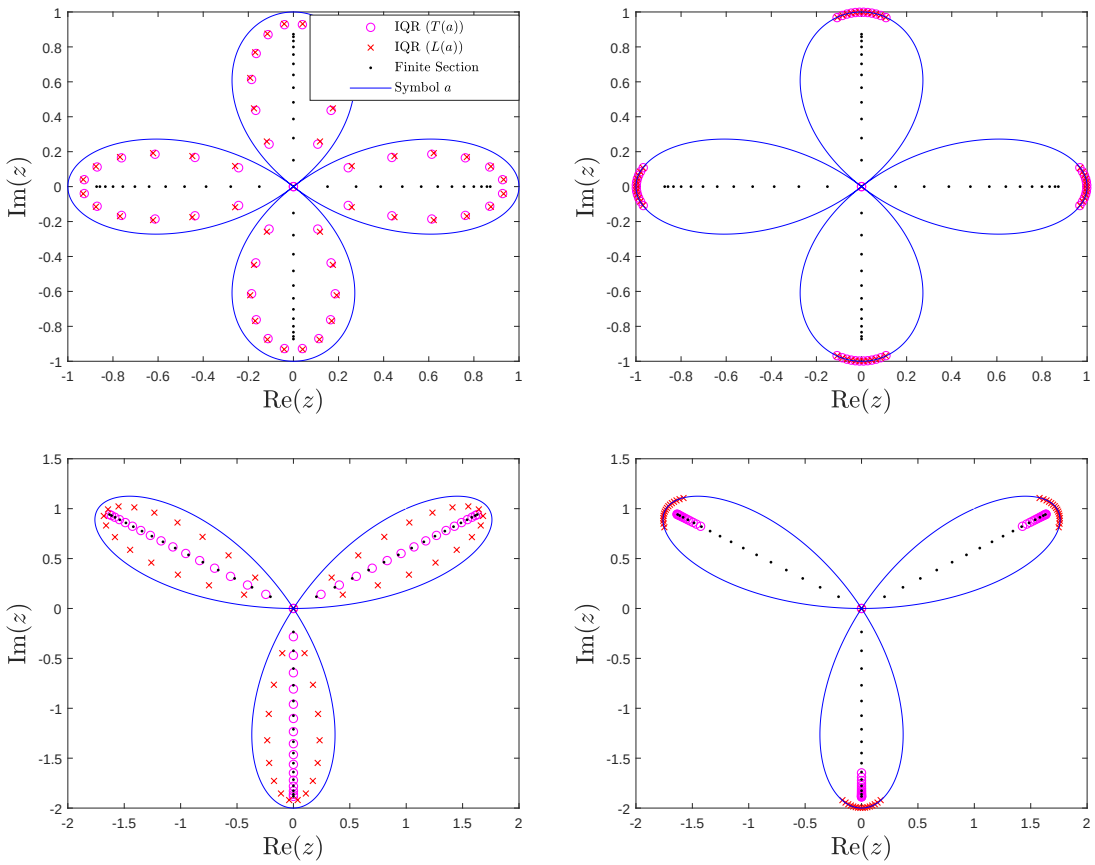


Figure 9.2: Top: Output of IQR and finite section on $T(a)$ and $L(a)$ for $m = 50$ and $n = 1$ (left), $n = 300$ (right). Bottom: Same but for the symbol $\tilde{a}$. In both cases for a given symbol b, $\mathrm{Sp}(L(b))$ is given by $\{b(z) : z \in \mathbb{T}\}$ (shown) and $\mathrm{Sp}(T(b))$ is given by $\mathrm{Sp}(L(b)) \cup \{z \in \mathbb{C}\backslash b(\mathbb{T}) : \mathrm{wind}(b, z) \neq 0\}$.

petals of $\mathrm{Sp}(L(a))$ converging as we increase n. It would be interesting to see if some form of Theorem 9.5.3 holds for the IQR algorithm (now taking $n \to \infty$). Given the examples presented here, such a statement would likely be quite complicated. However, we conjecture that if a normal operator has exactly one extreme point of its essential spectrum (and finitely many eigenvalues of magnitude greater than r_{ess}) then this extreme point will be recovered for large enough m.

Example 9.5.4 (IQR and avoiding spectral pollution)**.** In this example, we consider whether the IQR may be used as a tool to avoid spectral pollution. Sometimes when considering $\mathrm{Sp}(P_m T P_m)$, spectral pollution can be detected by changing m (edge states which correspond to spectral pollution are often unstable, but this is not always the case). In general, $\mathrm{Sp}(P_m Q_n^* T Q_n P_m)$ can be considered as a generalised version of finite section with a finite number (n) of IQR iterates being performed on the infinite-dimensional operator before truncation. If Q_n is unitary, then this changes the basis before truncation, and such a change may reduce (or change) spectral pollution allowing it to be detected. Here we consider

$$T_3 = \begin{pmatrix} 0 & 3 & & & \\ 3 & 0 & 1 & & \\ & 1 & 0 & 3 & \\ & & 3 & 0 & \ddots \\ & & & \ddots & \ddots \end{pmatrix}.$$

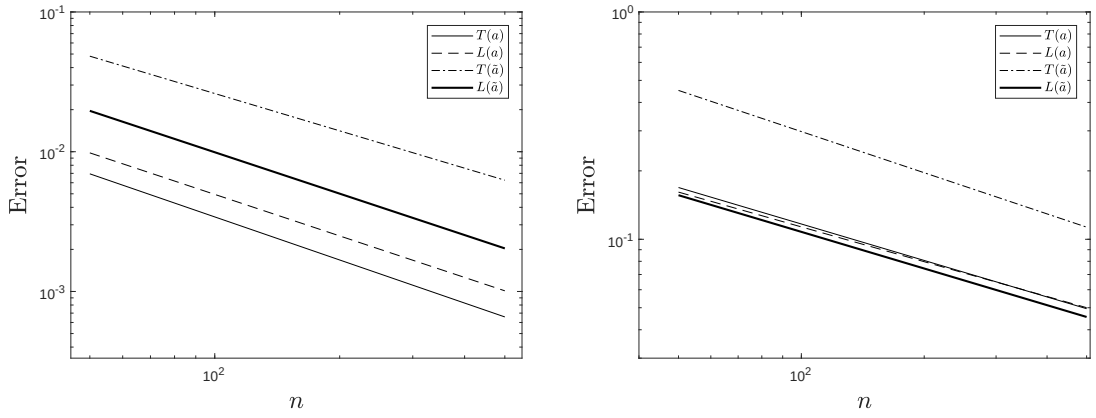


Figure 9.3: Left: Algebraic convergence to block diagonal operators. Right: Algebraic convergence to diagonal operators. In both cases we have plotted the difference in eigenvalues of the first block as we increase n.

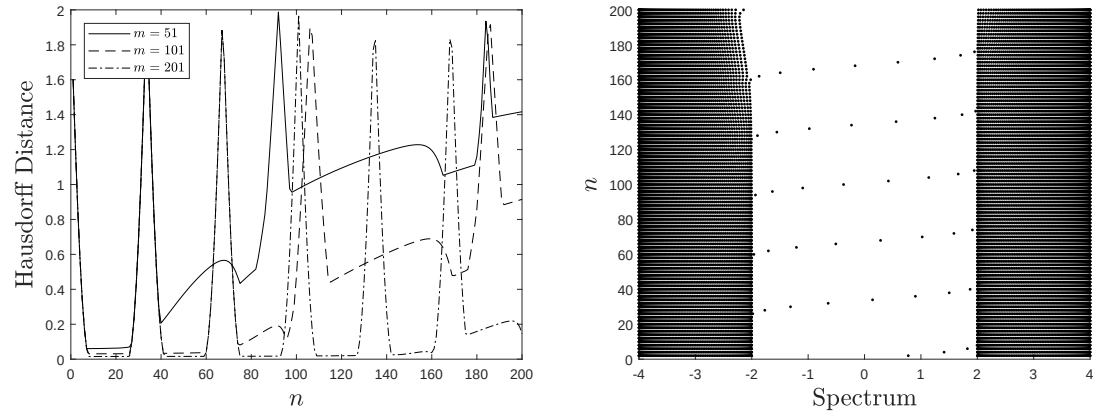


Figure 9.4: Left: $d_{\mathrm{H}}(\mathrm{Sp}(P_m Q_n^*(T_3+0.2I)Q_n P_m)-0.2, \mathrm{Sp}(T_3))$ as a function of n for different m. Right: $\mathrm{Sp}(P_m Q_n^*(T_3+0.2I)Q_n P_m)-0.2$ as a function of n for $m = 201$. Note the crossing of eigenvalues across the spectral gap.

The spectrum of T_3 is $[-4,-2] \cup [2,4]$. However, if m is odd then $0 \in \mathrm{Sp}(P_m T_3 P_m)$. We shifted the operator by considering $T_3 + 0.2I$ (and then shifted back for the spectrum). Figure 9.4 shows the Hausdorff distance between $\mathrm{Sp}(P_m Q_n^*(T_3 + 0.2I)Q_n P_m) - 0.2$ and $\mathrm{Sp}(T_3)$ as n varies for different m. The spikes in the distance correspond to eigenvalues leaving the interval $[-4,-2]$ and crossing to $[2,4]$ (also shown in Figure 9.4). The increase in distance as m decreases (for large n) is due less of the interval $[-4,-2]$ being approximated. It appears that the IQR algorithm can be an effective tool at detecting spectral pollution - certainly a mixture of varying m and n will be more effective than just varying m.

Another example of this is given by the operator $L(a)$ considered previously. For fixed m we found that

$$\lim_{n\to\infty} \sup_{z\in\mathrm{Sp}(P_m Q_n^* L(a) Q_n P_m)} \mathrm{dist}(z, \mathrm{Sp}(L(a))) = 0.$$

However, for finite section, spectral pollution occurs for all large m

$$\lim_{m\to\infty} \sup_{z\in\mathrm{Sp}(P_m L(a) P_m)} \mathrm{dist}(z, \mathrm{Sp}(L(a))) > 0$$

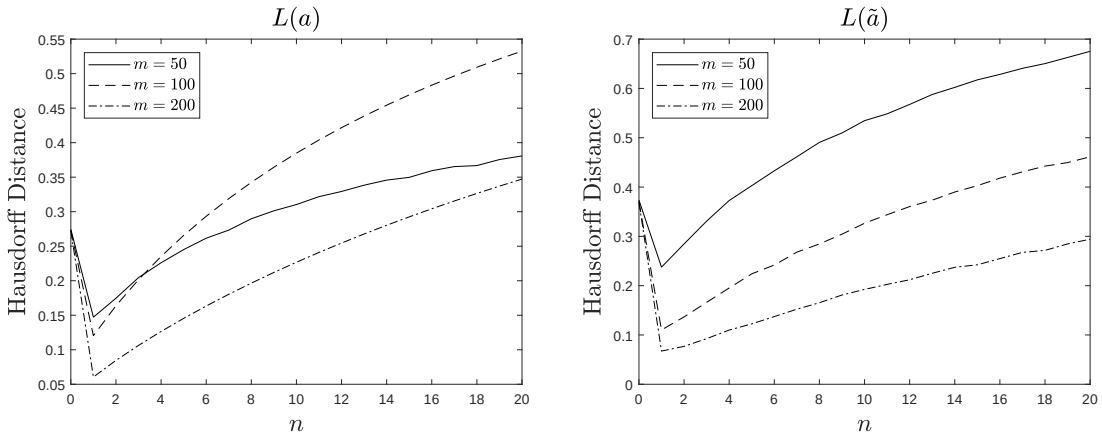


Figure 9.5: Left: $d_{\mathrm{H}}(\mathrm{Sp}(P_m Q_n^* L(a) Q_n P_m), \mathrm{Sp}(L(a)))$ as a function of n for different m. $d_{\mathrm{H}}(\mathrm{Sp}(P_m Q_n^* L(\tilde{a}) Q_n P_m), \mathrm{Sp}(L(\tilde{a})))$ as a function of n for different m.

and the IQR algorithm can only recover the extreme parts of the spectrum

$$\lim_{n\to\infty} d_{\mathrm{H}}(\mathrm{Sp}(P_m Q_n^* L(a) Q_n P_m), \mathrm{Sp}(L(a))) > 0.$$

Despite this, we found that for small fixed $n > 0$ it appears that

$$\lim_{m\to\infty} d_{\mathrm{H}}(\mathrm{Sp}(P_m Q_n^* L(a) Q_n P_m), \mathrm{Sp}(L(a))) = 0.$$

This is shown in Figure 9.5 with similar results for $L(\tilde{a})$.

9.5.2 Numerical examples II: non-normal operators

Although Theorem 9.2.9 considers normal operators, Theorems 9.2.13 and 9.2.15 suggest the IQR algorithm may also be useful for non-normal operators. Indeed, the results presented here demonstrate that in practice the IQR algorithm can work very well for non-normal problems. If an infinite matrix A has m isolated eigenvalues $\{\lambda_1, \ldots, \lambda_m\}$ (repeated according to multiplicity) outside $r_{\mathrm{ess}}(A)$ (the essential spectral radius), then Theorems 9.2.13 and 9.2.15 suggest that the eigenvalues will appear on the diagonal of $P_m Q_n^* A Q_n P_m$ as $n \to \infty$, i.e.

$$\mathrm{Sp}(P_m Q_n^* T Q_n P_m) \longrightarrow \{\lambda_1, \ldots, \lambda_m\}, \qquad \text{as } n \to \infty.$$

We will verify this numerically in the next examples. However, we will see that not only do we get convergence to the eigenvalues, but often we also pick up parts of the boundary of the essential spectrum (this was the case when considering $T(a)$ but appeared not to be the case for $T(\tilde{a})$). This phenomenon is not accounted for in the previous exposition where normality was crucial for proving Theorem 9.2.9.

Example 9.5.5 (Recovering the extremal part of the spectrum)**.** For example, let

$$A = \begin{pmatrix} a_1 & i & & & \\ 1 & a_2 & i & & \\ & 1 & a_3 & i & \\ & & 1 & a_4 & \ddots \\ & & & \ddots & \ddots \end{pmatrix},$$

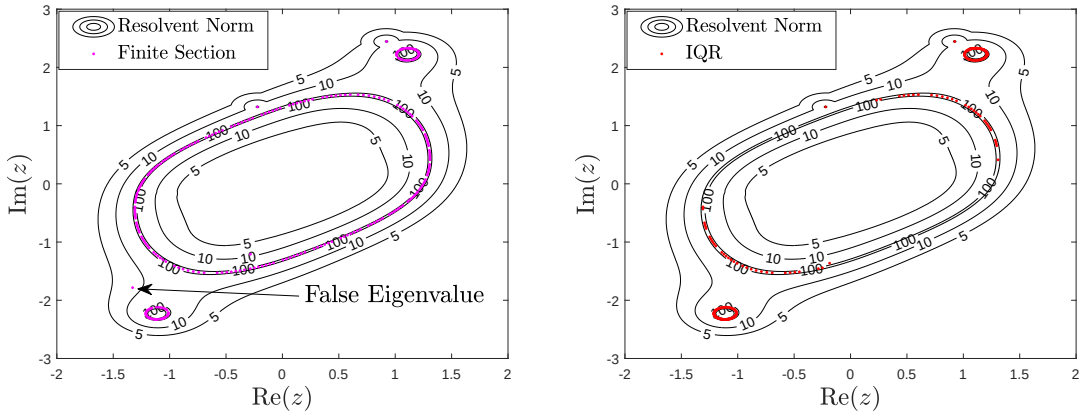


Figure 9.6: Left: $\mathrm{Sp}_\epsilon(A)$ plotted as contours of the resolvent norm, as well as $\mathrm{Sp}(P_m A P_m)$ for $m = 300$ with the false eigenvalue (recall that $\mathrm{Sp}(A) \subseteq \mathrm{Sp}_\epsilon(A)$). Right: Output of the IQR algorithm $\mathrm{Sp}(P_m Q_n^* A Q_n P_m)$ for $m = 300$ and $n = 1000$.

where $a_j = 5\cos(j)/4 + 2i\sin(j)$. To gain an accurate picture of the spectrum, note that A is banded and hence we can compute approximates to the pseudospectrum via the methods in Chapter 3. It is possible to detect spectral pollution outside of $\mathrm{Sp}_\epsilon(A)$ if we can approximate it well. The phenomenon of spectral pollution occurs for A: namely, the computed spectrum $\mathrm{Sp}(P_m A P_m)$ contains elements that have nothing to do with $\mathrm{Sp}(A)$. This is visualised in Figure 9.6 (left), an example with spectral pollution $z \notin \mathrm{Sp}_{1/10}(A)$ where the same phenomenon occurs for larger m. The spectral pollution phenomenon discussed in detail in §7.1 and §7.3.2.

