Eigenvalue Statistics for Random Block Operators

Daniel Flint Schmidt

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Mathematics

Alexander Elgart, Chair
Joseph Ball
George Hagedorn
Martin Klaus

March 25, 2015
Blacksburg, Virginia

Keywords: Eigenvalue statistics, Wegner estimate, Minami estimate, Schrödinger operators

Eigenvalue Statistics for Random Block Operators

Daniel Flint Schmidt

(ABSTRACT)

The Schrödinger Hamiltonian for a single electron in a crystalline solid with independent, identically distributed (i.i.d.) single-site potentials has been well studied. It has the form of a diagonal potential energy operator, which contains the random variables, plus a kinetic energy operator, which is deterministic. In the less-understood cases of multiple interacting charge carriers, or of correlated random variables, the Hamiltonian can take the form of a random block-diagonal operator, plus the usual kinetic energy term. Thus, it is of interest to understand the eigenvalue statistics for such operators.

In this work, we establish a criterion under which certain random block operators will be guaranteed to satisfy Wegner, Minami, and higher-order estimates. This criterion is phrased in terms of properties of individual blocks of the Hamiltonian. We will then verify the input conditions of this criterion for a certain quasiparticle model with i.i.d. single-site potentials. Next, we will present a progress report on a project to verify the same input conditions for a class of one-dimensional, single-particle alloy-type models. These two results should be sufficient to demonstrate the utility of the criterion as a method of proving Wegner and Minami estimates for random block operators.

This work was supported in part by NSF grant DMS-1210982.

Dedication

“Of making many books there is no end, and much study wearies the body.
Now all has been heard; here is the conclusion of the matter:
Fear God and keep his commandments, for this is the duty of all mankind.”

Ecclesiastes 12:12-13 (NIV)

Acknowledgments

I would like to thank my adviser, Alexander Elgart, for his guidance over the last four years, and George Hagedorn, Joseph Ball, and Martin Klaus for serving on my committee. I should also thank Ezra Brown for his help at various times, and Alan Lattimer for covering my class to make some of my conference travel possible. Finally, I thank my family for being a supportive presence throughout this process.

Attribution

Chapter 2 is based on a paper coauthored by Alexander Elgart and Daniel Schmidt, which has been accepted for publication in the *Journal of Spectral Theory* [12]. Chapters 3 and 4 are from a manuscript coauthored by Alexander Elgart and Daniel Schmidt, which is as-yet unpublished.

Contents

Chapter 1

Introduction

1.1 Background

The study of random Schrödinger operators is an attempt to understand certain questions arising in the mathematical foundations of quantum mechanics. The central question in this study concerns the reasons for the presence or absence of diffusion of various particles or disturbances in solids. This field of study traces its origin largely to a paper by P.W. Anderson in 1958 [3].

Anderson's original work focused on the absence of diffusion of spin waves in certain solids, and explained this absence in terms of a random component of the electric potential. Similar reasoning has been applied to the diffusion of electrons and other charge carriers. In both contexts, a sufficiently strong random component of the potential can cause severe interference between various multiple-scattering paths, which in some cases is sufficient to halt the long range propagation that might have occurred in a strictly periodic potential.

The standard Anderson model is a "single particle" model, in the sense that it does not include electron-electron interactions. It also assumes that the random variables that determine the potential are independent and identically distributed. In this work, we will consider models which generalize the Anderson model in various ways. In one case, we will modify the model by considering the motion of a quasiparticle, instead of the usual single electron. In a second case, we will modify the model by allowing the random variables in the potential to be correlated to some degree, instead of the usual case of i.i.d. random variables.

1.2 Mathematical Setting of the Problem

We begin with an overview of the mathematical setting of this work. (Note that this chapter will be only an abbreviated account of this background. Excellent references are available for further

details—see for example [19], [27], and [8].)

Consider a solid with a crystal structure, which we may represent by a graph. Suppose ions of various charges (perhaps including both positive and negative charges) are arranged randomly at the lattice points of this crystal. The electrons in this solid will in reality be influenced by electrostatic forces from both the lattice ions and the other electrons, but as a simplifying assumption, it is standard practice to ignore electron-electron interactions. This gives us a so-called "single particle" model, as described above. There are at least two major questions we could ask about the electrons in this system. First, what energy levels can an electron occupy? Second, given some energy level, will an electron at this energy be in a bound state or in a free state?

Normally, one would address questions about the dynamical behavior of electrons in any given state by using the time-dependent Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\psi(\mathbf{r},t) = H\psi(\mathbf{r},t) \tag{1.2.1}$$

Where $\psi(\mathbf{r},t)$ is the electron's state function and $H = H_0 + V$ is the Hamiltonian for the system, also referred to as the Schrödinger operator. Here H_0 is the Laplacian, or kinetic energy operator, and V is a multiplication operator representing the electric potential energy. Note that any randomness in the arrangement of ions will appear only in the potential energy term V.

Similarly, one would find the electron's "stationary states" and the corresponding energy levels by solving the time-independent Schrödinger equation:

$$H\psi(\mathbf{r}) = E\psi(\mathbf{r}) \tag{1.2.2}$$

Note that this equation takes the form of an eigenvalue problem, and thus it would be natural to expect that the energy levels and state functions would be given by the eigenvalues and eigenfunctions, respectively. In fact, the matter is more subtle than this intuition suggests. Since the Hamiltonian may be an operator on an infinite-dimensional vector space, its spectrum can include points which are not, strictly speaking, eigenvalues. This is not just a technical point—the distinction between various types of points in the spectrum will turn out to be crucial to determining the dynamical behavior of electrons, and we will revisit that matter below.

In this work, we will use the tight binding approximation, in which we model the electrons as jumping from one lattice point to another, rather than traveling freely through all of space. This means that the space the electrons occupy is $\mathbb{Z}^d$ or some other graph, rather than $\mathbb{R}^d$. This in turn means that the appropriate form of the Laplacian H_0 is the so-called "discrete Laplacian" or "graph Laplacian" given below.

$$(H_0\psi)(n) = \sum_{d(n,m)=1} (\psi(m) - \psi(n))$$

Here $d(m,n) = ||m-n||_1$ is a metric defined in terms of the ℓ^1 norm.

Also, in our context, the potential operator is the sum of the single-site potentials $u_i(n)$ generated by the individual lattice ions.

$$V(n) = \sum_{i \in \mathbb{Z}^d} u_i(n)$$

This gives the form of the Schrödinger equation that is relevant to our situation. In principle, all we need to do is to insert the potential for any given model, and solve the two forms of the Schrödinger equation to answer our questions about the energy levels and dynamical behaviors of electrons. Unfortunately, while it is possible to solve the Schrödinger equation exactly for some simple systems, this will be unfeasible for a large crystal lattice with a very complex potential. Furthermore, in our context we want the arrangement of ions at the lattice points to be random, so finding the solutions for any single configuration would not be sufficient anyway. Instead, we want statistical information about how the electrons and their energy levels will behave in "almost all" configurations, where "almost all" is understood to be a measure-theoretic statement with respect to the probability measure that determines the distribution of ions.

1.3 Spectral Decomposition

Our first major tool will be a decomposition of the spectrum $\sigma(H)$ of the Hamiltonian. It will take several steps to define the spectral decomposition, (and the corresponding decomposition of the state space) but when we are finished, there will be a close connection between the part of the spectrum in which an energy level lies, and the dynamical behavior of the corresponding state.

Given an operator H, define a **projection-valued spectral measure** to be any function μ_H which maps subsets of the spectrum $\sigma(H)$ to projections on the state space $\mathcal{H}$, satisfying the following:

1. $\mu_H(A)$ is an orthogonal projection, for any set $A \subset \sigma(H)$.
2. $\mu_H(\emptyset) = 0$, where 0 is to be understood as an operator.
3. $\mu_H(\sigma(H)) = I$, where I is the identity operator on $\mathcal{H}$.
4. $\mu_H(A \cap B) = \mu_H(A)\mu_H(B)$, for any $A, B \subset \sigma(H)$.
5. $\mu_H(\bigcup_{n=0}^{\infty} A_n) = \sum_{n=0}^{\infty} \mu_H(A_n)$ for any collection of pairwise disjoint sets $A_n \subset \sigma(H)$.

More specifically, we may define a particular projection-valued spectral measure μ_H as follows. For a given set of energies $A \subset \sigma(H)$, let $\mu_H(A)$ be the projection operator onto the subspace of $\mathcal{H}$ spanned by the state functions whose energies are in A. (See [19], [8].)

The purpose of the projection-valued spectral measure is to define the following (real-valued) measure:

$$\mu_\psi(A) = \langle \psi, \mu_H(A)\psi \rangle$$

This gives a different real-valued measure for each function $\psi \in \mathcal{H}$. This measure can be loosely thought of as quantifying the dependence of a state function ψ on the "basis states." Hence if $\psi = c_1\psi_1 + c_2\psi_2 + c_3\psi_3$, where the functions ψ_i are stationary states corresponding to eigenvalues E_i, respectively, then $\mu_\psi(\{E_1, E_2\}) = |c_1|^2 + |c_2|^2$, for example.

Strictly speaking, a rigorous treatment of the use of stationary states as a basis for $\mathcal{H}$ will require a discussion of rigged Hilbert spaces. For our purposes, this will not be necessary, and the above paragraph may be interpreted simply as a guide to intuition. For a treatment of rigged Hilbert spaces, see [23].

We will split the spectrum into three parts by applying a modified form of the Lebesgue decomposition theorem to the spectral measure μ_ψ. For reference, the usual version of that theorem is stated below. (See [26].)