Remark 9.5.6. *This example demonstrates that, in general, the finite section method is not always suitable for computing spectra. Rather than working with square sections of the infinite matrix A, one should work with $P_n A P_m$, where the parameters n and m are allowed to vary independently. Indeed, this idea was used in previous chapters (see §3.1.3) and also used implicitly in the IQR algorithm (see §9.3).*

We have also run the IQR algorithm with $n = 1000$ for A, shown in Figure 9.6 (right). We see that if one takes a finite section after running the IQR algorithm, then part of the boundary of the essential spectrum also appears, along with the discrete spectrum $\mathrm{Sp}_d(A)$. Note that the part of the boundary that is captured is the extreme part (points with largest modulus). It seems that after running the IQR algorithm, the spectral information from the largest isolated eigenvalues and the largest approximate point spectrum is 'squeezed up' to the upper and leftmost portions of the matrix. This is not completely counter-intuitive given (9.1.6), and is what normally happens in finite dimensions. The IQR iterates, in this case, converge to an upper triangular matrix (analogous to the finite-dimensional case) in agreement with Theorems 9.2.13 and 9.2.15.

Example 9.5.7 (PT-symmetry in quantum mechanics)**.** Finally, we consider a so called PT-symmetric operator (non-normal), demonstrating the same phenomena. Recall from §3.6.3 that 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 $\hat{p} \to -\hat{p}, i \to -i$. Further discussion of these operators is also given in §3.6.3. We consider an operator on $l^2(\mathbb{Z})$ of the form

$$(H_1 x)_n = x_{n-1} + x_{n+1} + V_n x_n.$$

This commutes with (the discrete version of) PT precisely when the potential has even real part and odd

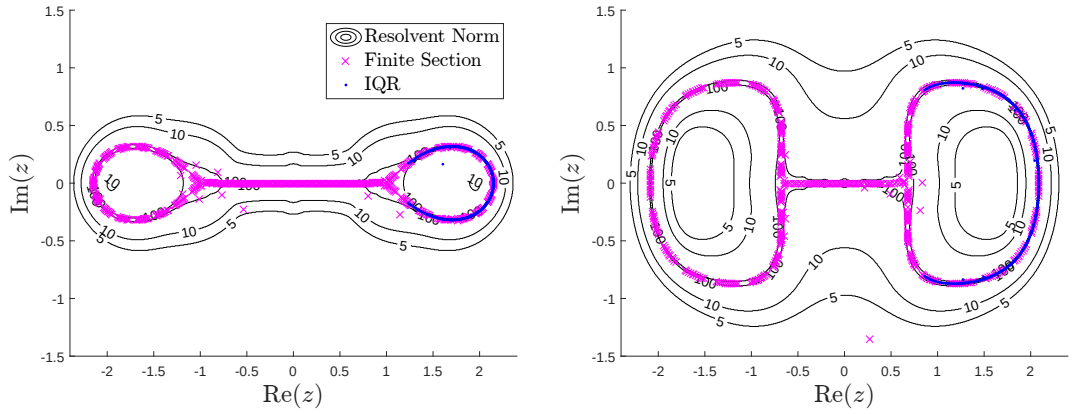


Figure 9.7: The plots show finite sections $\mathrm{Sp}(P_m H_1 P_m)$ (magenta) and (shifted) $\mathrm{Sp}(P_m Q_n^* H_1 Q_n P_m)$ IQR iterates (blue) along with converged resolvent norm contours for $\gamma = 1$ (left) and $\gamma = 2$ (right). Both figures are for $m = 500, n = 3000$ and show the convergence to the extremal parts of the spectrum.

imaginary part. We tested the IQR algorithm on the potential

$$V_n = \begin{cases} \cos(n) + i\gamma \sin(n), & \mod(n,2) = 0 \\ 0, & \mod(n,2) = 1 \end{cases},$$

and found similar results for other potentials. Figure 9.7 shows the same qualitative behaviour as the last example for $\gamma = 1, 2$ at $m = 500, n = 3000$. We shifted by 2.2 and 2.15 for $\gamma = 1, 2$ respectively. For comparison we have shown converged resolvent norms. We found that spectral pollution with no IQR iterates was consistent as we varied m. However, for a fixed m, increasing the number of iterates ($n \to \infty$) caused $\mathrm{Sp}(P_m Q_n^* H_1 Q_n P_m)$ to approach the extremal part of the spectrum.

9.5.3 Numerical examples III: random non-Hermitian operators and boundary conditions

In this final section, we explore examples where the $P_m Q_n^* T Q_n P_m$ naturally give rise to periodic boundary conditions (this was already seen for some examples of Laurent operators in §9.5.1). Both examples discussed here are physically motivated random tridiagonal operators on the lattice $\mathbb{Z}$ and are pseudoergodic - a class that will be explored in detail in Chapter 10. One of the key applications of studying such random operators can be found in condensed matter physics. The discrete models below have been used to study conductivity of disordered media, flux lines in superconductors and asymmetric hopping particles. Many such operators are also the discretisation of certain stochastic differential equations. As we will demonstrate, the IQR method can be a powerful way of avoiding spectral pollution caused by unnatural 'open' boundary conditions in forming the finite section $P_m T P_m$. In both of these examples, periodic boundary conditions are natural (this will be justified in Chapter 10), and we find that taking finite sections after iterating the IQR algorithm captures periodic boundary conditions.

Example 9.5.8 (Hopping sign model in sparse neural networks)**.** The first example is a non-normal operator with random sub and super-diagonals, first studied by Feinberg and Zee [FZ99a, HOZ03, CWCL11]. The usual 'hopping sign model' is defined via

$$(H_2 x)_n = x_{n-1} + b_n x_{n+1},$$

with $b_n \in \{\pm 1\}$ (say independent Bernoulli with parameter $p = 1/2$). This describes a particle 'hopping' on $\mathbb{Z}$ and can be mapped into a (complex-valued) random walk. We consider a slightly different operator described by

$$(H_3 x)_n = s^-_{n-1} \exp(-g) x_{n-1} + s^+_n \exp(g) x_{n+1},$$

and appearing in [AHN16] in the context of sparse neural networks. We assume that g is real and non-negative and that $s^\pm_j$ are i.d.d. random variables with Bernoulli distribution p. In other words

$$\mathbb{P}(s^\pm_j = 1) = 1 - \mathbb{P}(s^\pm_j = -1) = p.$$

We only consider $g = 1/10$ and $p = 1/2$, but will vary p in an effort to compute the spectrum of H_3 which only depends on the support of the distribution of the $s^\pm_j$'s. It is easy to prove that the spectrum (and pseudospectrum) of H_3 is almost surely constant and that there is no inessential spectrum. Furthermore, one can show that $\mathrm{Sp}(H_3)$ is contained in the annulus $\{z \in \mathbb{C} : 2\sinh(g) \leq |z| \leq 2\cosh(g)\}$.

Finite section calculations associated with this operator have some interesting properties and are extensively studied in [AHN16]. If one projects using the standard basis of $l^2(\mathbb{Z})$ then one obtains matrices of the form

$$M^1_n = \begin{pmatrix} 0 & s^-_{-n+1}\exp(-g) & & \\ s^+_{-n+1}\exp(g) & 0 & \ddots & \\ & \ddots & \ddots & s^-_{n-1}\exp(-g) \\ & & s^+_{n-1}\exp(g) & 0 \end{pmatrix}.$$

If we use open boundary conditions (i.e. we simply project onto the space spanned by $\{e_{-n}, ..., e_n\}$) then one can 'gauge' away g by a similarity transformation, leading to

$$M'_n = \begin{pmatrix} 0 & s^-_{-n+1} & & \\ s^+_{-n+1} & 0 & \ddots & \\ & \ddots & \ddots & s^-_{n-1} \\ & & s^+_{n-1} & 0 \end{pmatrix}.$$

On the other hand, the use of periodic boundary conditions leads to the matrix

$$M^2_n = \begin{pmatrix} 0 & s^-_{-n+1}\exp(-g) & & s^+_n\exp(g) \\ s^+_{-n+1}\exp(g) & 0 & \ddots & \\ & \ddots & \ddots & s^-_{n-1}\exp(-g) \\ s^-_n\exp(-g) & & s^+_{n-1}\exp(g) & 0 \end{pmatrix},$$

which does not suffer from this setback.

In [AHN16] this phenomenon was studied via localisation of the eigenvalues of M^2_n, in particular using the Lyapunov exponent $\kappa(z)$ which is equal to the inverse of the localisation length. An eigenfunction ψ with eigenvalue z localised around x_0 behaves approximately as

$$|\psi(x)| \sim \exp(-\kappa(z)\,|x - x_0|).$$

If one defines recursively

$$y_{n+1}(z) = \exp(g)\frac{\psi_{n+2}}{\psi_{n+1}} = -(s^-_{n-1}/s^+_n)/y_n(z) + z/s^+_n$$

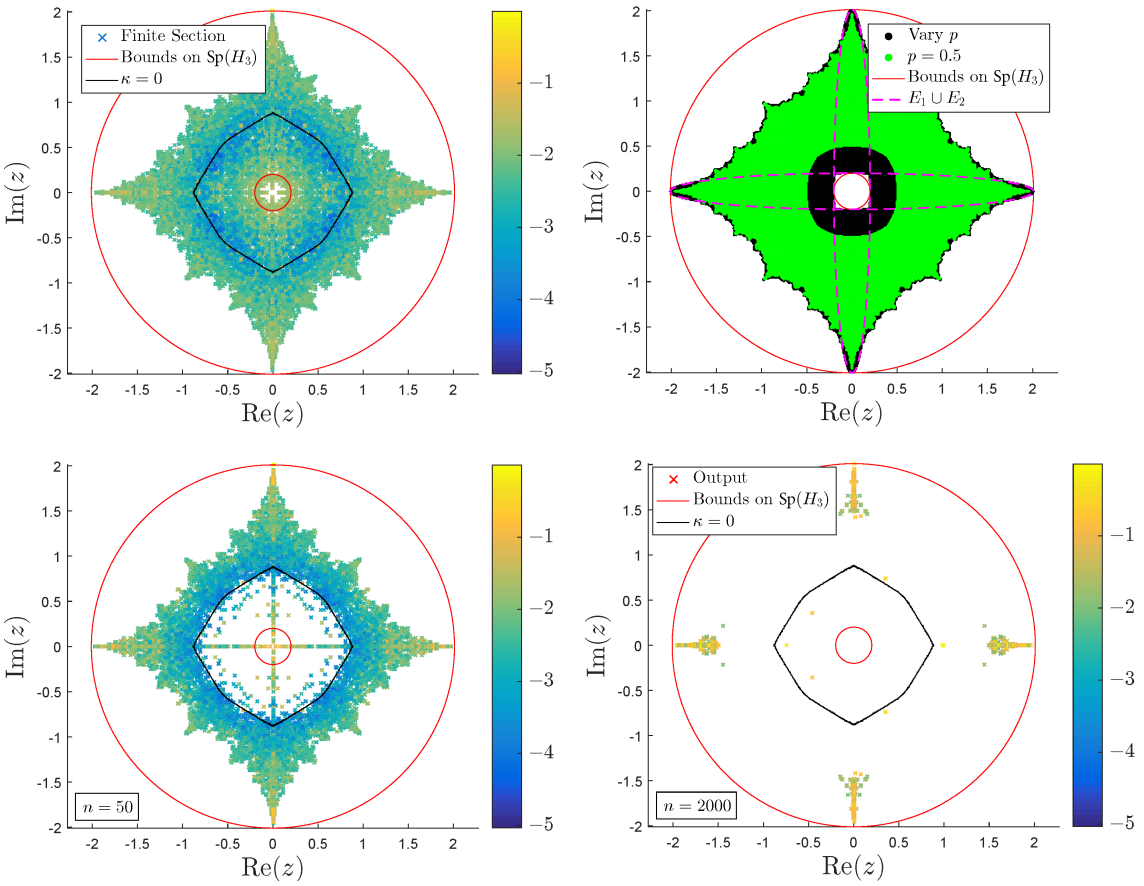


Figure 9.8: Top: Output of finite section over a random sample of 200 matrices of size 200 (left) and the estimates using pseudospectral techniques (right). Bottom: The output of IQR over 200 samples computing $\mathrm{Sp}(P_m Q_n^* H_3 Q_n P_m)$ for $m = 200$ and $n = 50$ (left), $n = 2000$ (right). Note that a few iterates seem to agree with periodic boundary conditions. Increasing the number of iterates further leads to convergence to the extremal parts of the essential spectrum.

then (in the limit of large system sizes)

$$\kappa(z; g) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{j=-N}^{N} \big(\log|y_j(z)| - g\big).$$

This is known as the transfer matrix approach. For fixed z, as we increase g, $\kappa(z; g)$ becomes negative. The heuristic is that a hole opens up in the spectrum corresponding to a negative Lyapunov exponent. Eigenvalues of M_n^2 inside the hole are swept up and become delocalised moving to the rim of the hole, whereas those outside remain largely undisturbed. Eigenvalues of M_n^1 inside the negative κ zone correspond to edge states due to the finite system size approximation.

Figure 9.8 shows the output of a sample of 200 finite sections with open boundary conditions and matrix size 200. We have also shown the annular region that bounds the spectrum, as well as the contour $\kappa = 0$. In order to calculate κ, we calculated the above sum on a grid with large N to ensure convergence. The colour bar corresponds to the inverse participation ratio (log scale) of normalised eigenfunctions defined by

$$1/P \equiv \frac{\sum_j |\psi_i|^4}{\sum_j |\psi_i|^2}.$$

Note that this has a maximum value of 1 (localised) and a minimum value of $1/N$ (delocalised), N being the size of the matrix. Open boundary conditions produce spectral pollution in the hole with localised

eigenfunctions, and the contour $\kappa = 0$ corresponds to the delocalised region. In order to compare to the spectrum of the infinite operator on $l^2(\mathbb{Z})$, we have plotted $\mathrm{Sp}_\epsilon(H_3)$, for $\epsilon = 10^{-2}$, calculated using matrix sizes of order 10^5. We note that the spectrum is independent of $p \in (0,1)$ so we have also shown the union of these estimates over $p = \{k/100\}_{k=1}^{99}$. Although the algorithm used to compute the pseudospectrum is guaranteed to converge to $\mathrm{Sp}_\epsilon(H_3)$, there are regions in the complex plane where this convergence is very slow. Taking unions over p is simply a way to speed up this convergence. We found upon taking ϵ smaller that the spectrum appeared to have a fractal-like nature. It also appears that the hole in the spectrum corresponds to the boundary of two ellipses. It is easy to prove that the ellipse

$$E_1 = \{\exp(g + i\theta) + \exp(-g - i\theta) : \theta \in [0, 2\pi)\}$$

is contained in $\mathrm{Sp}(H_3)$ and that the spectrum (and pseudospectrum) of H_3 has fourfold rotational symmetry. Denoting the rotation of E_1 by $\pi/4$ as E_2 we have shown $E_1 \cup E_2$ in the figure.

Figure 9.8 also shows the effect of IQR iterations over random samples of size 200 for $m = 200$ and $n = 50$ and 2000. Remarkably, as we increase n, a few iterations are enough to capture periodic boundary conditions and sweep away the localised edge states. We have also shown the inverse participation ratio which, although now is defined with respect to a new basis, still gives an indication of how 'diagonal' the matrix $P_m Q_n^* H_3 Q_n P_m$ is. If we increase n further, the output approaches the edge of the spectrum with eigenvectors becoming more localised (in the new basis). We found exactly the same phenomena to occur if we shifted the operator H_3 with convergence to the corresponding extremal part of the essential spectrum.

Example 9.5.9 (NSA Anderson model in superconductors)**.** Finally, we consider a non-normal operator with no inessential spectrum where the IQR algorithm does not seem to converge to the boundary of the essential spectrum, but rather to a curve associated with periodic boundary conditions in the large system size limit.

We revisit the NSA Anderson model from §3.6.2, introduced by Hatano and Nelson in the context of vortex pinning in type-II superconductors [HN96]. The operator in $\mathcal{B}(l^2(\mathbb{Z}))$ can be written as

$$(H_4 x)_n = \exp(-\tau) x_{n-1} + \exp(\tau) x_{n+1} + V_n x_n, \tag{9.5.1}$$

where $\tau > 0$ and V is a random potential. This operator also has applications in population biology [NS98] and the self-adjoint version of this model is widely studied for the phenomenon of Anderson localisation (absence of diffusion of waves) [And58, BJZ$^+$08]. In the NSA case, complex values of the spectrum indicate delocalisation. Note that we now have randomness on the diagonal with fixed coupling coefficients $\exp(\pm\tau)$.