Theorem 1.3.1. Lebesgue Decomposition Theorem *For any two $\sigma-$finite measures μ and ν on a measure space (Ω, Σ), there exist two sigma finite measures ν_s and ν_{ac} on (Ω, Σ) such that:*

1. $\nu = \nu_s + \nu_{ac}$
2. *The measure ν_{ac} is absolutely continuous with respect to μ.*
3. *The measure ν_s is singular with respect to μ.*

Furthermore, the measures ν_s and ν_{ac} are unique.

In applying this theorem, we will let $\mu = m$, the Lebesgue measure. Also, we will decompose the measure ν_s further into two more measures ν_{sc} and ν_{pp}. The first of these is called a *singular continuous measure*, which means that it is singular with respect to Lebesgue measure, but still satisfies $\nu_{sc}(x) = 0$ for any singleton. The second measure ν_{pp} is called a *pure point measure*, and is a sum of Dirac measures (i.e. it is supported only on singletons).

Now we may decompose the state space $\mathcal{H}$ according to the spectral decomposition.

1. If μ_ψ is absolutely continuous with respect to Lebesgue measure, then ψ is in the space $\mathcal{H}_{ac}$.
2. If μ_ψ is singular continuous with respect to Lebesgue measure, then ψ is in the space $\mathcal{H}_{sc}$.
3. If the measure μ_ψ is pure point, then ψ is in the space $\mathcal{H}_{pp}$.

Note that under these definitions, the original state space is not a union of the three new spaces, but rather a direct sum of the three, since a spectral measure μ_ψ may be a sum of components of all three types. Thus, we have:

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

We have used the decomposition of measures to define a decomposition of the state space. Now

we use the latter to define a decomposition of the spectrum of H. Define

$$\sigma_{pp} := \sigma(H|\mathcal{H}_{pp})$$
$$\sigma_{ac} := \sigma(H|\mathcal{H}_{ac})$$
$$\sigma_{sc} := \sigma(H|\mathcal{H}_{sc})$$

The set σ_{pp} is in fact the closure of the set of eigenvalues of H.

We now have $\sigma(H) = \sigma_{pp} \cup \sigma_{ac} \cup \sigma_{sc}$. This decomposition turns out to be physically significant, since the spectral type of an energy level will be related to the dynamical properties of the associated state.

1.4 Physical Significance of the Spectral Decomposition

A theorem by Ruelle, Amrein, Georgescu and Enss, which we will not state in full rigor here, clarifies the physical significance of the spectral decomposition.

1. Energies in σ_{pp} correspond to bound states of electrons.
2. Energies in σ_{ac} correspond to free states in which electrons travel to infinity.
3. Energies in σ_{sc} correspond to states in which electrons travel to infinity in the time average.

For further details, see [19].

For discrete random Schrödinger operators, we know that the spectrum itself is nonrandom: with probability 1, we have

$$\sigma = \{V(x) : x \in \Lambda\} + [0, 4d]$$

Here Λ is a finite box restriction of the lattice, and d is the dimension of the lattice. However, we do not know in general where σ_{pp}, σ_{ac}, and σ_{sc} lie within the spectrum, or even whether any given set among these three is non-empty.

We can describe several vague, often unproven, expectations for what the spectrum of a Schrödinger operator should look like in various situations. For example, in a model with random potentials and no long-range correlation of random variables, σ_{sc} is expected to be empty. Note that this is not true in general: the spectrum for a Hamiltonian describing a quasicrystal—in which the single-site potentials have long-range, non-periodic order—may include all three spectral types [19].

Also, we expect the spectrum to have a band structure, with pure point spectrum near the band edges and absolutely continuous spectrum near the center of each band, at least for weak disorder in dimension $d \geq 3$. In stronger disorder regimes, the pure point spectrum can extend all the way to the center of the band [1]. In dimension $d = 1$, the Hamiltonian is known to have dense pure

point spectrum for any nonzero disorder. Random Schrödinger operators on dimension $d = 2$ are an open problem [4].

Once again, not all of these expected results are known rigorously. In particular, the existence of σ_{ac} has not been proved in most cases. On the other hand, we do have ways to prove the existence of σ_{pp} for some models. Two common methods are the fractional moment method and the method of multiscale analysis (MSA). The latter is an inductive procedure on the size of a finite box restriction of the lattice. The use of this method requires an initial-scale estimate for the initial step, and a Wegner estimate (to be described below) for the inductive step.

A Wegner estimate is an inequality of the following form:

$$\mathbb{P}((\sigma(H^\Lambda) \cap I) \neq \emptyset) \leq C|\Lambda||I| \tag{1.4.1}$$

If we assume, without loss of generality, that the interval I has the form $I = (-\epsilon, \epsilon)$, then this can be rewritten as:

$$\mathbb{P}(\mathcal{C}_\epsilon(H^\Lambda) \geq 1) \leq C|\Lambda|\epsilon \tag{1.4.2}$$

where $\mathcal{C}_\epsilon$ is the function that counts eigenvalues, according to algebraic multiplicity, in the interval $(-\epsilon, \epsilon)$.

Note that the Wegner estimate counts eigenvalues only. Since we are using a finite box restriction of the Hamiltonian, rather than the original Hamiltonian, the operator whose spectrum we consider acts on a finite-dimensional vector space. Thus, while H itself may have pure point, absolutely continuous, and/or singular continuous spectrum, all points in the spectrum of H^Λ are eigenvalues.

In light of the above discussion, proving a Wegner estimate is a key to proving existence of the pure point spectrum, which in turn gives localization of electrons. The Wegner estimate turns out to have other uses as well, such as proving Poisson statistics of eigenvalues. In that case, we also need to have a Minami estimate:

$$\mathbb{P}(\mathcal{C}_\epsilon(H^\Lambda) \geq 2) \leq C|\Lambda|\epsilon^2 \tag{1.4.3}$$

Thus, we can determine the dynamical properties of electrons (among other things) in a given system if we have information about the statistical distribution of the eigenvalues for a finite box restriction of the Hamiltonian. The remainder of this work will be devoted to studying the eigenvalue statistics for certain types of Hamiltonians. In some cases, the Wegner and Minami estimates are already known. For example, the simplest kind of random Schrödinger operator would be a single particle model in which the single site potentials are i.i.d. random variables (i.e. the standard Anderson model). In this case, Wegner and Minami estimates have been proved already [24].

It is more difficult to establish the eigenvalue statistics for models with correlated random variables, or (in a mathematically similar problem) models that involve multiple interacting charge carriers. In both cases, the Schrödinger operator can take the form of a block operator, in which different

blocks depend on different random variables, but the random variables within a block may be correlated.

Loosely speaking, the more correlated variables the model includes, the more difficult the proofs tend to be. In the extreme case in which all the variables are equal, the Minami estimate is actually false. In the other extreme, in which there are no correlations, we have the standard Anderson model, and the estimates are already known. The open questions lie between these two extremes.

1.5 Organization of this Work

In chapter 2, we establish a general criterion under which random block operators are guaranteed to satisfy Wegner and Minami estimates. We then apply this criterion to prove the two estimates for a certain quasiparicle model arising from the theory of dirty superconductors.

In fact, the first major theorem of chapter 2 is actually a more abstract result from linear algebra. We use this theorem to demonstrate the eigenvalue statistics criterion, which is the main application theorem of that chapter. After that criterion is established, the proof that it applies to the specific case of the quasiparticle model is actually straightforward.

In chapter 3, the situation is reversed: the implications of the criterion may be assumed by this time, but the proof that the criterion applies to the specific model is challenging. Chapter 3 is a progress report on an attempt to apply this criterion to a certain class of alloy-type models.

In order to prove that, we need to prove another abstract theorem, this time from algebraic geometry. This theorem shows that the solution set to a system of n polynomial equations in $n + m$ variables is almost always of dimension m, and its measure is normally not too small, in a sense to be made precise later. Chapter 4 gives the proof of this theorem.

Chapter 2

Eigenvalue counting inequalities, with applications to Schrödinger operators

2.1 Introduction

2.1.1 Small eigenvalues and the Green function

Let A be an invertible Hermitian $N \times N$ matrix with inverse A^{-1}, and let I_N be the $N \times N$ identity matrix. Let $G(x, y)$ denote the matrix element in the (x, y) position of A^{-1}, also known as the Green function of A. Our first objective in this work is to relate information about the small eigenvalues of A to the behavior of $G(x, y)$. Let us denote by $C_\epsilon(A)$ the number of eigenvalues (counting multiplicities) of A in the interval $I_\epsilon := (-\epsilon, \epsilon)$. As a first step, let us ask the most basic question: Does A have at least one eigenvalue in the interval I_ϵ? A well known result in the matrix analysis says that

$$C_\epsilon(A) > 0 \iff \|A^{-1}\| > \frac{1}{\epsilon}.$$

Since $\|B\|_{\max} \leq \|B\| \leq N\|B\|_{\max}$ for any $N \times N$ matrix B with

$$\|B\|_{\max} = \max_{x,y} |B(x, y)|,$$

we obtain the relations

$$C_\epsilon(A) > 0 \Rightarrow \text{There exists a pair } \{x, y\} \text{ such that } |G(x, y)| > \frac{1}{N\epsilon}\,; \tag{2.1.1}$$

$$|G(x, y)| > \frac{1}{\epsilon} \text{ for some pair } \{x, y\} \Rightarrow C_\epsilon(A) > 0\,.$$

It is natural to try to quantify these relations further, viz. to detect whether the matrix A has at least m small eigenvalues from the behavior of $G(x, y)$. To this end, we prove the following result.