Standard finite section produces real eigenvalues since the matrix $P_m H_4 P_m$ is similar to a real symmetric matrix. However, truncating the operator and adopting periodic boundary conditions gives rise to the famous 'bubble and wings'. If $V = 0$ then the spectrum is an ellipse $E = \{\exp(\tau + i\theta) + \exp(-\tau - i\theta) : \theta \in [0, 2\pi)\}$, but as we increase the randomness wings appear on the real axis. For a study of this phenomenon and the described phase transition, we refer the reader to [FZ99a]. Goldsheid and Khoruzhenko have studied the convergence of the spectral measure in the periodic case as $N \to \infty$ in [GK98], N being the number of sites. In general, this can be very different from the spectrum of the operator on $l^2(\mathbb{Z})$ given by (9.5.1), highlighting the difficulty in computing the spectrum.

We consider the case $\tau = 1/2$ with V_n i.i.d. Bernoulli random variables taking values in $\{\pm 1\}$ with equal probability $p = 1/2$. Again, there is no inessential spectrum and the spectrum/pseudospectrum is

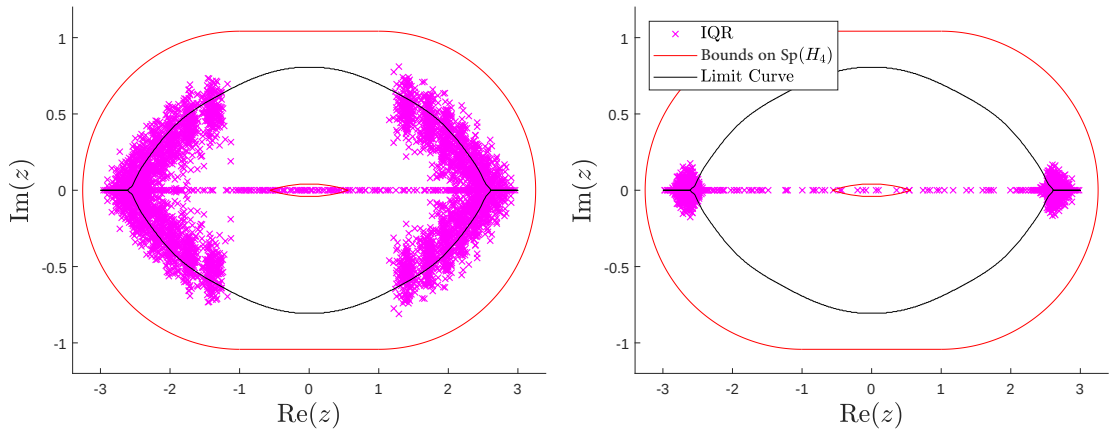


Figure 9.9: The output of IQR over 200 samples computing $\mathrm{Sp}(P_m Q_n^* H_4 Q_n P_m)$ for $m = 30$ and $n = 15$ (left), $n = 300$ (right). Note that we appear to recover the periodic limit curve and increasing the number of iterates causes the IQR algorithm to converge to the extremal parts. Applying shifts allows us to recover the extremal parts of the limit curves.

constant almost surely, depending only on the support of the distribution of the V_n. The following inclusion is also known, which bounds the spectrum:

$$\mathrm{Sp}(H_4) \subset (\overline{\mathrm{conv}}(E) + [-1, 1]) \cap (E + B_1),$$

where $\overline{\mathrm{conv}}(E)$ its closed convex hull of E and B_1 denotes the closed unit disk. The choice of τ ensures the spectrum has a hole in it. One may calculate the Lyapunov exponent, either by the transfer matrix approach or by calculating a potential related to the density of states. The limiting distribution of the eigenvalues of finite section with periodic boundary conditions is given by the complex curve

$$\{z \in \mathbb{C} \backslash \mathbb{R} : \kappa(z) = 0\} \cup \left\{ x \in \mathrm{supp}(dN) : \lim_{\delta \downarrow 0} \kappa(x + i\delta) > 0 \right\},$$

where dN denotes the density of states.

The output of the IQR algorithm for $m = 30$ and $n = 15$ and $n = 300$ over 200 random samples are shown in Figure 9.9. Note that if we took $n = 0$, the spectrum would be real in stark contrast to Figure 9.9. Taking a small number of IQR iterates approximates the bubble and wings with a few remaining real eigenvalues. However, upon increasing n, the output does not seem to converge to the extremal parts of the spectrum, but seems to remain stuck on the limit curve with the operator $P_m Q_n^* H_4 Q_n P_m$. Shifting by $+4iI$ caused the output to recover the top part of the limit curve.

Remark 9.5.10. *For any operator A that has Q_n unitary, the essential spectrum and spectrum of $Q_n^* A Q_n$ are equal to that of A. As the above two examples suggest, taking a small value of n could be used as a method of testing eigenvalues of finite section methods that correspond to finite system size effects, such as open boundary conditions. This could be used in quasiperiodic systems or systems with very few symmetries, where there is no obvious choice of appropriate boundary conditions.*

Chapter 10

Pseudoergodic Operators and Finite Section

There is a growing literature on random non-self-adjoint infinite matrices with motivation ranging from condensed matter physics to neural networks. Many of these operators fall into the class of 'pseudoergodic' operators (already seen in §9.5.3), which allows the elimination of probabilistic arguments when studying spectral properties.

In this chapter, based on [Col19c], we demonstrate a wide class of operators where the computation of the pseudospectrum is easier than that of the spectrum. We prove that, for pseudoergodic operators, the pseudospectra of finite section approximates with periodic boundary conditions converge to the pseudospectrum of the full infinite-dimensional operator. Numerical evidence of this was given in §3.6.2 and §9.5.3, and we refer the reader to [Col19c] for further examples. Our results hold in any dimension, not just for banded bi-infinite matrices, and can be considered as a generalisation of the well-known classical result for banded Laurent operators and their circulant approximations.

In terms of the SCI hierarchy, this gives a Σ_1 classification for the pseudospectral problem. Chapter 3 extends this far beyond the pseudoergodic case. Nevertheless, the results of this chapter are interesting in their own right, since they use the classical finite section method (as opposed to the local resolvent approach of Chapter 3). Furthermore, we show that the result carries over to the non-Hilbertian setting of pseudoergodic operators acting on vector-valued l^p spaces for arbitrary $p \in [1, \infty]$ (which also provides a Π_2 tower for the spectrum if we then shrink the pseudospectrum as $\epsilon \downarrow 0$).

10.1 Pseudoergodic Operators

Random matrices appear in a wide number of contexts throughout the sciences, ranging from applied physics through to areas of pure mathematics such as number theory [Meh04, BK99, Mui82, Ede88]. In particular, the study of random Jacobi operators can be traced back at least as far as the famous Anderson model [And58, And61]. Over the past two decades, there has been a considerable amount of interest in the study of random non-self-adjoint (NSA) operators on separable Hilbert spaces. As well as their interesting mathematical properties, motivation for studying such operators can be found in condensed matter physics: conductivity of disordered media, flux lines in superconductors and asymmetric hopping

particles, and even in population biology [FZ99a, HN97, NS98, HOZ03, AHN16]. One is often interested in how the spectrum and pseudospectrum behave under truncation of the operators to finite matrices [FZ99b, GK00, LR12, TE05, CWCL13], which can also lead to algorithms computing spectral properties numerically. Many of the operators studied in the above papers are pseudoergodic (also sometimes referred to as stochastic Laurent matrices in the $l^2(\mathbb{Z})$ case), which roughly means that every possible finite pattern in the matrix elements appears somewhere up to arbitrary precision (see Definition 10.1.1). This allows the treatment of such random operators in a deterministic fashion, leading to simplified proofs of spectral properties which often depend only weakly on the distribution of matrix elements (e.g. on its support).

The core result of this chapter is that in the limit of increasing system size, pseudospectra converge to the pseudospectra of the full operator if we apply periodic boundary conditions. This result was conjectured in [DNS99] for a particular one-dimensional lattice model but has so far remained an open problem.[1] The result presented here holds for any dimension and any finite range interaction pseudoergodic operator (not just tridiagonal). In other words, the passage from finite volumes to the infinite volume case is continuous with respect to the pseudospectrum. This can be considered as a complement to the well-known corresponding result for banded Laurent operators - it is precisely the fact that pseudoergodic operators 'look the same' under translational shifts that allows us to prove this result. The fact that we can approximate the pseudospectrum of the full infinite-dimensional operator using square matrices allows the numerical computation of pseudospectra on l^p spaces with $p \neq 2$, and we prove the convergence of pseudospectra in this case also. This is in contrast to the methods developed in Chapter 3, which use rectangular matrices and for which no such generalisation is numerically possible. As well as being of interest from the finite section point of view, our results have practical significance. A numerical example was given in §3.6.2 and further examples can be found in the paper [Col19c].

It should be mentioned that, in sharp contrast to our results, spectra of finite sections are often very different from that of the full operator, particularly in the NSA case (this was discussed in §9.5.3). The classic pseudoergodic example is the widely studied NSA Anderson model [HN97, NS98, SN98, BZ98, BSB97, HWY02], which was studied in §3.6.2 and §9.5.3. Recall that this pseudoergodic operator acts on $l^2(\mathbb{Z})$ via

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

where $\tau > 0$ and V is a (real-valued) random potential. Truncating the operator to $\mathrm{span}\{e_{-n}, ..., e_n\}$ and adopting periodic boundary conditions gives a spectrum with the famous 'bubble and wings'. Goldsheid and Khoruzhenko have studied the convergence of the spectral measure in the periodic case as $n \to \infty$ in [GK98, GK00]. This can be very different from the spectrum of the operator on $l^2(\mathbb{Z})$, highlighting the difficulty in computing the spectrum. Applying no boundary conditions and simply taking the matrix[2] $P_n A P_n$ is even worse. In this case, the matrix is similar to a real symmetric matrix, hence has a completely real spectrum. Already we can see stability playing a role - as n increases, the condition number of the similarity transform increases exponentially when $\tau \neq 0$. There are certain cases where the obvious finite section $P_n A P_n$ behaves better, and we refer the reader to [CWCL13, CWCL11] for a thorough study of the famous 'hopping sign model' where, remarkably, this is the case.

[1] Most results in the literature consider either special cases of tridiagonal pseudoergodic operators or use the theory of limit operators to write the pseudospectrum of pseudoergodic operators acting on $l^2(\mathbb{Z})$ as the union of pseudospectra over all possible periodic submatrices (see for example [Hag16a, Hag16b]), which is not helpful from a numerical perspective.

[2] Throughout this chapter, P_n will denote the orthogonal projection onto $\mathrm{span}\{e_{-n}, e_{-n+1}, ..., e_n\}$ in the case of $l^2(\mathbb{Z})$.

10.1.1 Definitions and main results

Given $A \in \mathcal{B}(l^2(\mathbb{Z}^d))$ and $i, j \in \mathbb{Z}^d$, we will denote the inner product $\langle Ae_j, e_i \rangle$ with respect to the canonical basis by $A_{i,j}$.

Definition 10.1.1 (Pseudoergodicity)**.** *Let A be a bounded operator acting on $l^2(\mathbb{Z}^d)$. Given a collection $\underline{M} = \{M_k\}_{k \in \mathbb{Z}^d}$ of compact subsets $M_k \subset \mathbb{C}$, we say that $\underline{M}$ is permissible if only finitely many of the M_k are not $\{0\}$. Given a permissible $\underline{M}$, we say that A is pseudoergodic with respect to $\underline{M}$ if $A_{i,j} \in M_{i-j}$ and the following property holds. Given any $\epsilon > 0$, finite subsets $S_k \subset \mathbb{Z}^d$ and functions $F_k : S_k \to M_k$, there exists a translation T acting on $\mathbb{Z}^d$ such that*

$$\sup_{i \in S_k} \left| A_{T(i),T(i)-k} - F_k(i) \right| < \epsilon, \quad \forall k \in \mathbb{Z}.$$

We define $\mathcal{A}(\underline{M})$ be the class of pseudoergodic operators with respect to $\underline{M}$, and Ω^d to be the class of pseudoergodic operators acting on $l^2(\mathbb{Z}^d)$.

A few remarks are in order. Note first that in the case of $d = 1$, such an operator must be banded by the assumption that only finitely many of the M_k are not $\{0\}$. Second, it is also clear that such operators must be bounded for any d. This is also true when considering these infinite matrices as operators acting on $l^p(\mathbb{Z}^d)$ (see §10.3) for which we use the same definition of pseudoergodicity. Third, the same translation T is required to work for all the diagonals simultaneously, and it is clearly sufficient only to test those diagonals that are non-zero. The idea is that every possible finite pattern is realised up to an arbitrarily small error in each of the selected diagonals. In the case of the (one-dimensional) NSA Anderson model with i.i.d. diagonals with support M, A is clearly almost surely pseudoergodic with respect to $M_1 = \{e^{\tau}\}, M_0 = M$ and $M_{-1} = \{e^{-\tau}\}$ (with all other diagonals being zero). This can be extended to the hopping sign model, random tridiagonal operators and many other variants studied in the literature.

It is straightforward to show that the maps $\mathrm{Sp}(\cdot)$ and $\mathrm{Sp}_\epsilon(\cdot)$ are constant on each $\mathcal{A}(\underline{M})$ (see [Dav01] for the case of pseudoergodic potentials, exactly the same argument can be extended to the operators in this chapter). We then let A_n^{per} denote the nth order truncation of $A \in \Omega^d$ with natural periodic boundary conditions (see §10.2, in particular equation (10.2.5)). In the Hilbert space case of $l^2(\mathbb{Z}^d)$ our main result is the following.

Theorem 10.1.2. *Let $A \in \Omega^d$ and $\epsilon > 0$, then $\lim_{n\to\infty} \mathrm{Sp}_\epsilon(A_n^{per}) = \mathrm{Sp}_\epsilon(A)$ in the Hausdorff metric and $\mathrm{Sp}_\epsilon(A_n^{per}) \subset \mathrm{Sp}_\epsilon(A)$. Define the algorithm $\Gamma_n(A) = \texttt{PseudoSpecPer}(A, n, \epsilon)$, then $\lim_{n\to\infty} \Gamma_n(A) = \mathrm{Sp}_\epsilon(A)$ in the Hausdorff metric and $\Gamma_n(A) \subset \mathrm{Sp}_\epsilon(A)$. In particular, for a given $\mathcal{A}(\underline{M})$, the problem of computing Sp_ϵ lies in Σ_1^A.*

The routine alluded to in the above theorem is written in pseudocode as

Function `PseudoSpecPer`(A, n, ϵ)
Input : n, A pseudoergodic, $\epsilon > 0$
Output: $\Gamma \subset \mathbb{C}$, an approximation to $\mathrm{Sp}_\epsilon(A)$
$G = (\mathbb{Z} + i\mathbb{Z})/n \cap B_n(0)$
for $z \in G$ **do**
$B = A_n^{per} - zI, \quad C = (A_n^{per})^* - \bar{z}I$
$S = B^*B, \quad T = C^*C$
$p = \mathtt{IsPosDef}(S - \epsilon^2), \quad q = \mathtt{IsPosDef}(T - \epsilon^2)$
$\nu(z) = \min(p, q)$
end
$\Gamma = \bigcup\{z \in G \,|\nu(z) = 0\}$.
end

Here $B_n(0)$ denotes the closed ball of radius n around 0 and the `IsPosDef` routine determines whether a matrix is positive definite (returns 1) or not (returns 0). This can be done by using an incomplete Cholesky decomposition [CRH19] (chosen for stability and speed of computation - see §3.5.2 for a discussion on efficient implementation and various methods of computing the smallest singular values). If wanted, this can be altered to use only finitely many arithmetic operations and comparisons. It is also efficient to restrict/alter the ball G to be any region of the complex plane where one is interested in computing the pseudospectrum (e.g. near a rough approximation of the pseudospectrum).

The results can also be extended to the $l^p(\mathbb{Z}^d)$ case where the definition of pseudoergodicity remains the same, and we use a superscript to denote pseudospectra with respect to the corresponding operator norm.

Theorem 10.1.3. *Let $p \in [1, \infty]$ and $A \in \Omega^d$. Then $\lim_{n\to\infty} \mathrm{Sp}_\epsilon^p(A_n^{per}) = \mathrm{Sp}_\epsilon^p(A)$ in the Hausdorff metric and $\mathrm{Sp}_\epsilon^p(A_n^{per}) \subset \mathrm{Sp}_\epsilon^p(A)$. In particular, for a given $\mathcal{A}(\underline{M})$, the problem of computing Sp_ϵ^p lies in Σ_1^A.*

This can also be used in a similar routine to `PseudoSpecPer` (see §10.3). The $2-$norm and any other $p-$norm of an $n\times n$ matrix can only differ by a factor of $\sqrt{n}$ and hence the different notions of pseudospectra are only useful for large or infinite matrices. There are examples [JT98] where this difference is important. In particular, the $1-$norm and $\infty-$norm pseudospectra are relevant for probability theory [BM92, Ber03] and heat flow [Arn51].