Theorem 2.1.1. *Let $A = A^*$ be an $N \times N$ invertible matrix. Let $A[\alpha, \beta]$ denote the submatrix of A with rows indexed by index subset α and columns indexed by index subset β. Consider the following two assertions:*

I. $C_\epsilon(A) \geq m$;

II. There exist index subsets

$$\alpha_m = \{i_1, \ldots, i_m\}, \quad \beta_m = \{j_1, \ldots, j_m\}$$

of $\{1, \ldots, N\}$ *such that*

$$A^{-1}[\alpha_m, \beta_m]A^{-1}[\beta_m, \alpha_m] > \frac{K^2}{\epsilon^2} I_N[\alpha_m, \alpha_m] \text{ for some } K > 0, \tag{2.1.2}$$

where I_N is the $N \times N$ identity matrix.

Then (I) *implies* (II) *with*

$$K = \frac{C_m}{N}, \quad C_m = \frac{1}{m!\, 2^{m-1}} \tag{2.1.3}$$

Conversely, (II) *with $K = 1$ implies* (I)*.*

The constant C_m in (2.1.3) is not sharp for $m > 1$. However, the dependence on N is optimal (and we will be interested in small m, large N behavior in the application below).

It is often convenient to work with principal submatrices $A^{-1}[\gamma]$ of A^{-1}. One can tailor Theorem 2.1.1 somewhat differently to accommodate this requirement, at the cost of increasing the cardinality of the corresponding index subsets α_m, β_m. Namely, we have the following result:

Corollary 2.1.2. *Let $A = A^*$ be an $N \times N$ invertible matrix. Consider the following two assertions:*

I. $C_\epsilon(A) \geq m$;

II. There exists an index subset $\gamma_m = \{i_1, \ldots, i_{2m}\}$ of $\{1, \ldots, N\}$ such that for any subset $\gamma \supset \gamma_m$ for which the matrix $A^{-1}[\gamma]$ is invertible

$$C_{\epsilon/K}\left(\left(A^{-1}[\gamma]\right)^{-1}\right) \geq m \text{ for some } K > 0. \tag{2.1.4}$$

Then (I) *implies* (II) *with*

$$K = \frac{C_m}{N}, \quad C_m = \frac{1}{m!\, 2^{m-1}}$$

Conversely, (II) *with $K = 1$ implies* (I)*.*

Remark. *The matrix $\left(A^{-1}[\gamma]\right)^{-1}$ coincides with the Schur complement of $A[\gamma^c]$ in A, see* (2.2.1) *below for details. Here $\gamma^c = \{1, 2, \ldots, N\} \setminus \gamma$.*

2.1.2 Application to random Schrödinger operators

In quantum physics, the tight-binding approximation is often used as the prototypical model for the study of electron propagation in solids. In this model, the evolution of the wave function ψ on the d-dimensional lattice $\mathbb{Z}^d$ is given by the Schrödinger equation

$$i\hbar\dot{\psi}_t = H\psi_t; \quad \psi(0) = \psi_0, \tag{2.1.5}$$

where the self-adjoint Hamiltonian H is a sum of the hopping term H_0 and the potential V, of the form

$$(H\psi)(x) = (H_0\psi)(x) + V(x)\psi(x), \quad x \in \mathbb{Z}^d.$$

In this work we consider the random operators that have this functional form. Let us list few of these:

Anderson model H_A One of the best-studied models for disordered solids was introduced by P. H. Anderson in [3]. In this model the Hilbert space is $\ell^2(\mathbb{Z}^d)$, the hopping term H_0 is the discrete Laplacian Δ, and the potential V in H above is of the form $V(x) = g\sum a_{x-y}v(y)$, where the single site potentials $v(y)$ are independent random variables. The real parameter g is a coupling constant which describes the strength of the disorder.

Alloy-type Anderson model H_{alloy} Here the Hilbert space is also $\ell^2(\mathbb{Z}^d)$, the hopping term H_0 is a short range ergodic operator. The value of the potential $V(x)$ at a site $x \in \mathbb{Z}^d$ is generated from independent random variables $\{u(y)\}$ via the transformation

$$V(x) = g\sum_{y\in\mathbf{\Gamma}} a_{x-y}u(y)\ ,$$

where the index y takes values in some sub-lattice $\mathbf{\Gamma}$ of $\mathbb{Z}^d$. The Hamiltonian H_A coincides with H_{alloy} provided $H_0 = \Delta$, $\mathbf{\Gamma} = \mathbb{Z}^d$; $a_z = \delta_{|z|}$, where δ_x is Kronecker delta function: $\delta_0 = 1$; $\delta_x = 0$ for $x \neq 0$. In general, the random potential at sites x, y is correlated for this model. As its name suggest, H_{alloy} is used to describe (random) alloys in the tight binding approximation.

Random block operator H_{block} The Hilbert space is $\ell^2(\mathbb{Z}^d; \mathbb{C}^k) \cong \left(\ell^2(\mathbb{Z}^d)\right)^k$ (the space of square-summable functions $\psi\colon \mathbb{Z}^d \to \mathbb{C}^k$). The kernel $H_0(x, y)$ of the hopping term is a deterministic, translation invariant $k \times k$ matrix. The random potential $V(x)$ at each site is an independently drawn random $k \times k$ Hermitian matrix multiplied by g.

Previous Results

Anderson [3] argued that in the $g \gg 1$ regime, the solution of the initial value problem (2.1.5) for H_A stays localized in space for all times almost surely if the initial wave packet ψ_0 is localized. Mathematical study of Anderson localization is an active field; we refer readers to the recent reviews [19, 28] on the subject for the detailed bibliography. In this work, we focus our attention on a single aspect of Anderson localization—the so-called m-level Wegner estimate.

Let $|S|$ denote the cardinality of the set S. Let H_A^Λ be a restriction of the operator H_A to a finite box Λ. Then the m-level Wegner estimate is an upper bound on the probability of n eigenvalues being in the same energy interval $I_\epsilon := (E - \epsilon, E + \epsilon)$:

$$\mathbb{P}\left(\mathcal{C}_\epsilon(H_A^\Lambda - E) \geq m\right) \leq C_m(|\Lambda|\epsilon)^m,$$

for random variables $v(x)$ with a bounded density. As such, it gives some measure of the correlation between multiple eigenvalues. We will refer to the 1-level bound simply as the Wegner estimate (first established by F. J. Wegner in [31]). It plays the instrumental role in the proof of Anderson localization.

If localization occurs in some energy interval $I \subset \mathbb{R}$, the entire spectrum of H_A in I is pure point. It is then natural to study the distribution of the eigenvalues for H_A^Λ in this interval. Physicists expect that there is no energy level repulsion for states in the localized regime: that is, the eigenvalues should be distributed independently on the interval I. The first rigorous result in this direction, namely that the point process associated with the (rescaled) eigenvalues converges to a Poisson process, was obtained by Molchanov [25] in the setting of a one-dimensional continuum.

Minami [24] established the analogous result for H_A under the assumption that the distribution of every $v(n)$ has a bounded density. The key component in [24] is the 2-level Wegner estimate, which is consequently known as the Minami estimate.

By now the localization phenomenon for the original Anderson model H_A is well understood. In particular, the general m-level Wegner estimate is known to hold for essentially all distributions μ of the random potential $v(x)$; see [5, 16]. We refer the reader to [9] for the state of the art results concerning eigenvalue counting inequalities for H_A. However, the current understanding of many (in fact almost all) other random models of interest remains partial at best.

The Wegner estimate for a special class of alloy-type Anderson model H_{alloy} was first established by Kirsch in [18]. By now it is known to hold in fair generality (albeit not universally). See the recent preprint [22] for the extensive bibliography on the subject. The Wegner estimate for a random block operator H_{block}—with $V(x) = gv(x)A$ where $v(x)$ are independent random variables and A is a fixed invertible Hermitian matrix—holds in perturbative regimes. That is, it holds near the edges of the spectrum, [7] and in the strong disorder regime $1 \ll g$, [13]. A weaker bound (weaker in terms of the volume dependence) near the edges of the spectrum was established for the Fröhlich model, where the matrix-valued potential is given by $V(x) = gU(x)^* AU(x)$, where A is a fixed self-adjoint $k \times k$ matrix, and the $U(x)$ are independently chosen according to the Haar measure on $SU(k)$, [6].

On the other hand, not much progress has been made on extensions of the multi-level Wegner estimate, besides allowing for more general background operators H_0 than the discrete Laplacian. In particular, apart from two special examples below, all previous works require a non-correlated random potential. In [21], this limitation was partially removed in the continuum one-dimensional setting, allowing for positively correlated randomness. In [10], the authors announced the establishment of the Minami estimate and subsequently Poisson statistics for a general class of positively correlated random potentials. Unfortunately, although [10] contains a new elegant and efficient proof of Minami's estimate for H_A, its extension to the generalized setting has a significant gap, which so far has not been removed. Finally, let us mention the recent result [29], which established the Minami estimate for a special class of weakly correlated randomness for which one can transform the problem to the uncorrelated one.

The reader may wonder about the glaring disparity between the wealth of results on the 1-level Wegner estimate and the scarcity of results for its many-level counterparts. The reason can be traced to the direct (and frequently exploited) link between the former and the underlying Green function given by (2.1.1). The amenable nature of the Green function then allows one to establish a robust 1-level Wegner estimate in many situations of interest. In the present work, we harness the connection between the many-level Wegner estimate and the Green function given by Theorem 2.1.1 to establish an m-level Wegner estimate for a certain class of models with correlated randomness. Roughly speaking, our method works if the randomness in the system is sufficiently rich. (We will quantify this statement in the sequel.)

Although in most known applications (such as localization, simplicity of the spectrum, and Poisson statistics of eigenvalues) one is interested in the 1- and 2-level Wegner estimates, it is nonetheless natural from a mathematical perspective to investigate the general many-level case. From a practical perspective, it can yield some insight on the nonlinear Anderson model via multi-state resonance phenomenon, [14].