Remark 10.1.4. *By considering the limit $\epsilon \downarrow 0$, this provides a Π_2^A algorithm for computing the spectrum of pseudoergodic operators acting on $l^p(\mathbb{Z}^d)$. This is an interesting result since it holds in the non-Hilbertian setting when $p \neq 2$.*

10.2 The Hilbert Space Case

Throughout this section, we will use $\|\cdot\|$ to denote the standard l^2 norm and assume that $\underline{M}$ is permissible. Let $A \in \mathcal{B}(l^2(\mathbb{Z}^d))$ and define the injection modulus by

$$\sigma_1(A) := \inf\{\|Ax\| : x \in l^2(\mathbb{Z}^d), \|x\| = 1\},$$

which is equal to the smallest singular value in the case of finite matrices. Define the function

$$\psi_A(z) := \min\{\sigma_1(A - zI), \sigma_1(A^* - \bar{z}I)\}.$$

It is well-known that $\psi_A(z) = \|R(z,A)\|^{-1}$ and hence (since all the operators are bounded) we can characterise the pseudospectrum via

$$\mathrm{Sp}_\epsilon(A) = \{z \in \mathbb{C} : \psi_A(z) \leq \epsilon\}. \tag{10.2.1}$$

As part of the proof of Theorem 10.1.2, we will show that for n larger than the bandwidth of A

$$\limsup_{l\to\infty} \psi_{A_l^{per}}(z) \leq \psi_A(z) \leq \psi_{A_n^{per}}(A),$$

where A_n^{per} denotes the finite sections of A with appropriate periodic boundary conditions (see below). We begin with the simpler case of $d = 1$ and then discuss the generalisation to $d > 1$. These are then used to prove Theorem 10.1.2. Finally, we discuss the generalisation to vector-valued l^2 sequences (where matrix value entries $A_{i,j}$ are considered).

10.2.1 The case of $d = 1$

We will first deal with the case of $d = 1$ since it presents the key ideas without additional notational complexity. Given $A \in \mathcal{B}(l^2(\mathbb{Z}))$, let $A_n^o \in \mathbb{C}^{(2n+1)\times(2n+1)}$ denote the matrix formed by P_nAP_n with P_n the orthogonal projection onto $\mathrm{span}\{e_{-n}, e_{-n+1}, ..., e_n\}$. In other words, A_n^o is the matrix formed by standard finite section with open boundary conditions. Our first lemma states that in the limit $n \to \infty$, $\psi_{A_n^o}(z) \leq \psi_A(z)$ and uses only the properties of bandedness and boundedness of $A \in \Omega^1$.

Lemma 10.2.1. *Let $A \in \Omega^1$ with $A \in \mathcal{A}(\underline{M})$, then for any $z \in \mathbb{C}$, $\limsup_{n\to\infty} \psi_{A_n^o}(z) \leq \psi_A(z)$.*

Proof. Let $\delta > 0$, then by definition there exists some $\tilde{x} \in l^2(\mathbb{Z})$ of norm 1 such that $\|(A - zI)\tilde{x}\| \leq \sigma_1(A - zI) + \delta$. Let $x_k = P_k\tilde{x}/\|P_k\tilde{x}\|$ then, since $P_k\tilde{x} \to \tilde{x}$ as $k \to \infty$ and A is bounded, it follows that for large enough $k \geq k_0$, $\|(A - zI)x_k\| \leq \sigma_1(A - zI) + 2\delta$. Set $x = x_{k_0}$, which has norm one by construction. Since the support of x is finite and A is banded, we must have $(A_n^o - zI)x = (A - zI)x$ for $n \geq m + k_0$ where m is the bandwidth of A given by

$$m := \max\{|k| : M_k \neq 0\}. \tag{10.2.2}$$

Hence $\sigma_1(A_n^o - zI) \leq \|(A_n^o - zI)x\| \leq \sigma_1(A - zI) + 2\delta$. Since $\delta > 0$ was arbitrary, it follows that

$$\limsup_{n\to\infty} \sigma_1(A_n^o - zI) \leq \sigma_1(A - zI).$$

Since the adjoint is also banded, we can prove the same inequality replacing $\sigma_1(A_n^o - zI)$ by $\sigma_1((A_n^o)^* - \bar{z}I)$ and $\sigma_1(A - zI)$ by $\sigma_1(A^* - \bar{z}I)$ in exactly the same way. The result now follows. □

Given $A \in \mathcal{A}(\underline{M})$, let $L_n^{\mathrm{b.c.}}$ be a lower diagonal matrix, with matrix values uniformly bounded in n, such that $(L_n^{\mathrm{b.c.}})_{i,j} = 0$ if $j > i + m - (2n + 1)$, where i, j are indexed in $\{-n, -n + 1, ..., n\}$ and m is defined in (10.2.2). Similarly let $U_n^{\mathrm{b.c.}}$ be an upper diagonal matrix, with matrix values uniformly bounded in n, such that $(U_n^{\mathrm{b.c.}})_{i,j} = 0$ if $i > j + m - (2n + 1)$. The superscript $\mathrm{b.c.}$ stands for the boundary conditions being imposed, which are captured by these upper and lower diagonal matrices. Let $A_n^{\mathrm{b.c.}} = A_n^o + L_n^{\mathrm{b.c.}} + U_n^{\mathrm{b.c.}}$. For fixed m, letting $n \to \infty$, we can conclude in exactly the same way as the above that

$$\limsup_{n\to\infty} \psi_{A_n^{\mathrm{b.c.}}}(z) \leq \psi_A(z). \tag{10.2.3}$$

The point is that the boundary conditions only act locally. We denote periodic boundary conditions by a superscript per and in this case we fix the non-zero entries of L_n^{per} and U_n^{per} such that these are given by

$$(L_n^{per})_{i,j} \in M_{i-j-(2n+1)} \text{ if } j \leq i+m-(2n+1), \quad (U_n^{per})_{i,j} \in M_{i-j+(2n+1)} \text{ if } i \leq j+m-(2n+1).$$

Note that we are allowing any choice up to these constraints. The above ensure that the coupling between sites (i.e. the non-zero diagonals) is consistent if one defines the matrix

$$A_{n,N}^{per} := \underbrace{\begin{pmatrix} A_n^o & L_n^{per} & & & \\ U_n^{per} & A_n^o & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & L_n^{per} \\ & & & U_n^{per} & A_n^o \end{pmatrix}}_{N \text{ blocks}},$$

where each block is a $n \times n$ matrix. The following proposition is the key result in showing periodic boundary conditions are a good choice for calculating pseudospectra of pseudoergodic operators.

Proposition 10.2.2. *Consider the above set up with $A \in \mathcal{A}(\underline{M})$ (and $d = 1$). For all $n \geq m$ and all $z \in \mathbb{C}$ we have $\psi_{A_n^{per}}(z) \geq \psi_A(z)$.*

Proof. We will show that for $n \geq m$ and all $z \in \mathbb{C}$ we have $\sigma_1(A_n^{per} - zI) \geq \sigma_1(A - zI)$. Dealing with $\sigma_1(A_n^{per*} - \bar{z}I)$ is similar and together these give the result.

Let $\delta > 0$ and choose $x \in \mathbb{C}^{2n+1}$ such that $\|x\| = 1$ and $\|(A_n^{per} - zI)x\| \leq \sigma_1(A_n^{per} - zI) + \delta$. The idea is to extend x periodically and use pseudoergodicity. Extend x and $y = (A_n^{per} - zI)x$ periodically N times to obtain $x^N, y^N \in \mathbb{C}^{N(2n+1)}$ and consider the matrix $A_{n,N}^{per}$ above. Then we have

$$(A_{n,N}^{per} - zI)x^N = \begin{pmatrix} A_n^o & L_n^{per} & & \\ U_n^{per} & A_n^o & \ddots & \\ & \ddots & \ddots & L_n^{per} \\ & & U_n^{per} & A_n^o \end{pmatrix} \begin{pmatrix} x \\ x \\ \vdots \\ x \\ x \end{pmatrix} - zx^N = \begin{pmatrix} (A_n^o + L_n^{per})x \\ A_n^{per}x \\ \vdots \\ A_n^{per}x \\ (A_n^o + U_n^{per})x \end{pmatrix} - zx^N.$$

It follows that the vector

$$(A_{n,N}^{per} - zI)x^N - y^N = -\begin{pmatrix} U_n^{per}x \\ 0 \\ \vdots \\ 0 \\ L_n^{per}x \end{pmatrix} \tag{10.2.4}$$

has norm bounded by some constant, $D(m, \underline{M})$, independent of N and all x of norm 1. This is because the values of the non-zero entries of $(A_{n,N}^{per} - zI)x^N - y^N$ are uniformly bounded and there are at most $2m$ of them. The constant will in general depend on m and the maximum modulus over the set $\cup_k M_k$, but this dependence is not relevant for the argument. The idea is shown visually in Figure 10.1.

It follows that

$$\|(A_{n,N}^{per} - zI)x^N\| \leq \|y^N\| + D(m, \underline{M}) \leq N^{\frac{1}{2}}(\sigma_1(A_n^{per} - zI) + \delta) + D(m, \underline{M}).$$

$$y = \begin{pmatrix} 1 & b_1 & 0 \\ 0 & 1 & b_2 \\ b_3 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + b_1 x_2 \\ x_2 + b_2 x_3 \\ x_3 + b_3 x_1 \end{pmatrix}$$

A_2^{per} A_2^o L_2^{per}

$$A_{1,2}^{per} x^2 = \left(\begin{array}{ccc|ccc} 1 & b_1 & 0 & 0 & 0 & 0 \\ 0 & 1 & b_2 & 0 & 0 & 0 \\ 0 & 0 & 1 & b_3 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & b_1 & 0 \\ 0 & 0 & 0 & 0 & 1 & b_2 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + b_1 x_2 \\ x_2 + b_2 x_3 \\ x_3 + b_3 x_1 \\ x_1 + b_1 x_2 \\ x_2 + b_2 x_3 \\ \textcircled{x_3} \end{pmatrix}$$

Figure 10.1: Visualisation of matching across periodic extensions for $n = 1$ and $N = 2$, on $l^2(\mathbb{Z})$. For this example the diagonal takes the value 1 and the superdiagonal takes values $b_i \in M_{-1}$. The circled index corresponds to the discrepancy with y^N (in this case the missing $b_3 x_1$ term).

By construction, all entries of the periodic extension $A_{n,N}^{per}$ come from the set M_k of the corresponding diagonal with respect to which A is pseudoergodic. Hence by pseudoergodicity of A, for each desired accuracy $\epsilon > 0$ there is a desired $(2n+1)N \times (2n+1)N$ submatrix of A which is ϵ close to $A_{n,N}^{per}$. Hence we can shift the support of x^N and let $w^N \in l^2(\mathbb{Z})$ equal x^N on the corresponding $(2n+1)N$ entries and zero otherwise. Choosing ϵ sufficiently small we have $\|w^N\| = \|x^N\| = N^{\frac{1}{2}}$ and

$$\|(A - zI)w^N\| \leq N^{\frac{1}{2}}(\sigma_1(A_n^{per} - zI) + \delta) + \delta + D(m, \underline{M}).$$

It follows that

$$\sigma_1(A - zI) \leq \sigma_1(A_n^{per} - zI) + 2\delta + D(m, \underline{M})N^{-\frac{1}{2}}.$$

Letting $N \to \infty$ and then $\delta \downarrow 0$ gives $\sigma_1(A_n^{per} - zI) \geq \sigma_1(A - zI)$. □

10.2.2 The case of $d > 1$

In order to deal with $d > 1$, it is useful to introduce some notation. Given $n \in \mathbb{N}$ and $k \in \mathbb{Z}^d$ define the index sets

$$C_n := \{-n, n+1, ..., n\}^d, \quad C_{n,k} := C_n + (2n+1)k.$$

The $C_{n,k}$ partition $\mathbb{Z}^d$ and will be used to construct the relevant periodisations. Given $N \in \mathbb{N}$, we also define

$$C_N \otimes C_n := \bigcup_{k \in C_N} C_{n,k} = C_{2Nn+N+n}.$$

For $W \subset \mathbb{Z}^d$, we define the orthogonal projections $P_W, P_W^{\perp} : l^2(\mathbb{Z}^d) \to l^2(\mathbb{Z}^d)$ via

$$(P_W x)_j = \begin{cases} x_j, & \text{if } j \in W, \\ 0, & \text{otherwise} \end{cases}$$

and $P_W^{\perp} = P_{\mathbb{Z}^d \setminus W}$. We define the shift operator $S_{n,k} : l^2(\mathbb{Z}^d) \to l^2(\mathbb{Z}^d)$ via

$$(S_{n,k} x)_j = x_{j-(2n+1)k}, j \in \mathbb{Z}^d$$

which translates between the $C_{n,k}$'s. Given $A \in \Omega^d$, and with a slight abuse of notation, we can define the matrices A_n^o and A_n^{per} acting on the range of P_{C_n} (i.e. $l^2(C_n)$) via

$$A_n^o x = P_{C_n} A P_{C_n} x, \quad A_n^{per} x = \sum_{k \in C_1} P_{C_n} S_{n,-k} A P_{C_n} x. \tag{10.2.5}$$

Finally, the periodisation $A_{n,N}^{per}$ acting on $l^2(C_N \otimes C_n)$ is defined via

$$A_{n,N}^{per} x = P_{C_N \otimes C_n} \sum_{k \in C_N} S_{n,k} A P_{C_n} S_{n,-k} x.$$

All of these definitions also extend to the l^p case considered in §10.3. As before, we could have taken any values from the relevant M_k's in forming the above generalisations of L_n^{per} and U_n^{per}. However, the above definitions give a much cleaner presentation. The reader is referred to Figure 10.2 for the case of $d = 2$, which also explains the idea of the proof below. Note that the proof of Lemma 10.2.1 is identical for $d > 1$ and yields (10.2.3) for periodic boundary conditions with the general notion of bandedness given by

$$m := \max\{\|k\|_\infty : M_k \neq \{0\}\}.$$

Care is only needed for the argument in the proof of Proposition 10.2.2.