The blessing and the curse of the existing methods employed in proof of the Minami estimate (with the single exception of [21]) is that the nature of the background operator H_0 plays little if any role in the proofs. It is however clear that in the case of the correlated random potential in H_A one cannot hope to get the Minami estimate without exploiting the structure of H_0. Indeed, consider the one dimensional operator H_{alloy} with $H_0 = 0$, and the random potential at odd sites being i.i.d. random variables, while $v(2n) = v(2n-1)$. Its spectrum consists of (the closure of) the set of eigenvalues $\{\lambda_n\} = \{v(n)\}$, each one being degenerate. Consequently, even though H_{alloy} in this setting is perfectly localized, the probability of finding two closely lying eigenvalues is equal to 1.

The m-level Wegner estimate for the random block model H_{block}

We will consider the class of random block models H_{block} introduced earlier.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph with degree at most κ, such that the set of vertices (sites) $\mathcal{V}$ is finite with cardinality N. The main example of this model is the restriction of the lattice $\mathbb{Z}^d$ to the box, but

the greater generality does not require additional effort here.

Let $(\Omega, \mathbb{P})$ be a probability space. Let $\{V(x) = gA_\omega(x) : \ x \in \mathcal{V}, \omega \in \Omega\}$ be a collection of independent, identically distributed random $k \times k$ Hermitian matrices.

Basic Assumption We now state the main technical condition that we will use as an input for our application theorem below.

(A) *For an integer n, let S be a given set of $2nk$ distinct integers. Then the matrix $A_\omega(x) - a$ is invertible for all $a \in S$ and all $\omega \in \Omega$. Moreover, there exists an $\alpha > 0$ such that, for any integer $a \in S$, any $\epsilon \in [0, 1]$ and arbitrary Hermitian $k \times k$ matrix J the bound*

$$\mathbb{P}\left(\left|\det\left((A_\omega(x) - a)^{-1} + (J + a)^{-1}\right)\right| \leq \epsilon\right) \leq K\epsilon^\alpha \tag{2.1.6}$$

holds.

It guarantees that the randomness in the system is rich enough to imply the result below (Theorem 2.1.3). At the first glance, a more natural condition should concern the properties of the matrix $A_\omega(x) + J$ as it is the correct functional form of the corresponding Schur complement of H_ω (see Section 2.3 for details). However, this turns out to be an unsuitable choice because of the absence of an *a priori* bound on the norm of the background operator J (which encodes the information about the environment of the x-block in H_ω). On the other hand, for a sufficiently large set of numbers $\{a_i\}$ one can ensure that regardless of the norm of J, one of the matrices $\{(J + a_i)^{-1}\}$ is bounded in norm by 1 (see Proposition 2.3.2). We then exploit the fact that matrices $(A - a)^{-1} + (J + a)^{-1}$ (which appears in (3.1.2)) and $A + J$ are related:

$$(A + J)^{-1} = (A - a)^{-1} - (A - a)^{-1}\left((A - a)^{-1} + (J + a)^{-1}\right)^{-1}(A - a)^{-1}, \tag{2.1.7}$$

provided that $A - a$ and $J + a$ are invertible. One can readily verify (2.1.7) (which is in fact a particular case of Woodbury's matrix identity) by multiplying both sides by $A + J$.

We now introduce our single-particle Hamiltonian. Namely, let $H_\omega(g)$ be a random block operator H_{block} acting on $\ell^2(\mathcal{V}; \mathbb{C}^k)$ (the space of square-summable functions $\psi : \ \mathcal{V} \to \mathbb{C}^k$) as

$$(H_\omega(g)\psi)(x) = (H_0\psi)(x) + gA_\omega(x)\psi(x), \tag{2.1.8}$$

where $g > 0$ is a coupling constant, H_0 is an arbitrary deterministic self-adjoint operator on $\ell^2(\mathcal{V}; \mathbb{C}^k)$, and $A_\omega(x)$ is an independently drawn random $k \times k$ Hermitian matrix as above. We use the notation $H_\omega(g)$ instead of H_{block} to stress the random nature of this operator as well as the dependence on the parameter g.

Theorem 2.1.3. *Assume* **(A)**. *Then*

I. For any $E \in \mathbb{R}$ the operator $H_\omega(g) - E$ is almost surely invertible.

II. Moreover, there exist $\epsilon_0 > 0$ and $C > 0$ (which depend only on k, m, α) for which we have

$$\mathbb{P}\left(\mathcal{C}_\epsilon(H_\omega(g) - E) \geq m\right) \leq C\left|\ln(N\epsilon/g)(N\epsilon/g)^\alpha\right|^m \tag{2.1.9}$$

for any $E \in \mathbb{R}$, for any $\epsilon \in [0, \epsilon_0]$ and for all $m \leq n$. In the $m = 1$ case we can improve the above bound to

$$\mathbb{P}\left(\mathcal{C}_\epsilon(H_g - E) \geq 1\right) \leq C(N\epsilon/g)^\alpha. \tag{2.1.10}$$

Examples

• Anderson model H_A. As we mentioned earlier, the nontrivial Minami estimate is well understood only for the original Anderson model among all alloy-type models. It is therefore a litmus test to verify Assumption **(A)** for H_A.

Theorem 2.1.4. *Suppose that the distribution μ of the $v(x)$ variables in H_A is compactly supported on the interval $I = [-b, b]$ for some $b > 0$ and is β-regular, i.e. for any Lebesgue - measurable $S \subset I$ we have*

$$\mu(S) \leq C|S|^\beta.$$

Then Assumption **(A)** *holds with $\alpha = \beta$.*

Our approach to the Minami estimate is also meaningful for the $\mathbf{\Gamma}$-trimmed Anderson model introduced in [11], near the edges of the spectrum, in the sense that the assumption **(A)** can be verified for it.

• Fröhlich model and alloy type Anderson model H_{alloy}. Assumption **(A)** is either not satisfied for a single site x or is satisfied with a power α which is too small to make the result meaningful. However, the close inspection of the proof of Theorem 2.1.3 shows that the matrix J that appears in Assumption **(A)** is not required to be completely arbitrary. In fact, the relevant matrices J carry the structure of the Schrödinger operator (with arbitrary boundary conditions). It seems plausible (and is on our to-do list) that Assumption **(A)** can be verified for such J and sets of sites that include x and its neighbors.

• The third model arises from the study of dirty superconductors via the Bogoliubov - de Gennes equation. After a suitable change of the coordinate basis, the Bogoliubov-de Gennes (BdG) model above can be described in terms of the operator defined in (3.1.1), with $A_\omega(x) = \sigma_\omega(x) \in M_{2\times 2}$, where $\sigma_\omega(x)$ is a random Pauli matrix of the form

$$\sigma_\omega(x) = \begin{bmatrix} u_x & v_x \\ v_x & -u_x \end{bmatrix}; \quad u_x,\, v_x \text{ are random variables.} \tag{2.1.11}$$

The m-Wegner estimate for these models has only been established for $m = 1$ case, for a restricted class of joint distributions of u, v variables (absolutely continuous and with support in a half plane), and for a specific background operator H_0 in [20, 15]. We establish the robust Wegner and Minami estimates for this model.

Theorem 2.1.5. *Let each $A_\omega(x)$ be given by* (2.1.11)*. Suppose that the joint distribution μ of the u, v variables is supported on a unit disc O and is β-regular, i.e. for any Lebesgue - measurable $S \subset O$ we have*

$$\mu(S) \leq C|S|^\beta.$$

Then Assumption **(A)** *holds with $\alpha = \beta$.*

2.1.3 Paper's organization

We prove our main abstract result, Theorem 2.1.1, along with its corollary, in Section 2.2. We prove our result on eigenvalue estimates, Theorem 2.1.3, in Section 2.3. We consider the implication of the latter result for the random block operators in Section 2.4. These proofs depend on a number of auxiliary results, which we prove in Section 2.5.

2.2 Proof of Theorem 2.1.1 and Corollary 2.1.2

2.2.1 Notation

Let n be a positive integer, and let α and β be index sets, i.e., subsets of $\{1, 2, \ldots, n\}$. We denote the cardinality of an index set by $|\alpha|$ and its complement by $\alpha^c = \{1, 2, \ldots, n\} \setminus \alpha$. For an $n \times n$ matrix A, let $A[\alpha, \beta]$ denote the submatrix of A with rows indexed by α and columns indexed by β, both of which are thought of as increasing, ordered sequences, so that the rows and columns of the submatrix appear in their natural order. We will write $A[\alpha]$ for $A[\alpha, \alpha]$. If $|\alpha| = |\beta|$ and if $A[\alpha, \beta]$ is nonsingular, we denote by $A/A[\alpha, \beta]$ the Schur complement of $A[\alpha, \beta]$ in A, [32]:

$$A/A[\alpha, \beta] = A[\alpha^c, \beta^c] - A[\alpha^c, \beta] \left(A[\alpha, \beta]\right)^{-1} A[\alpha, \beta^c]. \tag{2.2.1}$$

We will frequently use Schur's complementation and its consequences in this work; we refer the reader to the comprehensive book [32] on this topic.

For a Hermitian matrix A and a positive number a we will write

$$\mathcal{B}_a(A) := |\sigma(A) \cap [a, \infty)|.$$

Let $P_\epsilon(A)$ denote the spectral projection of the Hermitian matrix A onto the interval $(-\epsilon, \epsilon)$ for $\epsilon > 0$.