Proof of extension of Proposition 10.2.2 to $d > 1$. Again we will only show that for $n \geq m$ and all $z \in \mathbb{C}$ we have $\sigma_1(A_n^{per} - zI) \geq \sigma_1(A - zI)$. Let $\delta > 0$ and choose $x \in l^2(C_n)$ such that $\|x\| = 1$ and $\|(A_n^{per} - zI)x\| \leq \sigma_1(A_n^{per} - zI) + \delta$. Define the periodisations

$$y^N := P_{C_N \otimes C_n} \sum_{l \in C_N} S_{n,l}(A_n^{per} - zI)x, \quad x^N := P_{C_N \otimes C_n} \sum_{l \in C^N} S_{n,l} x.$$

The key step of the proof is a result analogous to (10.2.4). We have that

$$\begin{aligned}
(A_{n,N}^{per} - zI)x^N - y^N &= \left(P_{C_N \otimes C_n} \sum_{k \in C_N} S_{n,k} A P_{C_n} S_{n,-k} x^N \right) - z x^N \\
&\quad - \left(P_{C_N \otimes C_n} \sum_{k \in C_N} S_{n,k} A_n^{per} x \right) + z x^N \\
&= P_{C_N \otimes C_n} \sum_{k \in C_N} \left[\left(S_{n,k} A P_{C_n} S_{n,-k} \sum_{l \in C_N} S_{n,l} x \right) - S_{n,k} \sum_{l \in C_1} P_{C_n} S_{n,-l} A P_{C_n} x \right].
\end{aligned}$$

Since $P_{C_n} x = x$, we can simplify the first term in the sum via

$$P_{C_n} S_{n,-k} \sum_{l \in C_N} S_{n,l} x = P_{C_n} x.$$

This yields

$$(A_{n,N}^{per} - zI)x^N - y^N = P_{C_N \otimes C_n} \sum_{k \in C_N} \left(S_{n,k} A P_{C_n} x - S_{n,k} \sum_{l \in C_1} P_{C_n} S_{n,-l} A P_{C_n} x \right). \tag{10.2.6}$$

Since $n \geq m$ we can write

$$A P_{C_n} x = \sum_{t \in C_1} P_{C_{n,t}} A P_{C_n} x.$$

Note that the terms corresponding to $t = 0$ cancel in (10.2.6). We also have the relation

$$P_{C_n} S_{n,-l} = S_{n,-l} P_{C_{n,-l}}.$$

Putting these together in (10.2.6), we arrive at

$$\begin{aligned}(A_{n,N}^{per} - zI)x^N - y^N &= P_{C_N \otimes C_n} \sum_{k \in C_N} S_{n,k} \left[\left(\sum_{t \in C_1} P_{C_{n,t}} \right) - \sum_{l \in C_1} S_{n,-l} P_{C_{n,-l}} \sum_{t \in C_1} P_{C_{n,t}} \right] P_{C_n}^{\perp} A P_{C_n} x \\ &= P_{C_N \otimes C_n} \sum_{k \in C_N} \sum_{t \in C_1} \left(S_{n,k} - S_{n,k} S_{n,t} \right) P_{C_{n,t}} P_{C_n}^{\perp} A P_{C_n} x \\ &= \sum_{t \in C_1 \setminus \{0\}} \sum_{k \in C_N} P_{C_N \otimes C_n} \left(S_{n,k} - S_{n,k+t} \right) P_{C_{n,t}} A P_{C_n} x.\end{aligned}$$

Given $t \in C_1 \setminus \{0\}$, the only terms remaining in

$$\sum_{k \in C_N} P_{C_N \otimes C_n} \left(S_{n,k} - S_{n,k+t} \right) P_{C_{n,t}}$$

after cancellations are

$$\sum_{k \in C_N, t-k \notin C_N} -P_{C_N \otimes C_n} S_{n,k+t} P_{C_{n,t}}.$$

We can also restrict the sum to $k \in C_N$ such that there exists $t \in C_1$ with $t - k \notin C_N$ and denote this set inclusion via $k \in \partial C_N$. Upon swapping the order of summations again, we arrive at

$$(A_{n,N}^{per} - zI)x^N - y^N = - \sum_{k \in \partial C_N} \sum_{\substack{t \in C_1 \setminus \{0\} \\ t-k \notin C_N}} S_{n,k+t} P_{C_{n,t}} A P_{C_n} x.$$

Given $k \in \partial C_N$, the vector

$$\sum_{\substack{t \in C_1 \setminus \{0\} \\ t-k \notin C_N}} S_{n,k+t} P_{C_{n,t}} A P_{C_n} x$$

is supported in $C_{n,-k}$ and has norm at most $3^d \|A\|$. Since these vectors have disjoint support over different k, it follows that

$$\|(A_{n,N}^{per} - zI)x^N - y^N\| \leq 3^d \|A\| \, |\partial C_N|^{\frac{1}{2}} = \mathcal{O}(N^{\frac{d-1}{2}}). \tag{10.2.7}$$

It follows that

$$\|(A_{n,N}^{per} - zI)x^N\| \leq \|y^N\| + \mathcal{O}(N^{\frac{d-1}{2}}) \leq |C_N|^{\frac{1}{2}} \left(\sigma_1(A_n^{per} - zI) + \delta \right) + \mathcal{O}(N^{\frac{d-1}{2}}),$$

since $\|y^N\| = |C_N|^{\frac{1}{2}} \|y\|$. The idea behind this part of the proof is shown in Figure 10.2 for the case of $d = 2$.

Now we use the pseudoergodicity property of A. Again by construction, all entries of the periodic extension $A_{n,N}^{per}$ come from the set M_k of the corresponding diagonal with respect to which A is pseudoergodic. Hence by pseudoergodicity of A, for each desired accuracy $\epsilon > 0$ there is a desired $(2(2Nn + N + n) + 1)^d \times (2(2Nn + N + n) + 1)^d$ submatrix of A which is ϵ close to $A_{n,N}^{per}$. Hence we can shift the support of x^N and let $w^N \in l^2(\mathbb{Z})$ equal x^N on the corresponding $(2(2Nn + N + n) + 1)^d$ entries and zero otherwise. Choosing ϵ sufficiently small we have $\|w^N\| = \|x^N\| = |C_N|^{\frac{1}{2}} = (2N + 1)^{\frac{d}{2}}$ and

$$\|(A - zI)w^N\| \leq (2N + 1)^{\frac{d}{2}} (\sigma_1(A_n^{per} - zI) + \delta) + \delta + \mathcal{O}(N^{\frac{d-1}{2}}).$$

It follows that

$$\sigma_1(A - zI) \leq \sigma_1(A_n^{per} - zI) + 2\delta + \mathcal{O}(N^{-\frac{1}{2}}). \tag{10.2.8}$$

Letting $N \to \infty$ and then $\delta \downarrow 0$ gives $\sigma_1(A_n^{per} - zI) \geq \sigma_1(A - zI)$. □

$Ax =$

	$A_n^{(-1,1)}x$	$A_n^{(0,1)}x$	$A_n^{(1,1)}x$	
	$A_n^{(-1,0)}x$	$A_n^o x$	$A_n^{(1,0)}x$	
	$A_n^{(-1,-1)}x$	$A_n^{(0,-1)}x$	$A_n^{(1,-1)}x$	

$x^N =$

x	x	x	x	x
x	x	x	x	x
x	x	x	x	x
x	x	x	x	x
x	x	x	x	x

$A_{n,N}^{per}x^N =$

without $(0,-1),(1,0),(1,-1)$	without $(0,-1)$	without $(0,-1)$	without $(0,-1)$	without $(0,-1),(-1,0),(-1,-1)$
without $(1,0)$	$A_n^{per}x$	$A_n^{per}x$	$A_n^{per}x$	without $(-1,0)$
without $(1,0)$	$A_n^{per}x$	$A_n^{per}x$	$A_n^{per}x$	without $(-1,0)$
without $(1,0)$	$A_n^{per}x$	$A_n^{per}x$	$A_n^{per}x$	without $(-1,0)$
without $(0,1),(1,0),(1,1)$	without $(0,1)$	without $(0,1)$	without $(0,1)$	without $(0,1),(-1,0),(-1,1)$

$A_{n,N}^{per}x^N - y^N = -$

with $(0,-1),(1,0),(1,-1)$	with $(0,-1)$	with $(0,-1)$	with $(0,-1)$	with $(0,-1),(-1,0),(-1,-1)$
with $(1,0)$	0	0	0	with $(-1,0)$
with $(1,0)$	0	0	0	with $(-1,0)$
with $(1,0)$	0	0	0	with $(-1,0)$
with $(0,1),(1,0),(1,1)$	with $(0,1)$	with $(0,1)$	with $(0,1)$	with $(0,1),(-1,0),(-1,1)$

Figure 10.2: Visualisation of how periodisation works for $d = 2$ and $z = 0$. We have $A_n^{per} = A_n^o + A_n^{(-1,1)} + A_n^{(0,1)} + A_n^{(1,1)} + A_n^{(1,0)} + A_n^{(1,-1)} + A_n^{(0,-1)} + A_n^{(-1,-1)} + A_n^{(-1,0)}$. Without $(-1,1)$ refers to this sum without $A_n^{(-1,1)}$ and with $(-1,1)$ refers to $A_n^{(-1,1)}$ etc. We clearly see that $(A_{n,N}^{per} - zI)x^N - y^N$ is supported on ∂C_N with at most 3^d terms (in fact 3 in this case) in each 'box'.

10.2.3 Proof of Theorem 10.1.2

Using these results, we can now prove Theorem 10.1.2.

Proof of Theorem 10.1.2. The inclusion $\mathrm{Sp}_\epsilon(A_n^{per}) \subset \mathrm{Sp}_\epsilon(A)$ follows from Proposition 10.2.2 and the characterisation in (10.2.1). For the convergence $\lim_{n\to\infty} \mathrm{Sp}_\epsilon(A_n^{per}) = \mathrm{Sp}_\epsilon(A)$, note that A is bounded, so there exists a compact set K such that $\mathrm{Sp}_\epsilon(A) \subset K$. By Proposition 10.2.2 we only need to prove convergence of the sets $\mathrm{Sp}_\epsilon(A_n^{per})$ to $\mathrm{Sp}_\epsilon(A)$ restricted to K which without loss of generality we assume to be a closed ball around the origin. For any bounded operators S, T we have

$$|\sigma_1(T) - \sigma_1(S)| \leq \|S - T\|$$

and it follows that for $n \geq m$, $\psi_{A_n^{per}}(z)$ is Lipschitz over $z \in K$ with Lipschitz constant independent of n. Proposition 10.2.2 and Lemma 10.2.1 together give that

$$\psi_A(z) \leq \liminf_{n\to\infty} \psi_{A_n^{per}}(z) \leq \limsup_{n\to\infty} \psi_{A_n^{per}}(z) \leq \psi_A(z).$$

It follows that $\psi_{A_n^{per}}(z)$ converges pointwise to $\psi_A(z)$ and hence uniform Lipschitz continuity upgrades this to uniform convergence over K. Now let $0 < \delta < \epsilon$ then the above shows that for large n we have

$$\mathrm{Sp}_{\epsilon-\delta}(A) \subset \mathrm{Sp}_\epsilon(A_n^{per}) \subset \mathrm{Sp}_\epsilon(A).$$

Finally, $\mathrm{Sp}_\eta(T)$ is continuous (with respect to the Hausdorff metric) in η for any fixed $T \in \mathcal{B}(l^2(\mathbb{Z}^d))$. Convergence now follows since $0 < \delta < \epsilon$ was arbitrary.

To see convergence of `PseudoSpecPer`, note that we have $\Gamma_n(A) \subset \mathrm{Sp}_\epsilon(A_n^{per})$ by construction. Choose a compact subset $K \subset \mathbb{C}$ with $\psi_{A_n^{per}}(z) > 2\epsilon$ for all $z \in \mathbb{C}\backslash K$ and for all n. By the uniform convergence and the Arzelá-Ascoli theorem we can choose $\delta_n \downarrow 0$ such that for all n,

$$|\psi_{A_n^{per}}(z) - \psi_{A_n^{per}}(w)| < \delta_n \text{ for all } z, w \in K \text{ with } |z - w| < 1/n.$$

Let n be large so that $K \subset [-n,n]+i[-n,n]$ and such that $\delta_n < \epsilon$. If this holds and $z_1 \in \mathrm{Sp}_{\epsilon-\delta_n}(A_n^{per})$ then there exists some $z_2 \in \frac{1}{n}(\mathbb{Z}+i\mathbb{Z}) \cap B_n(0)$ with $|z_1 - z_2| < 1/n$ and hence $|\psi_{A_n^{per}}(z_1) - \psi_{A_n^{per}}(z_2)| < \delta_n$. It follows that $z_2 \in \Gamma_n(A)$ and hence

$$\mathrm{Sp}_{\epsilon-\delta_n}(A_n^{per}) \subset \Gamma_n(A) + B_{1/n}(0) \subset \mathrm{Sp}_{\epsilon}(A_n^{per}) + B_{1/n}(0).$$

Let $\eta > 0$ with $\eta < \epsilon$ and choose n large such that $\epsilon - \delta_n > \eta$ then

$$\mathrm{Sp}_{\eta}(A_n^{per}) \subset \Gamma_n(A) + B_{1/n}(0) \subset \mathrm{Sp}_{\epsilon}(A_n^{per}) + B_{1/n}(0).$$

The right-hand side converges to $\mathrm{Sp}_{\epsilon}(A)$ and the left-hand side converges to $\mathrm{Sp}_{\eta}(A)$. Since $\eta < \epsilon$ was arbitrary and Sp_{η} is continuous in η, the desired convergence now follows. □

We have now shown why periodic boundary conditions are a natural choice for pseudoergodic operators. Although we may not have convergence of spectra (for example the 1D NSA Anderson model in §3.6.2), we do obtain convergence for pseudospectra.

10.2.4 Extension to vector-valued sequences

Here we briefly remark on the extension of the above arguments to vector-valued sequences. Consider the following generalisation of the standard lattice $\mathbb{Z}^d$. For some $d \in \mathbb{N}$ and finite set $\mathcal{S}$, set

$$\mathbb{X} = \mathbb{Z}^d \times \mathcal{S}.$$

We view this as the lattice $\mathbb{Z}^d$ with $|\mathcal{S}|$ sites attached to each point. In this case $l^2(\mathbb{X},\mathbb{C}) \sim l^2(\mathbb{Z}^d, \mathbb{C}^{|\mathcal{S}|})$. Enumerating a basis of $l^2(\mathbb{X})$ (each basis vector corresponding to a site) as $\{e_{i,a} : i \in \mathbb{Z}^d, a \in \mathcal{S}\}$ allows us, for $A \in \mathcal{B}(l^2(\mathbb{X}))$, to form matrix elements

$$A_{(i,a),(j,b)} = \langle Ae_{j,b}, e_{i,a}\rangle.$$

In complete generalisation of Definition 10.1.1 above (where $|\mathcal{S}| = 1$), we say that a collection $\underline{M} = \{M_{k,a,b} \subset \mathbb{C} : k \in \mathbb{Z}^d, \text{ and } a,b \in \mathcal{S}\}$ is permissible if there exists $m \in \mathbb{N}$ such that $M_{k,a,b} = \{0\}$ if $\|k\|_\infty > m$. Given permissible $\underline{M}$, we say A is (translationally) pseudoergodic with respect to $\underline{M}$ if $A_{(i,a),(j,b)} \in M_{i-j,a,b}$ for all $a,b \in \mathcal{S}$ and the following property holds. Given any $\epsilon > 0$, finite subsets $S_{k,a,b} \subset \mathbb{Z}^d \times \mathcal{S}^2$ and functions $F_{k,a,b} : S_{k,a,b} \to M_{k,a,b}$ (for $k \in \mathbb{Z}^d$, and $a,b \in \mathcal{S}$), there exists a translation T acting on $\mathbb{Z}^d$ such that

$$\sup_{(i,a,b)\in S_{k,a,b}} \left|A_{(T(i),a),(T(i)-k,b)} - F_{k,a,b}(i)\right| < \epsilon, \quad k \in \mathbb{Z}^d, \text{ and } a,b \in \mathcal{S}.$$

Denote the collection of such A by $\mathcal{A}(\underline{M})$ and the union of $\mathcal{A}(\underline{M})$ over permissible $\underline{M}$ by $\Omega^{\mathbb{X}}$. Note that $A_{(i,a),(j,b)} \in M_{i-j,a,b}$ implies that

$$A_{(i,a),(j,b)} = 0, \quad \text{if } \|i-j\|_\infty > m, \tag{10.2.9}$$

the generalised notion of bandedness.

To treat these operators, only a slight adjustment to the definitions in §10.2.2 are needed. We now define the index sets $C_n^{\mathcal{S}} := \{-n, n+1, ..., n\}^d \times \mathcal{S}$, $C_{n,k}^{\mathcal{S}} := \big(C_n + (2n+1)k\big) \times \mathcal{S}$ and

$$C_N^{\mathcal{S}} \otimes C_n^{\mathcal{S}} := \bigcup_{k\in C_N} C_{n,k}^{\mathcal{S}} = C_{2Nn+N+n}^{\mathcal{S}}.$$

For $W \subset \mathbb{X}$, we define the orthogonal projections $P_W, P_W^\perp : l^2(\mathbb{X}) \to l^2(\mathbb{X})$ via

$$(P_W x)_{(j,a)} = \begin{cases} x_{(j,a)}, & \text{if } (j,a) \in W, \\ 0, & \text{otherwise} \end{cases}$$

and $P_W^\perp = P_{\mathbb{X}\setminus W}$ as before. The shift operator $S_{n,k}^{\mathcal{S}} : l^2(\mathbb{X}) \to l^2(\mathbb{X})$ now acts via

$$(S_{n,k}^{\mathcal{S}} x)_{(j,a)} = x_{(j-(2n+1)k,a)}, \quad j \in \mathbb{Z}^d, a \in \mathcal{S}$$

which translates between the $C_{n,k}^{\mathcal{S}}$'s. The definitions of A_n^o, A_n^{per} and $A_{n,N}^{per}$ are as before with the relevant superscripts $\mathcal{S}$ on the projections and shifts:

$$A_n^o x = P_{C_n^{\mathcal{S}}} A P_{C_n^{\mathcal{S}}} x, \quad A_n^{per} x = \sum_{k \in C_1} P_{C_n^{\mathcal{S}}} S_{n,-k}^{\mathcal{S}} A P_{C_n^{\mathcal{S}}} x, \quad A_{n,N}^{per} x = P_{C_N^{\mathcal{S}} \otimes C_n^{\mathcal{S}}} \sum_{k \in C_N} S_{n,k}^{\mathcal{S}} A P_{C_n^{\mathcal{S}}} S_{n,-k}^{\mathcal{S}} x.$$

The proof of the generalisation of Proposition 10.2.2 to $|\mathcal{S}| > 1$ now follows through almost verbatim with the addition of the relevant superscripts $\mathcal{S}$. For instance, the same manipulations lead to

$$(A_{n,N}^{per} - zI)x^N - y^N = -\sum_{k \in \partial C_N} \sum_{\substack{t \in C_1 \setminus \{0\} \\ t-k \notin C_N}} S_{n,k+t}^{\mathcal{S}} P_{C_{n,t}^{\mathcal{S}}} A P_{C_n^{\mathcal{S}}} x,$$

from which the rest of the argument easily follows. Lemma 10.2.1 also holds and together these prove the generalisation of Theorem 10.1.2 to $\Omega^{\mathbb{X}}$ using the same arguments as in §10.2.3.