2.2.2 Proof of Theorem 2.1.1

Suppose that (I) holds. We will use the following assertion:

Proposition 2.2.1. *Let A be an $N \times N$ positive definite matrix, and suppose that $\mathcal{B}_a(A) = k$ for some $a > 0$. Then there exists an index subset $\alpha_k = \{i_1, i_2, \ldots, i_k\}$ of $\{1, \ldots, N\}$ such that $A[\alpha_k] \geq \frac{a}{k!\, 2^{k-1}\, N} I_N[\alpha_k]$.*

By Proposition 2.2.1 there exists α_m such that

$$P_\epsilon(A)[\alpha_m] \geq \frac{C_m}{N} I_N[\alpha_m].$$

with $C_m = \frac{1}{k!\, 2^{k-1}}$. Combining this bound with

$$A^{-2} > \frac{1}{\epsilon^2} P_\epsilon(A)$$

we obtain

$$A^{-2}[\alpha_m] > \frac{C_m}{N\epsilon^2}\, I_N[\alpha_m].$$

Since $\sigma(TT^*) \setminus \{0\} = \sigma(T^*T) \setminus \{0\}$ for any operator T, we deduce from the previous equation (with $T = I_N[\alpha_m, \alpha_N]A^{-1}$) that there exists an orthogonal projection Q of rank m such that

$$A^{-1} I_N[\alpha_m] A^{-1} > \frac{C_m}{N\epsilon^2}\, Q.$$

Applying now Proposition 2.2.1 once again, we conclude that there exists β_m such that (2.1.2) holds with K given by (2.1.3).

Conversely, suppose that (2.1.2) holds with $K = 1$. Since

$$A^{-2}[\alpha_m] \geq A^{-1}[\alpha_m, \beta_m] A^{-1}[\beta_m, \alpha_m],$$

the assertion follows from the Cauchy interlacing theorem for the Hermitian matrix A^{-2} and its principal submatrix $A^{-2}[\alpha_m]$. □

Proof of Proposition 2.2.1. The proof will proceed by induction in k. If $k = 1$, the result follows from the fact that A is positive, so tr $A = \sum_{\lambda \in \sigma(A)} \lambda \geq a$. Since the trace is at least a, there exists a diagonal entry which is greater than or equal to $\frac{a}{N}$.

Suppose we have established the induction hypothesis for $k = K$. We want to verify the induction step, i.e. the case $k = K + 1$. To this end, choose the index i_1 so that $A_{i_1 i_1} \geq A_{ii}$ for all i. Without loss of generality, let us assume that $i_1 = 1$. Then A is of the block form

$$A = \begin{bmatrix} A_{11} & u \\ u^* & B \end{bmatrix}.$$

Consider now the matrix $D = A/A_{11}$. It is positive definite by the Schur complement condition for positive definiteness (as A is positive definite). Also, the matrix

$$F = \begin{bmatrix} A_{11} & u \\ u^* & B - D \end{bmatrix}$$

is rank one (since $F/A_{11} = 0$), so by the rank one perturbation theory, $\mathcal{B}_a(A - F) \geq K$. But $\mathcal{B}_a(A - F) = \mathcal{B}_a(D)$. Using the induction hypothesis, we conclude that there exists an index set $\alpha_K = \{i_2, \ldots, i_K\}$ with $1 \notin \alpha_K$ such that $D[\alpha_K] \geq \frac{a}{K!\,2^{K-1}N} I_N[\alpha_K]$.

The induction step (with $\alpha_{K+1} = \alpha_K \cup \{1\}$) now follows from the following assertion:

Lemma 2.2.2. *Let A be an $l \times l$ positive definite matrix of the block form*

$$A = \begin{bmatrix} A_{11} & u \\ u^* & B \end{bmatrix}. \tag{2.2.2}$$

Suppose that in addition $A_{11} \geq A_{ii}$ for all $i \in \{1, \ldots, l\}$, and $A/A_{11} \geq a$ for some $a > 0$. Then $A \geq \frac{a}{2l}$.

□

Proof of Lemma 2.2.2. To show that $A - \frac{a}{2l} \geq 0$ it suffices to check (by the Schur complement condition for positive definiteness) that

$$A_{11} - \frac{a}{2l} \geq 0; \quad (A - \frac{a}{2l})/(A_{11} - \frac{a}{2l}) \geq 0. \tag{2.2.3}$$

Since $A/A_{11} \geq a$, we have $A_{ii} \geq a$ for all $i \geq 2$ as a is positive, so by assumption of the lemma $A_{11} \geq a$ as well (and hence we have established the first bound in (2.2.3)). Next we write

$$\begin{aligned} (A - \frac{a}{2l})/(A_{11} - \frac{a}{2l}) &= B - \frac{a}{2l} - \frac{u^*u}{A_{11} - \frac{a}{2l}} \\ = \quad \left(B - \frac{u^*u}{A_{11}}\right) - \frac{a}{2l} - \frac{au^*u}{2lA_{11}(A_{11} - a/2l)} \\ \geq \quad a - \frac{a}{2l} - a\frac{u^*u}{l(A_{11})^2}. \end{aligned} \tag{2.2.4}$$

Now observe that since A is positive, the contraction $A[\{1, i+1\}]$ is also positive for all i, and in particular $\det A[\{1, i+1\}] = A_{11}B_{ii} - |u_i|^2 \geq 0$. But $A_{11} \geq B_{ii}$ for all i, hence $|u_i|^2/(A_{11})^2 \leq 1$. We therefore can estimate

$$\|u^*u\| = \|u\|^2 \leq (l-1)(A_{11})^2.$$

Substitution of this estimate into (2.2.4) yields the second bound in (2.2.3). □

2.2.3 Proof of Corollary 2.1.2

We first observe that if sets of indices α, β satisfy $\alpha \subset \beta$, then $A^{-1}[\alpha]$ is a principal submatrix of $A^{-1}[\beta]$, and we have

$$C_\epsilon\left(\left(A^{-1}[\beta]\right)^{-1}\right) \geq C_\epsilon\left(\left(A^{-1}[\alpha]\right)^{-1}\right) \tag{2.2.5}$$

by the Cauchy interlacing theorem, provided the matrices $A^{-1}[\alpha], A^{-1}[\beta]$ are invertible. Therefore, it suffices to establish the corollary for the smallest set γ_{min} that contains γ_m and for which $A^{-1}[\gamma_{min}]$ is invertible. Without loss of generality we will assume that $\gamma_{min} = \gamma_m$.

Suppose that (I) holds. Then the assertion (II) of Theorem 2.1.1 holds with K given by (2.1.3). Construct now the set $\gamma_m = \alpha_m \cup \beta_m$ with α_m, β_m from the assertion (II) of Theorem 2.1.1. Let us consider the matrix

$$B := \left(A^{-1}[\gamma_m]\right)^{-1}.$$

Since $A^{-1}[\alpha_m, \beta_m]$ is a submatrix of $A^{-1}[\gamma_m]$ we see that the condition (II) of Theorem 2.1.1 is fulfilled for B. Hence we can apply Theorem 2.1.1 to B to conclude that $C_{\epsilon/K}(B) \geq m$.

Conversely, suppose that (2.1.4) holds with $K = 1$. Then (I) holds as well, as follows from (2.2.5) with $\alpha = \gamma_m$, $\beta = \{1, \ldots, N\}$. □

2.3 Proof of Theorem 2.1.3

We first observe that by scaling it suffices to prove the result for the $g = 1$ case. We will use the shorthand notation H_ω instead of $H_\omega(1)$ in the sequel.

Next, we prove the first assertion of the theorem, using induction in N. To initiate the induction, we consider the case $N = 1$, so that $H_\omega = A_\omega(x) + K$, where K is a deterministic Hermitian matrix. It follows from Assumption **(A)** and (2.1.7) that $H_\omega - E$ is invertible almost surely.

Suppose now that the induction hypothesis holds, i.e. the matrix $H_\omega - E$ is almost surely invertible for $N \leq M$ and all E. We want to establish the induction step ($N = M + 1$ case). To this end, let $\hat{\mathcal{V}}$ be any subset of $\mathcal{V}$ of cardinality M, and let $\hat{H}_\omega$ be a restriction of H_ω to $\hat{\mathcal{V}}$. By the induction hypothesis, $\hat{H}_\omega - E$ is invertible almost surely for all E. Let us consider some configuration ω for which $\hat{H}_\omega - E$ is invertible. Then $H_\omega - E$ is invertible if and only if the Schur complement of $\hat{H}_\omega - E$ in $H_\omega - E$, i.e. $(H_\omega - E)/(\hat{H}_\omega - E)$, is invertible, [32]. But $(H_\omega - E)/(\hat{H}_\omega - E)$ is a Hermitian $k \times k$ matrix of the form $A_\omega(x) + J$, where $\{x\} = \mathcal{V} \setminus \hat{\mathcal{V}}$, and J is a matrix independent of the randomness in $A_\omega(x)$. It follows by the same argument as in the $N = 1$ case that $(H_\omega - E)/(\hat{H}_\omega - E)$ is invertible for almost all values of the randomness in $A_\omega(x)$.

We now prove the second assertion of the theorem. We will only consider configurations ω in Ω such that $H_\omega - E$ is invertible (for the remaining set of configurations has measure zero by the first

assertion).

For the random operator T_ω, let $\mathcal{E}_\epsilon(T_\omega)$ be the event $\{\omega : \mathcal{C}_\epsilon(T_\omega) \geq m\}$. With this notation, we wish to estimate the size of the set $\mathcal{E}_\epsilon(H_\omega - E)$. If we enumerate the vertices $v \in \mathcal{V}$, we can think of H_ω as a $kN \times kN$ Hermitian matrix with a block form, i.e. the indices $\{lk - k + 1, \ldots, lk\}$ correspond to the vertex l in $\mathcal{V}$, with $l = 1, \ldots, N$.