10.3 The General l^p Case

In this section, we will prove that the results of §10.2 can be generalised to the case of viewing the pseudoergodic operator as acting on $l^p(\mathbb{X})$, where $\mathbb{X}$ is the generalisation of $\mathbb{Z}^d$ discussed in §10.2.4. Recall that due to the definition of pseudoergodicity, the operators are banded in the generalised sense with uniformly bounded matrix values - hence their matrices can be viewed as operators acting on $l^p(\mathbb{X})$ for any $p \in [1, \infty]$. For general Banach spaces, one needs to be careful of the definition of pseudospectrum, since the resolvent norm can be constant on open subsets of the resolvent [Sha08]. This does not occur for Banach spaces which have the strong maximum property (see [Sha08, Glo76] for a definition and the following theorem - the fact that $l^p(\mathbb{X})$ satisfies the required property is mentioned in [Sha08] with results from [Cla36, Glo75, Lešš88]) and the following theorem demonstrates that we do not have to worry about this in the cases considered in this chapter.

Theorem 10.3.1 ([Sha08, Glo76])**.** *Suppose that X is a Banach space such that at least one of X, X^* is complex uniformly convex or such that X is finite-dimensional. Then X has the strong maximum property. In particular, this holds for $l^p(\mathbb{X})$.*

This means that we shall take (1.4.1) as our definition of $\mathrm{Sp}_\epsilon^p(A)$ with the l^2 operator norm replaced by its l^p counterpart. Some authors differ in requiring $\{z \in \mathbb{C} : \|R(z,A)\|^{-1} < \epsilon\}$ (note the strict inequality) or the closure of such a set but in light of Theorem 10.3.1, we see that the closure definition and ours agree in this context. In proving results, we will find the following theorem useful. If B is a bounded operator on the Banach space X, then B^* is the adjoint operator defined on X^* (with the convention of taking anti-linear functionals following Kato [Kat95]). In our case, this means that if $1 \leq p < \infty$ then A^* is the matrix operator defined by the usual complex conjugate defined on $l^q(\mathbb{X})$ with $1/p + 1/q = 1$.

Theorem 10.3.2 (See for example [TE05]). *Let X be a Banach space with the strong maximum property and $A \in \mathcal{B}(X)$ then $\mathrm{Sp}^{\mathrm{X}}_{\epsilon}(A)$ (the $\epsilon-$pseudospectrum defined using the operator norm on $A \in \mathcal{B}(X)$) is the set of $z \in \mathbb{C}$ satisfying any of the following four equivalent definitions*

I. $\|R(z,A)\|^{-1} \leq \epsilon$,

II. $z \in \mathrm{Sp}(A+E)$ *for some* $E \in \mathcal{B}(X)$ *with* $\|E\| \leq \epsilon$,

III. $z \in \mathrm{Sp}(A)$ *or there exists* $x_n \in X$ *of norm* 1 *with* $\limsup_{n\to\infty} \|(A-zI)x_n\| \leq \epsilon$,

IV. There exists $x_n \in X$ *of norm* 1 *with* $\limsup_{n\to\infty} \|(A-zI)x_n\| \leq \epsilon$ *or there exists* $y_n \in X^*$ *of norm* 1 *with* $\limsup_{n\to\infty} \|(A^*-\bar{z}I)y_n\| \leq \epsilon$.

Following [Sei12], we define the injection and surjection modulus respectively by

$$j_X(A) = \sup\{\tau \geq 0 : \|Ax\| \geq \tau\|x\| \text{ for all } x \in X\} = \inf\{\|Ax\| : \|x\| = 1\}$$
$$q_X(A) = \sup\{\tau \geq 0 : A(B_X) \supset \tau B_X\}.$$

We then have $\|A^{-1}\|^{-1} = \min\{j(A), q(A)\}$, $j_{X^*}(A^*) = q_X(A)$ and $q_{X^*}(A^*) = j_X(A)$. Furthermore, if A is invertible then $j_X(A) = q_X(A)$. We define the functions

$$\psi^p_A(z) := \min\{j_{l^p}(A-zI), q_{l^p}(A-zI)\},$$
$$\psi^p_{A^{per}_n}(z) := \min\{j_{l^p}(A^{per}_n - zI), q_{l^p}(A^{per}_n - zI)\},$$

and note that we can characterise the pseudospectrum as $\mathrm{Sp}^p_\epsilon(A) = \{z \in \mathbb{C} : \psi^p_A(z) \leq \epsilon\}$. Assume for the remainder of this section that $\underline{M}$ is permissible. Note that we have not yet shown that $\mathrm{Sp}^p_\epsilon(A)$ is constant over all $A \in \mathcal{A}(\underline{M})$, however this follows from Theorem 4.7 (and Corollary 4.9) of [BLLS17]. Upon letting $\epsilon \downarrow 0$, this also proves that the spectrum is constant on $\mathcal{A}(\underline{M})$. Recalling the generalised bandwidth m of $A \in \mathcal{A}(\underline{M})$ in (10.2.9), we have the following Proposition which extends Proposition 10.2.2 to $p \neq 2$.

Proposition 10.3.3. *Let $p \in [1,\infty]$, $d \in \mathbb{N}$ and $A \in \mathcal{B}(l^p(\mathbb{X}))$ be pseudoergodic with respect to $\underline{M}$. For $n \geq m$ and all $z \in \mathbb{C}$ we have $\psi^p_{A^{per}_n}(z) \geq \psi^p_A(z)$.*

Proof. Assume that $A \in \mathcal{A}(\underline{M})$ and $n \geq m$. If $z \in \mathrm{Sp}^p(A)$, then $\psi^p_A(z) = 0$ and we have nothing to prove, so assume that $z \notin \mathrm{Sp}^p(A)$. This implies that

$$\psi^p_A(z) = j_{l^p}(A-zI) = q_{l^p}(A-zI).$$

Since A^{per}_n acts on a finite-dimensional vector space, we also have that

$$\psi^p_{A^{per}_n}(z) = j_{l^p}(A^{per}_n - zI) = q_{l^p}(A^{per}_n - zI).$$

Hence we must prove that

$$j_{l^p}(A-zI) \leq j_{l^p}(A^{per}_n - zI). \tag{10.3.1}$$

We begin with the case that $p < \infty$. To see that (10.3.1) holds in this case, we argue as in §10.2.2 (with the added notational complexity of §10.2.4 if $|\mathcal{S}| > 1$). The only real changes are that in (10.2.7) which now becomes

$$\|(A^{per}_{n,N} - zI)x^N - y^N\| \leq 3^d\|A\| \left|\partial C^{\mathcal{S}}_N\right|^{\frac{1}{p}} = \mathcal{O}(N^{\frac{d-1}{p}}),$$

and x^N, y^N which now have norms $\left|C_N^{\mathcal{S}}\right|^{\frac{1}{p}}$ and $\left|C_N^{\mathcal{S}}\right|^{\frac{1}{p}} \|y\|$ respectively. We now have

$$\left|C_N^{\mathcal{S}}\right|^{\frac{1}{p}} = (2N+1)^{\frac{d}{p}} \left|\mathcal{S}\right|^{\frac{1}{p}}.$$

The same arguments then yields

$$j_{l^p}(A_n^{per} - zI) \leq j_{l^p}(A - zI) + 2\delta + \mathcal{O}(N^{\frac{1}{p}})$$

in place of (10.2.8) and (10.3.1) then follows using exactly the same arguments.

Next we show that (10.3.1) holds for $p = \infty$. For this we consider the matrix adjoint $B = (A - zI)^*$ as an operator on $l^1(\mathbb{X})$. Note that this is not the same as the operator adjoint (which acts on a much larger space). Similarly, we consider the matrix adjoint $B_n = (A_n^{per} - zI)^*$ as an operator on $l^1(C_n^{\mathcal{S}})$. B is pseudoergodic and hence

$$j_{l^1}(B) \leq j_{l^1}(B_n).$$

(Note that the periodisation commutes with taking the matrix adjoint.) It follows that $q_{l^\infty}(B^*) = j_{l^1}(B)$ and $q_{l^\infty}(B_n^*) = j_{l^1}(B_n)$. Hence

$$j_{l^\infty}(A - zI) = q_{l^\infty}(A - zI) = q_{l^\infty}(B^*) \leq q_{l^\infty}(B_n^*) = q_{l^\infty}(A_n^{per} - zI) = j_{l^\infty}(A_n^{per} - zI),$$

which proves (10.3.1) for $p = \infty$. □

Theorem 10.3.4. *Let $p \in [1, \infty]$ and $A \in \mathcal{B}(l^p(\mathbb{X}))$ be pseudoergodic with respect to $\underline{M}$. Then $\psi^p_{A_n^{per}}(z)$ converges uniformly to $\psi^p_A(z)$ from above on compact subsets of $\mathbb{C}$. Hence $\lim_{n\to\infty} \mathrm{Sp}^p_\epsilon(A_n^{per}) = \mathrm{Sp}^p_\epsilon(A)$ in the Hausdorff metric and $\mathrm{Sp}^p_\epsilon(A_n^{per}) \subset \mathrm{Sp}^p_\epsilon(A)$, i.e. Theorem 10.1.3 and its extension to $l^p(\mathbb{X})$ hold.*

Proof. Suppose that we can prove pointwise convergence. Uniform convergence follows by a similar argument as Theorem 10.1.2 where we have uniform Lipschitz continuity from the definition of injection modulus (and hence the surjection modulus by considering the operator dual if $p < \infty$ or the matrix adjoint if $p = \infty$). By Proposition 10.3.3, convergence is from above and hence $\mathrm{Sp}^p_\epsilon(A_n^{per}) \subset \mathrm{Sp}^p_\epsilon(A)$. Using Theorem 10.3.1 and a straightforward compactness argument, it is easy to see that $\mathrm{Sp}^p_\epsilon(A)$ is continuous in ϵ. The uniform convergence of $\psi^p_{A_n^{per}}(z)$ now implies $\lim_{n\to\infty} \mathrm{Sp}^p_\epsilon(A_n^{per}) = \mathrm{Sp}^p_\epsilon(A)$ as in the proof of Theorem 10.1.2.

Hence we are left with proving pointwise convergence. By Proposition 10.3.3, it is enough to show that

$$\limsup_{n\to\infty} \psi^p_{A_n^{per}}(z) \leq \psi^p_A(z). \tag{10.3.2}$$

The truncation argument in the proof of Lemma 10.2.1 works for $p \in (1, \infty)$ (p and its dual must be finite) and hence we only have to consider the $p \in \{1, \infty\}$ cases. Note that the truncation argument shows that

$$\limsup_{n\to\infty} j_{l^1}(A_n^{per} - zI) \leq j_{l^1}(A - zI).$$

Applying this to the matrix adjoints and using the same argument as in the proof of Proposition 10.3.3 shows that

$$\limsup_{n\to\infty} q_{l^\infty}(A_n^{per} - zI) \leq q_{l^\infty}(A - zI).$$

Suppose that we can also show that

$$\limsup_{n\to\infty} j_{l^\infty}(A_n^{per} - zI) \leq j_{l^\infty}(A - zI). \tag{10.3.3}$$

Then the duality $(l^1)^* = l^\infty$ yields (again by matrix adjoints) that

$$\limsup_{n\to\infty} q_{l^1}(A_n^{per} - zI) \leq q_{l^1}(A - zI),$$

which finishes the proof of (10.3.2) and hence of the theorem.

We are thus left with proving (10.3.3) so assume for the remainder of the proof that $p = \infty$. Given $\delta > 0$, there exists $x \in l^\infty(\mathbb{X})$ of norm 1 such that $\|(A - zI)x\| \leq j_{l^\infty}(A - zI) + \delta$. Fix any $N \in \mathbb{N}$ and define

$$(x_N)_{(i,a)} = x_{(i,a)} \max\left\{0, 1 - \frac{\|i\|_\infty}{N}\right\}, \quad i \in \mathbb{Z}^d, a \in \mathcal{S}.$$

It is clear that x_N has finite support and $P_{C_n^S} x_N = x_N$ for large n. Now we use the fact that if $A_{(i,a),(j,b)} \neq 0$ then $\|i - j\|_\infty \leq m$ for some $m \in \mathbb{N}$. Consider the entry $((A_n^{per} - zI)x_N)_{(i,a)}$ where we assume that n is large so that this is equal to $((A - zI)x_N)_{(i,a)}$ for all (i, a). Since the operator is banded in the generalised sense, we must have

$$\left|((A - zI)x_N - \lambda_i(N)(A - zI)x)_{(i,a)}\right| \leq \frac{C(A, z)}{N}, \tag{10.3.4}$$

for some constant $C(A, z)$ independent of N and (i, a) where $\lambda_i(N)$ is the local factor

$$\lambda_i(N) = \max\left\{0, 1 - \frac{\|i\|_\infty}{N}\right\},$$

which converges to 1 as $N \to \infty$ for any i. Let y_N be defined by

$$(y_N)_{(i,a)} = \lambda_i(N)\left((A - zI)x\right)_{(i,a)}$$

then we have

$$\limsup_{n\to\infty} j_{l^\infty}(A_n^{per} - zI) \leq \frac{\|(A - zI)x_N\|}{\|x_N\|} \leq \frac{\frac{C(A,z)}{N} + \|y_N\|}{\|x_N\|}.$$

But $\lim_{N\to\infty} \|x_N\| = \|x\| = 1$ and

$$\lim_{N\to\infty} \|y_N\| = \|(A - zI)x\| \leq j_{l^\infty}(A - zI) + \delta.$$

Hence

$$\limsup_{n\to\infty} j_{l^\infty}(A_n^{per} - zI) \leq j_{l^\infty}(A - zI) + \delta.$$

Since $\delta > 0$ was arbitrary, this proves (10.3.3) and hence the theorem. □

Remark 10.3.5. *Bandedness was crucial in the above proof to obtain (10.3.4). One can in fact study $\|B\|$ and $\|B^{-1}\|$ for much more general operators B on l^∞ by looking at $\|B_0\|$ and $\|B_0^{-1}\|$, where B_0 is the restriction of B to the space of sequences convergent to zero (see [HLS16] Lemma 3.8), hence allowing similar arguments for B_0 as in the case of $p < \infty$. See also [Sei14] for further discussion of these so-called $\mathcal{P}$-techniques.*

Concluding Remarks

A detailed summary of the contributions of this thesis can be found in §1.2 of Chapter 1. Here, we provide a lookup table of the computational spectral problems discussed in this thesis and the corresponding theorems:

Problem	**Theorems**	**Section**	**Pages**
Spectra (and pseudospectra) of operators (including unbounded) on graphs with error control.	3.1.4	§3.1.1	44
Decision problem if spectra (and pseudospectra) of operators (including unbounded) on graphs intersect compact set.	3.1.6	§3.1.1	45
Spectra (and pseudospectra) of PDOs with error control.	3.1.10, 3.1.12	§3.1.2	47, 48
Approximate states.	3.4.1	§3.4	68
Spectral measures (projection-valued and scalar-valued), measure decompositions and projections.	4.3.1, 4.3.3	§4.3	96, 97
Functional calculus and Radon–Nikodym derivative of absolutely continuous part of measure.	4.4.1, 4.4.2	§4.4	104, 104
Functional calculus and Radon–Nikodym derivative of absolutely continuous part of measure with error control under local regularity assumptions.	4.5.3, 4.5.7, 4.5.9	§4.5.2	108, 111, 114
Spectral type (absolutely continuous, singular continuous, pure point).	5.1.1	§5.1	127
Discrete spectra and eigenvalue multiplicities.	6.1.1, 6.1.3, 6.1.4	§6.1.1	144, 144, 145
Spectral gap problem and generalisation.	6.1.5, 6.1.7	§6.1.2	145, 146
Spectral radii and essential spectral radii.	7.3.1, 7.3.3	§7.3.1	162, 162
Polynomial operator norms and capacity.	7.3.4	§7.3.1	163
Gaps in essential spectra and finite section failure.	7.3.8	§7.3.2	164
Lebesgue measure of spectra (and pseudospectra) and decision problem of when this is zero.	8.1.1, 8.1.3, 8.1.5	§8.1.1	181, 181, 182
Fractal dimensions (Hausdorff and box-counting) of spectra.	8.1.7, 8.1.10	§8.1.2	183, 184
Convergence theorems of IQR algorithm.	9.2.9, 9.2.13, 9.2.15	§9.2	215, 218, 221
Implementation theorems of IQR algorithm.	9.3.3, 9.3.8	§9.3	225, 228
IQR: extremal parts of spectrum, full spectrum, dominant invariant subspaces.	9.4.4, 9.4.6, 9.4.6	§9.4	231, 232, 233
Convergence of finite section pseudospectra with periodic boundary conditions for pseudoergodic operators.	10.1.1, 10.1.3	§10.1.1	249, 250

Table 10.1: List of spectral problems solved in this thesis (see theorems for classifications).

Open problems and future work

We end with some remarks on open problems and future work. The resolvent approach in Chapter 3 can easily be extended to operator pencils. This raises the following questions:

- The above results could lead to efficient finite element computation of spectra. A resolvent-based approach could work in tandem with current finite element codes and thus be applicable to a large and active community. Note that the problem of error control for finite element methods (even in the case when algorithms converge) is well documented [Zha15]. This approach may also be possible in other set-ups such as boundary element methods or methods that have a different representation for the domain and range spaces (such as the ultraspherical spectral method).

- Resolvent techniques have also been used for finite-dimensional nonlinear eigenvalue problems. The extension of resolvent methods to infinite-dimensional nonlinear problems is currently under investigation but is likely to be very challenging. The computational foundations for such problems are completely untouched and merit investigation.

- The extension of these methods to Banach spaces may be difficult since one would need an algorithm for computing the injection modulus of an operator. Computing general l^p norms of finite matrices is NP-hard (for $p \neq 1, 2, \infty$) and hence direct approaches may be intractable. Are there tractable methods that approximate injection moduli of truncated (finite rank) operators on Banach spaces?

The resolvent method to compute measures and decompositions in Chapters 4 and 5 raises the following questions:

- We introduced a class of rational kernels that locally accelerate the convergence of approximations. Many questions remain, such as the choice of optimal poles. The convergence is faster near smoother parts of the measure. Is there a way to subtract the singular part of the measure to make it smoother? This could involve a mixture of global and local approaches.

- Examples of evolution equations solved using contour integration of the resolvent were given. Future work will explore the use of Proposition 4.2.1 and rectangular systems/resolvent contour approaches to construct stable solvers, including for unbounded operators arising as generators of strongly continuous semigroups. More generally, the techniques of this thesis may be extendable to uncovering the foundations of solving linear and nonlinear PDEs, particularly on unbounded domains.

Chapter 9, particularly §9.5, raises the following questions regarding the IQR algorithm:

- What conditions are needed on a possibly non-normal operator for the IQR algorithm to pick up the extreme points of the essential spectrum? We conjecture that there may be a large class of operators for which this holds.[3] For example, if the set of extremal points of the essential spectrum has size one. Is the convergence rate to non-isolated points of the spectrum algebraic?

- Is there a way of combining the finite section method and IQR algorithm to avoid spectral pollution? For example, for which operators that do not have a trivial QR decomposition is there a way of choosing $n = n(m)$ such that $\mathrm{Sp}(P_m Q_{n(m)}^* A Q_{n(m)} P_m)$ converges to the spectrum as $m \to \infty$?

[3] This is false in general as is easily seen by considering the shift operator.

- More generally, the original paper of Deift, Li and Tomei [DLT85] shows that the IQR algorithm samples Toda flows at integer times. This immediately raises the question of whether there is a better sampling strategy? Moreover, there are many other differential flows of this kind associated with forward and inverse eigenvalue problems. Future work will look at the approaches of this thesis in the context of such flows in infinite dimensions. For example, are there any such flows that are useful for unbounded operators?

Beyond the topics of this thesis, there are natural links with optimisation, neural networks, PDEs and computer-assisted proofs. Such areas present promising avenues for future work.

Bibliography

[AA80] Serge Aubry and Gilles André. Analyticity breaking and Anderson localization in incommensurate lattices. *Ann. Israel Phys. Soc*, 3(133):18, 1980.

[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.

[Adv06] LLC Advanpix. Multiprecision computing toolbox for MATLAB. *YoNohama, Japan*, 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.

[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.

[AHN16] Ariel Amir, Naomichi Hatano, and David R. Nelson. Non-Hermitian localization in biological networks. *Physical Review E*, 93(4):042310, 2016.

[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.

[AM93] Michael Aizenman and Stanislav Molchanov. Localization at large disorder and at extreme energies: an elementary derivation. *Communications in Mathematical Physics*, 157(2):245–278, 1993.

[And58] Philip W. Anderson. Absence of diffusion in certain random lattices. *Physical Review*, 109(5):1492, 1958.

[And61] Philip W. Anderson. Localized magnetic states in metals. *Physical Review*, 124(1):41, 1961.

[Arn51] Walter E. Arnoldi. The principle of minimized iteration in the solution of the matrix eigenvalue problem. *Quart. Appl. Math.*, 9:17–29, 1951.

[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.

[AW15] Michael Aizenman and Simone Warzel. *Random operators*, volume 168 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2015.

[BACH+19] 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:1508.03280v4*, 2019.

[BAMR20a] Jonathan Ben-Artzi, Marco Marletta, and Frank Rösler. Computing scattering resonances. *arXiv:2006.03368*, 2020.

[BAMR20b] Jonathan Ben-Artzi, Marco Marletta, and Frank Rösler. Computing the sound of the sea in a seashell. *arXiv:2009.02956*, 2020.

[Bar05] Thabit Barakat. The asymptotic iteration method for the eigenenergies of the anharmonic oscillator potential $V(x) = Ax^{2\alpha} + Bx^2$. *Physics Letters A*, 344(6):411–417, 2005.

[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.

[BBM+14] Sabine Bögli, B. Malcolm Brown, Marco Marletta, Christiane Tretter, and Markus Wagenhofer. Guaranteed resonance enclosures and exclosures for atoms and molecules. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 470(2171):20140488, 2014.

[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.

[Ber01] Michael Berry. Fractal modes of unstable lasers with polygonal and circular mirrors. *Optics communications*, 200(1-6):321–330, 2001.

[Ber03] Michael Berry. Mode degeneracies and the Petermann excess-noise factor for unstable lasers. *Journal of Modern Optics*, 50(1):63–81, 2003.

[Ber04] Michael Berry. Physics of nonhermitian degeneracies. *Czechoslovak journal of physics*, 54(10):1039–1047, 2004.

[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.

[BG05] Albrecht Böttcher and Sergei M. Grudsky. *Spectral properties of banded Toeplitz matrices*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.

[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.

[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.

[BJZ+08] Juliette Billy, Vincent Josse, Zhanchun Zuo, Alain Bernard, Ben Hambrecht, Pierre Lugan, David Clément, Laurent Sanchez-Palencia, Philippe Bouyer, and Alain Aspect. Direct observation of Anderson localization of matter waves in a controlled disorder. *Nature*, 453(7197):891–894, 2008.

[BK99] Michael Berry and Jonathan P. Keating. The Riemann zeros and eigenvalue asymptotics. *SIAM Review*, 41(2):236–266, 1999.

[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.

[BLLS17] Siegfried Beckus, Daniel Lenz, Marko Lindner, and Christian Seifert. On the spectrum of operator families on discrete groups over minimal dynamical systems. *Math. Z.*, 287(3-4):993–1007, 2017.

[BM92] Christine Bernardi and Yvon Maday. *Approximations spectrales de problèmes aux limites elliptiques*, volume 10 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Paris, 1992.

[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.

[BS10] Albrecht Böttcher and Ilya M. Spitkovsky. A gentle guide to the basics of two projections theory. *Linear Algebra Appl.*, 432(6):1412–1459, 2010.

[BSB97] P. W. Brouwer, P. G. Silvestrov, and C. W. J. Beenakker. Theory of directed localization in one dimension. *Physical Review B*, 56(8):R4333, 1997.

[BSVS01] Michael Berry, Cornelis Storm, and Wim Van Saarloos. Theory of unstable laser modes: edge waves and fractality. *Optics communications*, 197(4-6):393–402, 2001.

[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.

[BZ98] Édouard Brézin and A. Zee. Non-Hermitian delocalization: Multiple scattering and bounds. *Nuclear Physics B*, 509(3):599–614, 1998.

[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.

[CH19a] Matthew J. Colbrook and Anders C. Hansen. The foundations of spectral computations via the solvability complexity index hierarchy: Part I. *arXiv:1908.09592*, 2019.

[CH19b] 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.

[CHT20] Matthew J. Colbrook, Andrew Horning, and Alex Townsend. Computing spectral measures of self-adjoint operators. *arXiv:2006.01766*, 2020.

[CL55] Earl A. Coddington and Norman Levinson. *Theory of ordinary differential equations*. McGraw-Hill Book Company, Inc., New York-Toronto-London, 1955.

[CL90] René Carmona and Jean Lacroix. *Spectral theory of random Schrödinger operators*. Probability and its Applications. Birkhäuser Boston, Inc., Boston, MA, 1990.

[Cla36] James A. Clarkson. Uniformly convex spaces. *Trans. Amer. Math. Soc.*, 40(3):396–414, 1936.

[CM91] R. N. Chaudhuri and M. Mondal. Improved Hill determinant method: General approach to the solution of quantum anharmonic oscillators. *Physical Review A*, 43(7):3241, 1991.

[Col19a] Matthew J. Colbrook. Computing spectral measures and spectral types. *arXiv:1908.06721v2*, 2019.

[Col19b] Matthew J. Colbrook. The foundations of spectral computations via the solvability complexity index hierarchy: Part II. *arXiv:1908.09598*, 2019.

[Col19c] Matthew J. Colbrook. Pseudoergodic operators and periodic boundary conditions. *Mathematics of Computation*, 2019.

[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.

[Cup80] Jan J. M. Cuppen. A divide and conquer method for the symmetric tridiagonal eigenproblem. *Numerische Mathematik*, 36(2):177–195, 1980.

[CW13] Snorre H. Christiansen and Ragnar Winther. On variational eigenvalue approximation of semidefinite operators. *IMA J. Numer. Anal.*, 33(1):164–189, 2013.

[CWCL11] Simon N. Chandler-Wilde, Ratchanikorn Chonchaiya, and Marko Lindner. Eigenvalue problem meets Sierpinski triangle: computing the spectrum of a non–self–adjoint random operator. *Operators and Matrices*, 5(4):633–648, 2011.

[CWCL13] Simon N. Chandler-Wilde, Ratchanikorn Chonchaiya, and Marko Lindner. On the spectra and pseudospectra of a class of non-self-adjoint random matrices and operators. *Oper. Matrices*, 7(4):739–775, 2013.

[Dam09] David Damanik. The spectrum of the almost Mathieu operator. *arXiv:0908.1093*, 2009.

[Dav98] E. Brian Davies. Spectral enclosures and complex resonances for general self-adjoint operators. *LMS J. Comput. Math.*, 1:42–74, 1998.

[Dav99] E. Brian Davies. Pseudo-spectra, the harmonic oscillator and complex resonances. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.*, 455(1982):585–599, 1999.

[Dav01] E. Brian Davies. Spectral theory of pseudo-ergodic operators. *Communications in Mathematical Physics*, 216(3):687–704, 2001.

[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.

[DE15] Alexander Dubbs and Alan Edelman. Infinite random matrix theory, tridiagonal bordered Toeplitz matrices, and the moment problem. *Linear Algebra and its Applications*, 467:188–201, 2015.

[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.

[DM89] Peter Doyle and Curt McMullen. Solving the quintic by iteration. *Acta Mathematica*, 163(3-4):151–180, 1989.

[DN86] Joanne Dombrowski and Paul Nevai. Orthogonal polynomials, measures and recurrence relations. *SIAM Journal on Mathematical Analysis*, 17(3):752–759, 1986.

[DNS99] Karin A. Dahmen, David R. Nelson, and Nadav M. Shnerb. Population dynamics and non-Hermitian localization. In *Statistical mechanics of biocomplexity*, pages 124–151. Springer, 1999.

[Don63] William F. Donoghue, Jr. On a problem of Nieminen. *Inst. Hautes Études Sci. Publ. Math.*, pages 31–33, 1963.

[DS00] Jack Dongarra and Francis Sullivan. Guest editors' introduction: The top 10 algorithms. *Computing in Science & Engineering*, 2(1):22–23, 2000.

[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.

[Dur70] Peter L. Duren. *Theory of H^p spaces*. Pure and Applied Mathematics, Vol. 38. Academic Press, New York-London, 1970.

[DVET+05] Alessandro Della Villa, Stefan Enoch, Gérard Tayeb, Vincenzo Pierro, Vincenzo Galdi, and Filippo Capolino. Band gap formation and multiple scattering in photonic quasicrystals with a Penrose-type lattice. *Physical Review Letters*, 94(18):183903, 2005.

[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.

[Ede88] Alan Edelman. Eigenvalues and condition numbers of random matrices. *SIAM Journal on Matrix Analysis and Applications*, 9(4):543–560, 1988.

[EDML06] S. Even-Dar Mandel and Ron Lifshitz. Electronic energy spectra and wave functions on the square Fibonacci tiling. *Philosophical Magazine*, 86(6-8):759–764, 2006.

[EDML08] S. Even-Dar Mandel and Ron Lifshitz. Electronic energy spectra of square and cubic Fibonacci quasicrystals. *Philosophical Magazine*, 88(13-15):2261–2273, 2008.

[EE87] David E. Edmunds and W. Desmond Evans. *Spectral theory and differential operators*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 1987.

[EHW+13] T. Eichelkraut, R. Heilmann, S. Weimann, S. Stützer, F. Dreisow, D. N. Christodoulides, S. Nolte, and A. Szameit. Mobility transition from ballistic to diffusive transport in non-Hermitian lattices. *Nature Communications*, 4, 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.

[Ens78] Volker Enss. Asymptotic completeness for quantum mechanical potential scattering. I. Short range potentials. *Communications in Mathematical Physics*, 61(3):285–291, 1978.

[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.

[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.

[FWM+14] Liang Feng, Zi Jing Wong, Ren-Min Ma, Yuan Wang, and Xiang Zhang. Single-mode laser by parity-time symmetry breaking. *Science*, 346(6212):972–975, 2014.

[FZ99a] Joshua Feinberg and A. Zee. Non-Hermitian localization and delocalization. *Physical Review E*, 59(6):6433, 1999.

[FZ99b] Joshua Feinberg and A. Zee. Spectral curves of non-Hermitian Hamiltonians. *Nuclear Physics B*, 552(3):599–623, 1999.

[GEB+15] Tiejun Gao, E. Estrecho, K. Y. Bliokh, T. C. H. Liew, M. D. Fraser, Sebastian Brodbeck, Martin Kamp, Christian Schneider, Sven Höfling, Y. Yamamoto, et al. Observation of non-Hermitian degeneracies in a chaotic exciton-polariton billiard. *Nature*, 526(7574):554–558, 2015.

[GG13] Andre K. Geim and Irina V. Grigorieva. Van der Waals heterostructures. *Nature*, 499(7459):419–425, 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.

[GK00] Ilya Ya Goldsheid and Boris A. Khoruzhenko. Eigenvalue curves of asymmetric tridiagonal random matrices. *Electron. J. Probab.*, 5:no. 16, 28, 2000.

[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.

[Glo75] Josip Globevnik. On complex strict and uniform convexity. *Proc. Amer. Math. Soc.*, 47:175–178, 1975.

[Glo76] Josip Globevnik. Norm-constant analytic functions and equivalent norms. *Illinois J. Math.*, 20(3):503–506, 1976.

[GMM09] David Gabai, Robert Meyerhoff, and Peter Milley. Minimum volume cusped hyperbolic three-manifolds. *Journal of the American Mathematical Society*, 22(4):1157–1215, 2009.

[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.

[GS97] 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.

[GS03] V. Girardin and R. Senoussi. Semigroup stationary processes and spectral representation. *Bernoulli*, 9(5):857–876, 2003.

[GSD+09] A. Guo, G. J. Salamo, D. Duchesne, R. Morandotti, M. Volatier-Ravat, V. Aimez, G. A. Siviloglou, and D. N. Christodoulides. Observation of PT-symmetry breaking in complex optical potentials. *Physical Review Letters*, 103(9):093902, 2009.

[GSS15] Philippe J. Gaudreau, Richard M. Slevinsky, and Hassan Safouhi. Computing energy eigenvalues of anharmonic oscillators using the double exponential Sinc collocation method. *Annals of Physics*, 360:520–538, 2015.

[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.