Size reduction We first reduce the dimensionality of the original problem using Corollary 2.1.2. This assertion gives us the existence of the index subset γ_m with $|\gamma_m| = 2m$ such that inclusion

$$\mathcal{E}_\epsilon(H_\omega - E) \subset \mathcal{E}_{\epsilon/K}\left(\left((H_\omega - E)^{-1}[\gamma]\right)^{-1}\right) \tag{2.3.1}$$

holds for any index set $\gamma \supset \gamma_m$ for which $(H_\omega - E)^{-1}[\gamma]$ is invertible, with K given by (2.1.3). (To be precise, the matrix size N in that corollary gets replaced by kN.)

In general, the submatrix $(H_\omega - E)^{-1}[\gamma]$ can be a complicated object, so it is not immediately clear that such a reduction is helpful. However, if the set γ happens to consist of the indices that agree with the block structure of H_ω, something interesting happen. More precisely, suppose that $i \in \gamma \Rightarrow (j \in \gamma$ for any j with $\lfloor j/k \rfloor = \lfloor i/k \rfloor)$, where $\lfloor \cdot \rfloor$ is the floor function. In this case we can associate γ with a subset $\mathcal{V}'$ of the original vertex set $\mathcal{V}$. Then the submatrix $((H_\omega - E)^{-1}[\gamma])^{-1}$ retains the same block form as H_ω, in the following sense: If we go back to the vertex representation for $((H_\omega - E)^{-1}[\gamma])^{-1}$ (which is possible due to the special form of the set γ), then for any $\psi \in \ell^2(\mathcal{V}'; \mathbb{C}^k)$ and any $x \in \mathcal{V}'$ we have

$$\left(\left((H_\omega - E)^{-1}[\gamma]\right)^{-1}\psi\right)(x) = (T_0\psi)(x) + A_\omega(x)\psi(x). \tag{2.3.2}$$

This can be seen from the fact that the matrix $((H_\omega - E)^{-1}[\gamma])^{-1}$ coincides with the Schur complement of $(H_\omega - E)[\gamma^c]$ in $H_\omega - E$,

$$\left((H_\omega - E)^{-1}[\gamma]\right)^{-1} = (H_\omega - E)/(H_\omega - E)[\gamma^c].$$

It is important to note that the operator T_0 in (2.3.2) is independent of the randomness associated with matrices $\{A_\omega(x)\}_{x\in\mathcal{V}'}$ (though it does depend on the other random variables). We also note that the matrix $((H_\omega - E)^{-1}[\gamma])$ is almost surely invertible (as follows from the first part of the theorem).

Combining these observations, we conclude that it is beneficial (and sufficient) to consider the sets γ in (2.3.1) that respect the block structure of H_ω and therefore contain up to $2km$ indices (i.e. up to $2m$ vertices in $\mathcal{V}'$, as in Corollary 2.1.2, and exactly k indices per vertex, to preserve blocks). Thus, we have obtained the following intermediate result:

Lemma 2.3.1. *Suppose that the second assertion of Theorem 2.1.3 holds for all $N \leq 2m$. Then it holds for any N.*

Norm reduction The deterministic part of H_ω—namely the operator H_0—can be arbitrary large in norm (even if $\|H_0\| \leq C$ for the original H_ω, the size reduction process indicated above creates a new background operator T_0 with uncontrollable norm). Our next step in the proof will require that the background operator is bounded in norm by a constant, say by $1/2$. We achieve this by means of the following transformation.

Proposition 2.3.2. *Let $B_{1,2}$ be a pair of Hermitian $L \times L$ matrices with $\|B_1\| \leq 1$. Consider the matrices*

$$B = B_1 + B_2, \quad \hat{B} = (B_1 - aI_L)^{-1} + (B_2 + aI_L)^{-1}$$

where $a \in \mathbb{R}$. Then there exists an integer $a \in [-L-3, -3] \cup [3, L+3]$ (which depends on B_2 but not on B_1) and $\epsilon_0 > 0$ (which depends only on L) such that for any $\epsilon < \epsilon_0$

$$\max\left(\left\|(B_1 - aI_L)^{-1}\right\|, \left\|(B_2 + aI_L)^{-1}\right\|\right) \leq \frac{1}{2}; \tag{2.3.3}$$

$$\mathcal{C}_{\epsilon/(225L^4)}\left(\hat{B}\right) \leq \mathcal{C}_\epsilon(B) \leq \mathcal{C}_{7L^2\epsilon}\left(\hat{B}\right). \tag{2.3.4}$$

We will apply this proposition to the operator $H_\omega - E$ by choosing $B_2 = H_0 - E$, $B_1 = H_\omega - H_0$. By the hypothesis of Theorem 2.1.3, the assumptions of Proposition 2.3.2 are satisfied, with $L = kN$. Combining this observation with the size reduction, we obtain the second intermediate result.

Lemma 2.3.3. *Assume* **(A)**. *Let $\hat{H}_\omega$ be an operator acting on $\ell^2(\mathcal{V}; \mathbb{C}^k)$ as*

$$(\hat{H}_\omega \psi)(x) = (H_0\psi)(x) + (A_\omega(x) - a)^{-1}\psi(x). \tag{2.3.5}$$

Suppose that $\|H_0\| \leq 1/2$. If there exist $\epsilon_0 > 0$ and $b > 0$ (which depend on k, m, α) so that for all integers $a \in [-km-3, -3] \cup [3, km+3]$ and all $N \leq 2m$ the bound

$$\mathbb{P}\left(\mathcal{C}_{7k^2m^2\epsilon}(\hat{H}_\omega) \geq m\right) \leq b|\ln\epsilon|^m\epsilon^\alpha$$

holds uniformly in H_0 and $\epsilon < \epsilon_0$, then the second assertion of Theorem 2.1.3 holds.

Reduction to the determinant Suppose that $\mathcal{C}_{7k^2m^2\epsilon}(\hat{H}_\omega) \geq m$. Since by construction $\|\hat{H}_\omega\| \leq 1$, the operator $\hat{H}_\omega$ can have no more than $kN - m$ large eigenvalues, and each of these can have an absolute value no larger than 1. As a result, we obtain the bound:

$$|\det \hat{H}_\omega| \leq (7k^2m^2\epsilon)^m. \tag{2.3.6}$$

We may now employ the following lemma to calculate the probability of the aforementioned bound on the determinant. Its proof can be found in Section 2.5.

Lemma 2.3.4. *Assume* **(A)**. *Let $\hat{H}_\omega$ be as in* (2.3.5)*, and let $\hat{\mathcal{E}}_\delta$ be the event*

$$\hat{\mathcal{E}}_\delta = \{\omega \in \Omega : \det(\hat{H}_\omega) \leq \delta\}.$$

Let

$$\delta_0 := \exp\left(2K\alpha^{1+1/N}\right).$$

Then for any $\delta \in [0, \delta_0]$ we have

$$\mathbb{P}(\hat{\mathcal{E}}_\delta) \leq (2K\alpha)^N \ln^N(\delta^{-1})\delta^\alpha. \tag{2.3.7}$$

Using this result in conjunction with (2.3.6) we obtain the there exist $\epsilon_0 > 0$ and $b > 0$ that depend on k, m, α so that

$$\mathbb{P}\left(\mathcal{C}_{7k^2m^2\epsilon}(\hat{H}_\omega) \geq m\right) \leq b|\ln\epsilon|^m\epsilon^{m\alpha}, \tag{2.3.8}$$

for $N \leq m$ and for $\epsilon < \epsilon_0$. The combination of Lemma 2.3.3 and (2.3.8) yields (3.1.3).

Improvement on the Wegner bound We want to improve on this bound for the special case that $m = 1$. In this case we need to verify the (improved) input for Lemma 2.3.1 for $N = 1$ and $N = 2$. In the former case, the bound (3.1.4) follows from **(A)** and Proposition 2.3.2 (where we choose $B_2 = H_0 - E$, $B_1 = H_\omega - H_0$). So for the rest of the argument we will assume that $N = 2$.

Let $\mathcal{E}_\epsilon$, $\mathcal{S}_\epsilon$ be the events

$$\begin{aligned} \mathcal{E}_\epsilon &= \{\omega : \mathcal{C}_\epsilon(H_\omega - E) \geq 1\}; \\ \mathcal{S}_\epsilon &= \{\omega : \mathcal{C}_{\epsilon^{2/3}}(H_\omega - E) \geq 2\}. \end{aligned}$$

We first observe that it follows from (3.1.3) (which we already established earlier) that

$$\mathbb{P}(\mathcal{E}_\epsilon \cap \mathcal{S}_\epsilon) \leq \mathbb{P}(\mathcal{S}_\epsilon) \leq C\left|\ln(\epsilon^{2\alpha/3})\epsilon^{2\alpha/3}\right|^2 \leq C\epsilon^\alpha$$

for ϵ sufficiently small. Therefore, to get (3.1.4) it suffices to show that $\mathbb{P}(\mathcal{E}_\epsilon \setminus \mathcal{S}_\epsilon) \leq C\epsilon^\alpha$. To this end, suppose that $\omega \in \mathcal{E}_\epsilon \setminus \mathcal{S}_\epsilon$. Then

$$(H_\omega - E + \epsilon)^{-1} + 2\epsilon^{-2/3} > 0; \;\; \left\|(H_\omega - E + \epsilon)^{-1} + 2\epsilon^{-2/3}\right\| \geq \frac{1}{2\epsilon}. \tag{2.3.9}$$

If $\mathcal{V} = \{x, y\}$, let us denote by P_x (P_y) the rank k projection onto the site x (accordingly y). The positivity of the left-hand side can be exploited by means of Lemma 2.3.5 below with choices $P_1 = P_x$, $P_2 = P_y$.