[HA50] David Hilbert and Wilhelm Ackermann. *Principles of mathematical logic*, volume 69. American Mathematical Soc., 1950.

[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.

[Hag16a] Raffael Hagger. *Fredholm Theory with Applications to Random Operators*. PhD thesis, Technische Universität Hamburg, 2016.

[Hag16b] Raffael Hagger. On the spectrum and numerical range of tridiagonal random operators. *Journal of Spectral Theory*, 6(2):215–266, 2016.

[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 multi-dimensional integrals. *Numerische Mathematik*, 2:84–90, 1960.

[Hal63] Paul R. Halmos. What does the spectral theorem say? *Amer. Math. Monthly*, 70:241–247, 1963.

[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.

[Han08a] Anders C. Hansen. *On the approximation of spectra of linear Hilbert space operators*. PhD thesis, University of Cambridge, 2008.

[Han08b] Anders C. Hansen. On the approximation of spectra of linear operators on Hilbert spaces. *Journal of Functional Analysis*, 254(8):2092–2126, 2008.

[Han10] Anders C. Hansen. Infinite-dimensional numerical linear algebra: theory and applications. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 466(2124):3539–3559, 2010.

[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.

[HK87] Tetsuo Hatakeyama and Hiroshi Kamimura. Electronic properties of a Penrose tiling lattice in a magnetic field. *Solid State Comm.*, 62(2):79–83, 1987.

[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.

[HLS16] Raffael Hagger, Marko Lindner, and Markus Seidel. Essential pseudospectra and essential norms of band-dominated operators. *J. Math. Anal. Appl.*, 437(1):255–291, 2016.

[HMH+14] Hossein Hodaei, Mohammad-Ali Miri, Matthias Heinrich, Demetrios N. Christodoulides, and Mercedeh Khajavikhan. Parity-time–symmetric microring lasers. *Science*, 346(6212):975–978, 2014.

[HN96] Naomichi Hatano and David R. Nelson. Localization transitions in non-Hermitian quantum mechanics. *Physical Review Letters*, 77(3):570, 1996.

[HN97] Naomichi Hatano and David R. Nelson. Vortex pinning and non-Hermitian quantum mechanics. *Physical Review B*, 56(14):8651, 1997.

[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.

[HOZ03] Daniel E. Holz, Henri Orland, and A. Zee. On the remarkable spectrum of a non-Hermitian random matrix model. *Journal of Physics A: Mathematical and General*, 36(12):3385, 2003.

[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.

[HT20] Andrew Horning and Alex Townsend. Feast for differential eigenvalue problems. *SIAM Journal on Numerical Analysis*, 58(2):1239–1262, 2020.

[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.

[Hun90] Fern Hunt. Error analysis and convergence of capacity dimension algorithms. *SIAM Journal on Applied Mathematics*, 50(1):307–321, 1990.

[HWY02] Naomichi Hatano, Takahiro Watanabe, and Junko Yamasaki. Localization, resonance and non-Hermitian quantum mechanics. *Physica A: Statistical Mechanics and its Applications*, 314(1):170–176, 2002.

[Ind61] Jack Indritz. An inequality for Hermite polynomials. *Proc. Amer. Math. Soc.*, 12:981–983, 1961.

[Jit99] Svetlana Jitomirskaya. Metal-insulator transition for the almost Mathieu operator. *Annals of Mathematics*, 150(3):1159–1175, 1999.

[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.

[JT98] Gudbjorn F. Jónsson and Lloyd N. Trefethen. A numerical analyst looks at the "cutoff phenomenon" in card shuffling and other Markov chains. In *Numerical analysis 1997 (Dundee)*, volume 380 of *Pitman Res. Notes Math. Ser.*, pages 150–178. Longman, Harlow, 1998.

[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.

[Kat49] Tosio Kato. On the upper and lower bounds of eigenvalues. *Journal of the Physical Society of Japan*, 4:334–339, 1949.

[Kat95] Tosio Kato. *Perturbation theory for linear operators*. Classics in Mathematics. Springer-Verlag, Berlin, 1995.

[KGM08] Shachar Klaiman, Uwe Günther, and Nimrod Moiseyev. Visualization of branch points in PT-symmetric waveguides. *Physical Review Letters*, 101(8):080402, 2008.

[Kir07] Werner Kirsch. An invitation to random Schrödinger operators. *arXiv:0709.3707*, 25:1–119, 2007.

[KK47] Mark G. Krein and Mark A. Krasnoselski. Fundamental theorems concerning the extension of Hermitian operators and some of their applications to the theory of orthogonal polynomials and the moment problem. *Uspekhi Mat. Nauk.*, 2:60–106, 1947.

[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.

[KKL03] Rowan Killip, Alexander Kiselev, and Yoram Last. Dynamical upper bounds on wavepacket spreading. *American journal of mathematics*, 125(5):1165–1198, 2003.

[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.

[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.

[KM07] Werner Kirsch and Bernd Metzger. The integrated density of states for random Schrödinger operators. In *Spectral theory and mathematical physics: a Festschrift in honor of Barry Simon's 60th birthday*, volume 76 of *Proc. Sympos. Pure Math.*, pages 649–696. Amer. Math. Soc., Providence, RI, 2007.

[KNO19] Marek Kaluba, Piotr W. Nowak, and Narutaka Ozawa. $\mathrm{Aut}(\mathbb{F}_5)$ has property (T). *Mathematische annalen*, 375(3):1169, 2019.

[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.

[KS03] R. Killip and B. Simon. Sum rules for Jacobi matrices and their applications to spectral theory. *Ann. Math.*, pages 253–321, 2003.

[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.

[Las96] Yoram Last. Quantum dynamics and decompositions of singular continuous spectra. *Journal of Functional Analysis*, 142(2):406–445, 1996.

[Leš88] Gorazd Lešnjak. Complex convexity and finitely additive vector measures. *Proc. Amer. Math. Soc.*, 102(4):867–873, 1988.

[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.

[Lon09] Stefano Longhi. Bloch oscillations in complex crystals with PT symmetry. *Physical Review Letters*, 103(12):123601, 2009.

[LR12] Marko Lindner and Steffen Roch. Finite sections of random Jacobi operators. *SIAM Journal on Numerical Analysis*, 50(1):287–306, 2012.

[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.

[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.

[Lub08] Christian Lubich. *From quantum to classical molecular dynamics: reduced models and numerical analysis*. Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS), Zürich, 2008.

[MÖ92] 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.

[Mat86] Daniel C. Mattis. The few-body problem on a lattice. *Reviews of Modern Physics*, 58(2):361, 1986.

[Mat95] Pertti Mattila. *Geometry of sets and measures in Euclidean spaces*, volume 44 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995.

[MBLM+10] B. Malcolm Brown, Matthias Langer, Marco Marletta, Christiane Tretter, and Markus Wagenhofer. Eigenvalue enclosures and exclosures for non-self-adjoint problems in hydrodynamics. *LMS Journal of Computation and Mathematics*, 13:65–81, 2010.

[McM87] Curt McMullen. Families of rational maps and iterative root-finding algorithms. *Annals of Mathematics (2)*, 125(3):467–493, 1987.

[McM88] Curt McMullen. Braiding of the attractor and the failure of iterative algorithms. *Invent. Math.*, 91(2):259–272, 1988.

[MEGCM08] Konstantinos G. Makris, Ramy El-Ganainy, Demetrios N. Christodoulides, and Ziad H. Musslimani. Beam dynamics in PT symmetric optical lattices. *Physical Review Letters*, 100(10):103904, 2008.

[Meh04] Madan L. Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.

[Mog91] Alexander Mogilner. Hamiltonians in solid state physics as multiparticle discrete Schrödinger operators. *Advances in Societ Mathematics*, 5:139–194, 1991.

[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.

[Mui82] Robb J. Muirhead. *Aspects of multivariate statistical theory*. John Wiley & Sons, Inc., New York, 1982.

[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.

[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.

[NH13] Yuji Nakatsukasa and Nicholas J Higham. Stable and efficient spectral divide and conquer algorithms for the symmetric eigenvalue decomposition and the SVD. *SIAM Journal on Scientific Computing*, 35(3):A1325–A1349, 2013.

[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.

[Nov11] K. S. Novoselov. Nobel lecture: Graphene: Materials in the flatland. *Reviews of Modern Physics*, 83(3):837, 2011.

[NS98] David R. Nelson and Nadav M. Shnerb. Non-Hermitian localization and population biology. *Physical Review E*, 58(2):1383, 1998.

[NYWM01] G. H. C. New, M. A. Yates, J. P. Woerdman, and G. S. McDonald. Diffractive origin of fractal resonator modes. *Optics communications*, 193(1-6):261–266, 2001.

[Olv18] Sheehan Olver. ApproxFun.jl v0.8. *github (online) https://github.com/JuliaApproximation/ApproxFun.jl*, 2018.

[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.

[OT14] Sheehan Olver and Alex Townsend. A Practical Framework for Infinite-dimensional Linear Algebra. In *Proceedings of the 1st First Workshop for High Performance Technical Computing in Dynamic Languages*, HPTCDL '14, pages 57–62, Piscataway, NJ, USA, 2014. IEEE Press.

[OW18] Sheehan Olver and Marcus Webb. SpectralMeasures.jl. *github (online) https://github.com/JuliaApproximation/SpectralMeasures.jl*, 2018.

[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.

[PEF15] Charles Puelz, Mark Embree, and Jake Fillman. Spectral approximation for quasiperiodic Jacobi operators. *Integral Equations and Operator Theory*, 82(4):533–554, 2015.

[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.

[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.

[RBM+12] Alois Regensburger, Christoph Bersch, Mohammad-Ali Miri, Georgy Onishchukov, Demetrios N. Christodoulides, and Ulf Peschel. Parity-time synthetic photonic lattices. *Nature*, 488(7410):167–171, 2012.

[Rem98] Christian Remling. The absolutely continuous spectrum of one-dimensional Schrödinger operators with decaying potentials. *Communications in Mathematical Physics*, 193(1):151–170, 1998.

[RGSE18] José A. Rivera, Thomas C. Galvin, Austin W. Steinforth, and J. Gary Eden. Fractal modes and multibeam generation from hybrid microlaser resonators. *Nature communications*, 9(1):1–8, 2018.

[RMB+13] Alois Regensburger, Mohammad-Ali Miri, Christoph Bersch, Jakob Näger, Georgy Onishchukov, Demetrios N. Christodoulides, and Ulf Peschel. Observation of defect states in PT-symmetric optical lattices. *Physical Review Letters*, 110(22):223902, 2013.

[RMEG+10] Christian E. Rüter, Konstantinos G. Makris, Ramy El-Ganainy, Demetrios N. Christodoulides, Mordechai Segev, and Detlef Kip. Observation of parity–time symmetry in optics. *Nature Phys.*, 6(3):192–195, 2010.

[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.

[SBGC84] Dan Shechtman, Ilan Blech, Dennis Gratias, and John W. Cahn. Metallic phase with long-range orientational order and no translational symmetry. *Physical Review Letters*, 53:1951–1953, Nov 1984.

[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.

[Sch13] Henning Schomerus. Topologically protected midgap states in complex photonic lattices. *Optics Letters*, 38(11):1912–1914, 2013.

[Sei12] Markus Seidel. On (N, ϵ)-pseudospectra of operators on Banach spaces. *Journal of Functional Analysis*, 262(11):4916–4927, 2012.

[Sei14] Markus Seidel. Fredholm theory for band-dominated and related operators: a survey. *Linear Algebra Appl.*, 445:373–394, 2014.

[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.

[Sha08] Eugene Shargorodsky. On the level sets of the resolvent norm of a linear operator. *Bull. Lond. Math. Soc.*, 40(3):493–504, 2008.

[Sha13] Eugene Shargorodsky. On the limit behaviour of second order relative spectra of self-adjoint operators. *Journal of Spectral Theory*, 3(4):535–552, 2013.

[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.

[Sim95] Barry Simon. Operators with singular continuous spectrum. I. General operators. *Annals of Mathematics (2)*, 141(1):131–145, 1995.

[Sim00] Barry Simon. Schrödinger operators in the twenty-first century. *Mathematical physics*, 2000:283–288, 2000.

[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.

[SLZ+11] Joseph Schindler, Ang Li, Mei C. Zheng, Fred M. Ellis, and Tsampikos Kottos. Experimental study of active LRC circuits with PT symmetries. *Physical Review A*, 84(4):040101, 2011.

[Sma81] Steve Smale. The fundamental theorem of algebra and complexity theory. *Bull. Amer. Math. Soc. (N.S.)*, 4(1):1–36, 1981.

[Sma85] Steve Smale. On the efficiency of algorithms of analysis. *Bull. Amer. Math. Soc. (N.S.)*, 13(2):87–121, 1985.

[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.

[Sma98] Steve Smale. The work of Curtis T. McMullen. In *Proceedings of the International Congress of Mathematicians, Vol. I (Berlin, 1998)*, pages 127–132, 1998.

[SN98] Nadav M. Shnerb and David R. Nelson. Winding numbers, complex currents, and non-Hermitian localization. *Physical Review Letters*, 80(23):5172, 1998.

[SS60] Palle Schmidt and Frank Spitzer. The Toeplitz matrices of an arbitrary Laurent polynomial. *Math. Scand.*, 8:15–38, 1960.

[Sta65] Joseph G. Stampfli. Hyponormal operators and spectral density. *Trans. Amer. Math. Soc.*, 117:469–476, 1965.

[Sta12] Zbigniew M. Stadnik. *Physical properties of quasicrystals*, volume 126. Springer Science & Business Media, 2012.

[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.

[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.

[TD99] F. Tisseur and J. Dongarra. A parallel divide and conquer algorithm for the symmetric eigenvalue problem on distributed memory architectures. *SIAM Journal on Scientific Computing*, 20(6):2223–2236, 1999.

[TDGG15] Duc-Thanh Tran, Alexandre Dauphin, Nathan Goldman, and Pierre Gaspard. Topological Hofstadter insulators in a two-dimensional quasicrystal. *Physical Review B*, 91(8):085125, 2015.

[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.

[Tho83] D. J. Thouless. Bandwidths for a quasiperiodic tight-binding model. *Physical Review B*, 28(8):4272, 1983.

[Tho90] D. J. Thouless. Scaling for the discrete Mathieu equation. *Communications in mathematical physics*, 127(1):187–193, 1990.

[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.

[Tsy08] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Science & Business Media, 2008.

[TT91] D. J. Thouless and Yong Tan. Total bandwidth for the Harper equation. III. corrections to scaling. *Journal of Physics A: Mathematical and General*, 24(17):4055, 1991.

[TT13] Alex Townsend and Lloyd N. Trefethen. An extension of Chebfun to two dimensions. *SIAM J. Sci. Comput.*, 35(6):C495–C518, 2013.

[TTK15] Shinichi Takemura, Nayuta Takemori, and Akihisa Koga. Valence fluctuations and electric reconstruction in the extended Anderson model on the two-dimensional Penrose lattice. *Physical Review B*, 91(16):165114, 2015.

[Tuc11] Warwick Tucker. *Validated numerics: a short introduction to rigorous computations*. Princeton University Press, 2011.

[Tur36] Alan M. Turing. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proc. London Math. Soc. (2)*, 42(3):230–265, 1936.

[Tur10] Alexander V. Turbiner. Double well potential: perturbation theory, tunneling, WKB (beyond instantons). *International Journal of Modern Physics A*, 25(02n03):647–658, 2010.

[VM04] Julien Vidal and Rémy Mosseri. Quasiperiodic tilings in a magnetic field. *Jour. non-cryst. sol.*, 334:130–136, 2004.

[VNA13] Z. Valy Vardeny, Ajay Nahata, and Amit Agrawal. Optics of photonic quasicrystals. *Nature Phot.*, 7(3):177–187, 2013.

[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.

[Web17] Marcus Webb. *Isospectral algorithms, Toeplitz matrices and orthogonal polynomials*. PhD thesis, University of Cambridge, 2017.

[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.

[WO17] Marcus Webb and Sheehan Olver. Spectra of Jacobi operators via connection coefficient matrices. *arXiv:1702.03095*, 2017.

[WRM+15] Martin Wimmer, Alois Regensburger, Mohammad-Ali Miri, Christoph Bersch, Demetrios N. Christodoulides, and Ulf Peschel. Observation of optical solitons in PT-symmetric lattices. *Nature Communications*, 6, 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.

[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.

[ZHI+15] Bo Zhen, Chia Wei Hsu, Yuichi Igarashi, Ling Lu, Ido Kaminer, Adi Pick, Song-Liang Chua, John D. Joannopoulos, and Marin Soljačić. Spawning rings of exceptional points out of Dirac cones. *Nature*, 525(7569):354–358, 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.