Lemma 2.3.5. *Let $A > 0$, and let $P_{1,2}$ be orthogonal projections that satisfy $P_1P_2 = 0$. Let $P = P_1 + P_2$. Then we have*

$$\|PAP\| \leq 2\max(\|P_1AP_1\|, \|P_2AP_2\|). \tag{2.3.10}$$

Using (2.3.9) and (2.3.10), we infer that $\omega \in \mathcal{R}_\epsilon$ (and thus $\mathcal{E}_\epsilon \setminus \mathcal{S}_\epsilon \subset \mathcal{R}_\epsilon$), where

$$\mathcal{R}_\epsilon = \left\{ \omega : \max_{i=x,y} \left(\left\| P_i (H_\omega - E + \epsilon)^{-1} P_i \right\| + 2\epsilon^{-2/3} \right) \geq \frac{1}{4\epsilon} \right\}.$$

But

$$P_x (H_\omega - E + \epsilon)^{-1} P_x = \Big((H_\omega - E + \epsilon) / P_x (H_\omega - E + \epsilon) P_x \Big)^{-1} = (A_\omega(x) + J)^{-1}$$

by the block inversion formula. Here the operator J depends on $A_\omega(y)$ but not on $A_\omega(x)$. Hence we can deduce from **(A)** and Proposition 2.3.2 that

$$\mathbb{P}\left(\left\| P_x (H_\omega - E + \epsilon)^{-1} P_x \right\| \geq \frac{1}{5\epsilon} \right) \leq \tilde{C}\epsilon,$$

with $\tilde{C}$ that depends on k, α but not on ϵ. The same bound holds with P_x replaced by P_y. Hence we infer that for ϵ small enough

$$\mathbb{P}(\mathcal{R}_\epsilon) \leq C\epsilon^\alpha,$$

and the result follows. □

Proof of Lemma 2.3.5. Let $A_1 = P_1 A P_1$, $A_2 = P_2 A P_2$, and $A_{12} = P_1 A P_2$. Then by Schur complement condition for positive definiteness

$$A_2 > 0\,; \quad A_1 \geq A_{12} A_2^{-1} A_{21}.$$

Since A_2 is positive, $A_2^{-1} \geq 1/\|A_2\|$, hence

$$A_1 \geq A_{12} A_{21} / \|A_2\|\,,$$

and so

$$\|A_1\| \|A_2\| \geq \|A_{12}\|^2\,,$$

where in the last step we have used $A_{12} = A_{12}^*$. Since

$$\|PAP\| \leq \max(\|A_1\| + \|A_{12}\|, \|A_2\| + \|A_{21}\|)\,,$$

the result follows. □

2.4 Proof of Theorems 2.1.4 and 2.1.5

2.4.1 Proof of Theorem 2.1.4

Let a be an integer that satisfies $a - b \geq 2$, and let j be arbitrary fixed real number. Then if $\epsilon \in [0, 1/(2a)]$, the inequality

$$\left| \frac{1}{v_x - a} + \frac{1}{j + a} \right| < \epsilon \tag{2.4.1}$$

for v_x has solutions in I only if $0 < j + a < 1$. Since equation

$$\left| \frac{1}{v_x - a} + \frac{1}{j + a} \right| = \epsilon$$

define the pair of points

$$v_x = a + \frac{1}{(j + a)^{-1} \pm \epsilon},$$

the set of v_x for which (2.4.1) holds is the interval $\hat{I}$ of length

$$|\hat{I}| = \frac{1}{(j + a)^{-1} - \epsilon} - \frac{1}{(j + a)^{-1} + \epsilon} = \frac{2\epsilon}{(j + a)^{-2} - \epsilon^2} < 4\epsilon,$$

where in the last step we used $0 < a + j < 1$, $\epsilon < 1/(2a) < 1/4$. Hence

$$\mathbb{P}\left(\left| \frac{1}{v_x - a} + \frac{1}{j + a} \right| < \epsilon \right) \leq C\epsilon^\beta$$

by the hypothesis of the theorem. □

2.4.2 Proof of Theorem 2.1.5

We first establish the bounds

$$\mathbb{P}(|\det(\sigma_\omega(x) + J)| \leq \epsilon) \leq C\epsilon^\beta; \tag{2.4.2}$$

$$\mathbb{P}(\mathcal{C}_\epsilon(\sigma_\omega(x) + J) \neq 0) \leq C\epsilon^\beta \tag{2.4.3}$$

for $\epsilon \in [0, 1]$. Indeed, note that $\det(\sigma_\omega(x) + J) = c^2 - (u_x - a)^2 - (v_x - b)^2$ with some constants a, b, c originating from J. Therefore the set $|\det(\sigma_\omega(x) + J)| \leq \epsilon$ is an intersection I of the disc O with the annulus centered at a, b and with radii $R_- = \sqrt{\max(c^2 - \epsilon, 0)}$, $R_+ = \sqrt{c^2 + \epsilon}$. The area of this set therefore cannot exceed $\pi(R_+^2 - R_-^2) \leq 2\pi\epsilon$ and (2.4.2) follows. To establish (2.4.3), we note that

$$\mathbb{P}(\mathcal{C}_\epsilon(\sigma_\omega(x) + J) \neq 0) = \mathbb{P}(\|(\sigma_\omega(x) + J)^{-1}\| \geq 1/\epsilon).$$

The value of $\|(\sigma_\omega(x) + J)^{-1}\|$ however can be evaluated explicitly and is given by

$$\|(\sigma_\omega(x) + J)^{-1}\| = \left| |c| - \sqrt{(u_x - a)^2 + (v_x - b)^2} \right|^{-1},$$

with the same constants a, b, c as before. Hence the set of the points in O that satisfy $\|(\sigma_\omega(x) + J)^{-1}\| \geq 1/\epsilon$ is an intersection $\hat{I}$ of the disc O with the annulus centered at a, b and with radii $R_- = ||c| - \epsilon|$, $R_+ = |c| + \epsilon$. The area of $\hat{I}$ cannot exceed the circumference of the unit circle times the maximal thickness 2ϵ of the annulus, so $|\hat{I}| \leq 4\pi\epsilon$, and (2.4.3) follows.

The assertion of the theorem follows now from Lemma 2.4.1 below (whose proof can be found in Section 2.5) and bounds (2.4.2) - (2.4.3).

□

Lemma 2.4.1. *Let A and J be Hermitian $k \times k$ matrices, and let a be some real number that satisfies $|a| \geq 2$. If $\|A\| \leq 1$ and*

$$\left|\det\left((A-a)^{-1}+(J+a)^{-1}\right)\right| \leq \frac{1}{16|a|}\left\{\frac{|a|-1}{2(|a|+1)^2}\right\}^{k-1}$$

then we have

$$\left(2(|a|+1)^2\right)^{-k}\left|\det\left(A+J\right)\right| \leq \left|\det\left((A-a)^{-1}+(J+a)^{-1}\right)\right|$$

or

$$\frac{1}{16|a|}\left\{2(|a|+1)^2\left\|(A+J)^{-1}\right\|\right\}^{1-k} \leq \left|\det\left((A-a)^{-1}+(J+a)^{-1}\right)\right|.$$

2.5 Proofs

Proof of Proposition 2.3.2. If $|a| \geq 3$, then since $||B_1|| \leq 1$, we have:

$$\nu := \left\|(B_1-a)^{-1}\right\|, \quad (L+4)^{-1} \leq \nu \leq 1/2\,. \tag{2.5.1}$$

Since B_2 is $L \times L$, it has at most L distinct eigenvalues. On the other hand, for every set S of real numbers with $|S| = L$ there exists an integer $a \in [-L-3,-3] \cup [3, L+3]$ so that $\mathrm{dist}(S,-a) \geq 2$, hence we can choose a that satisfies (2.3.3).

With this choice of a, consider the block matrix W of the form

$$W = \begin{bmatrix} (B_2+aI_n)^{-1} & (B_1-aI_n)^{-1} \\ (B_1-aI_n)^{-1} & -(B_1-aI_n)^{-1} \end{bmatrix}.$$

Note that the Schur complements to the upper and lower diagonal blocks are

$$\begin{aligned} -W/W_{11} &= (B_1-a)^{-1} + (B_1-a)^{-1}(B_2+a)(B_1-a)^{-1}\,; \\ W/W_{22} &= (B_2+a)^{-1} + (B_1-a)^{-1}\,. \end{aligned}$$

Let

$$T = (B_1-a)^{-1}/\nu, \tag{2.5.2}$$

where ν is given by (2.5.1).

In what follows, we will need two lemmas:

Lemma 2.5.1.

1. *Let $D = D^*, \tilde{D} = \tilde{D}^* \in M_{n,n}$. Then*

$$\mathcal{C}_\epsilon(D) \leq \mathcal{C}_{2\epsilon}(\tilde{D})\,,$$

provided $\|D - \tilde{D}\| \leq \epsilon$.

2. $\mathcal{C}_\epsilon(A) \leq \mathcal{C}_\epsilon(BAB)$ *whenever*

$$A = A^*, B = B^*, \|B\| \leq 1$$

Lemma 2.5.2. *Suppose $D = D^* \in M_{n,n}$ is of the form*

$$D = \begin{bmatrix} A & V \\ V^* & B \end{bmatrix},$$

with $A \in M_{k,k},\ B \in M_{m,m}$, $\|V\| \leq 1/2$, and

$$\mathcal{C}_{2\epsilon}(B) = 0\,. \tag{2.5.3}$$

Then

$$\mathcal{C}_\epsilon(D/B) \leq \mathcal{C}_\epsilon(D)\,; \tag{2.5.4}$$

$$\mathcal{C}_\epsilon(D) \leq \mathcal{C}_{\beta\epsilon}(D/B)\,, \tag{2.5.5}$$

with $\beta = 2(\|B^{-1}\| + 1)^2$.

Armed with these results, we can infer that

$$\begin{aligned}\mathcal{C}_\epsilon(B) = \mathcal{C}_{\nu^2\epsilon}\left(\nu^2 B\right) \leq \mathcal{C}_{\nu^2\epsilon}\left(\nu^2 TBT\right) &= \mathcal{C}_{\nu^2\epsilon}(W/W_{11}) \\ &\leq \mathcal{C}_{\nu^2\epsilon}(W) \leq \mathcal{C}_{\beta\nu^2\epsilon}(W/W_{22}) = \mathcal{C}_{\beta\nu^2\epsilon}\left(\hat{B}\right),\end{aligned} \tag{2.5.6}$$

where in the second step we have used Lemma 2.5.1 and in the remaining steps we have used Lemma 2.5.2. Here

$$\beta = 2\left(\left\|(W_{22})^{-1}\right\| + 1\right)^2 \leq 2(L+5)^2 \leq 25L^2$$

for $L \geq 2$. (It is straightforward to check that the relation $\mathcal{C}_\epsilon(B) \leq \mathcal{C}_{\beta\nu^2\epsilon}\left(\hat{B}\right)$ holds for $L = 1$.) Plugging in the upper bounds for ν, β we get the second inequality in (2.3.4):

$$\mathcal{C}_\epsilon(B) \leq \mathcal{C}_{7L^2\epsilon}\left(\hat{B}\right).$$

On the other hand, let

$$U = \kappa \begin{bmatrix} B_2 + aI_n & B_1 - aI_n \\ B_1 - aI_n & -B_1 + aI_n \end{bmatrix}; \quad \kappa = \frac{1}{2\|B_1 - a\|}\,.$$

Then

$$\begin{aligned} U/U_{22} &= \kappa B\,; \\ -U/U_{11} &= \kappa\left(B_1 - a + (B_1 - a)(B_2 + a)^{-1}(B_1 - a)\right). \end{aligned}$$

Similarly to (2.5.6), we obtain

$$\begin{aligned}\mathcal{C}_{\nu^2\epsilon/(25L^2)}\left(\hat{B}\right) &= \mathcal{C}_{\epsilon/(25L^2)}\left(\kappa^{-1}T\left(U/U_{11}\right)T\right)\\ &\leq \mathcal{C}_{\kappa\epsilon/(25L^2)}\left(U/U_{11}\right) \leq \mathcal{C}_{\kappa\epsilon/(25L^2)}\left(U\right)\\ &\leq \mathcal{C}_{\kappa\beta\epsilon/(25L^2)}\left(U/U_{22}\right) \leq \mathcal{C}_{\epsilon}(B)\,,\end{aligned}$$

with T given by (2.5.2) and

$$\beta = 2\left(\left\|(U_{22})^{-1}\right\| + 1\right)^2 \leq 2\left(\|B_1 - a\| + 1\right)^2 \leq 25L^2\,.$$

Since

$$\frac{\nu^2\epsilon}{25L^2} \geq \frac{\epsilon}{25L^2(L+4)^2} \geq \frac{\epsilon}{225L^4},$$

the first inequality in (2.3.4) follows.

□

Proof of Lemma 2.5.1. For the first part, we use the Weyl's theorem, cf. Theorem 4.3.1 in [17], which states that if

$$\sigma(A) = \{\lambda_i(A)\}_{i=1}^n\,,\ \ \sigma(B) = \{\lambda_i(B)\}_{i=1}^n\,,\ \ \sigma(A+B) = \{\lambda_i(A+B)\}_{i=1}^n$$

for Hermitian $A,\ B$, with the eigenvalues arranged in increasing order, then

$$\lambda_k(A) + \lambda_1(B) \leq \lambda_k(A+B) \leq \lambda_k(A) + \lambda_n(B)\,,\quad k = 1, ..., n\,.$$

Therefore, every number $\lambda_k(A+B)$ which lies in the interval $[-\epsilon, \epsilon]$ can be matched with $\lambda_k(A) \in [-2\epsilon, 2\epsilon]$, provided that $\|B\| \leq \epsilon$.

For the second part, observe that there exists a Hermitian matrix $\hat{A}$ such that

1. $\|\hat{A} - A\| \leq \epsilon$;
2. $\mathrm{nul}(\hat{A}) = \mathcal{C}_\epsilon(A)$;
3. $\hat{A}$ has no non-zero eigenvalues in the interval $(-\epsilon, \epsilon)$.

Then Sylvester's law of inertia implies that $\mathrm{nul}(\hat{A}) \leq \mathrm{nul}(B\hat{A}B)$ (with equality in the case of nonsingular B). Since $\|BAB - B\hat{A}B\| \leq \epsilon$ we can use Weyl's theorem again to conclude that

$$\mathcal{C}_\epsilon(BAB) \geq \mathrm{nul}(B\hat{A}B) \geq \mathrm{nul}(\hat{A}) = \mathcal{C}_\epsilon(A).$$

□

Proof of Lemma 2.5.2. The relation (2.5.4) follows from the interlacing theorem for inverses of Hermitian matrices—see Lemma 2.3 in [32], which is itself a simple consequence of the the Schur complement formula and Cauchy interlacing theorem for Hermitian matrices.

To prove (2.5.5) note that there exists a matrix

$$\hat{D} := \begin{bmatrix} \hat{A} & \hat{V} \\ \hat{V}^* & \hat{B} \end{bmatrix}$$

such that

1. $\|\hat{D} - D\| \leq \epsilon$;
2. $\operatorname{nul} \hat{D} = \mathcal{C}_\epsilon(D)$;
3. $\hat{D}$ has no non-zero eigenvalues in the interval $[-\epsilon, \epsilon]$.

where $\operatorname{nul} \hat{D}$ is the multiplicity of the zero eigenvalue of $\hat{D}$, and equals zero if this eigenvalue is absent. One can readily prove the existence of $\hat{D}$ by diagonalizing D and replacing all eigenvalues less than or equal to ϵ with zeros. Using the Haynsworth inertia additivity formula, we get

$$\operatorname{nul} \hat{D} = \operatorname{nul} \hat{B} + \operatorname{nul} \left(\hat{D}/\hat{B} \right) .$$

Observe that the condition (1) above implies $\|\hat{B} - B\| \leq \epsilon$. We can therefore infer from (2.5.3) and Lemma 2.5.1 that

$$\mathcal{C}_\epsilon(\hat{B}) = 0 \, . \tag{2.5.7}$$

As a result we obtain the equality

$$\mathcal{C}_\epsilon(D) = \operatorname{nul} \hat{D} = \operatorname{nul} \left(\hat{D}/\hat{B} \right) . \tag{2.5.8}$$

Note now that

$$\begin{aligned}
&\|\hat{V}\hat{B}^{-1}\hat{V}^* - VB^{-1}V^*\| \\
&\leq \ \|(\hat{V} - V)\| \cdot \|\hat{B}^{-1}\hat{V}^*\| + \ \|V\| \cdot \|\hat{B}^{-1} - B^{-1}\| \cdot \|\hat{V}^*\| \\
&\quad + \ \|V\hat{B}^{-1}\| \cdot \|(\hat{V}^* - V^*)\| \\
&\leq \epsilon\|\hat{B}^{-1}\| \left(\frac{1}{2} + \epsilon \right) + \frac{1}{2}\|\hat{B}^{-1} - B^{-1}\| \left(\frac{1}{2} + \epsilon \right) + \frac{1}{2}\|\hat{B}^{-1}\|\epsilon \\
&= \left(\epsilon + \epsilon^2 \right) \|\hat{B}^{-1}\| + \left(\frac{1}{2}\epsilon + \frac{1}{4} \right) \|\hat{B}^{-1} - B^{-1}\| \\
&\leq \frac{3}{2}\epsilon\|\hat{B}^{-1}\| \ + \frac{1}{2}\|\hat{B}^{-1} - B^{-1}\| \, .
\end{aligned}$$

(We have assumed that $\epsilon \leq \frac{1}{2}$.) Using the first resolvent identity, we get the bound

$$\|\hat{B}^{-1} - B^{-1}\| \ = \ \|\hat{B}^{-1}(\hat{B} - B)B^{-1}\| \ \leq \ \epsilon\|B^{-1}\| \cdot \|\hat{B}^{-1}\| \ , \tag{2.5.9}$$

which in turn implies the estimate

$$\|\hat{B}^{-1}\| \leq \|B^{-1}\| + \epsilon\|B^{-1}\| \cdot \|\hat{B}^{-1}\| \leq 2\|B^{-1}\|,$$

where we have used (2.5.7) in the last step. Inserting the last inequality into the right-hand side of (2.5.9), we finally obtain

$$\|\hat{B}^{-1} - B^{-1}\| \leq 2\epsilon\|B^{-1}\|^2.$$

As a result, we arrive at

$$\|\hat{V}\hat{B}^{-1}\hat{V}^* - VB^{-1}V^*\| \leq \frac{3}{2}\epsilon\|B^{-1}\| + \epsilon\|B^{-1}\|^2,$$

hence

$$\left\|\hat{D}/\hat{B} - D/B\right\| \leq \epsilon(1 + \frac{3}{2}\|B^{-1}\| + \|B^{-1}\|^2) < \epsilon(\|B^{-1}\| + 1)^2 =: \frac{\epsilon\beta}{2}.$$

Consequently, we get

$$\mathcal{C}_\epsilon(D) = \mathrm{nul}\left(\hat{D}/\hat{B}\right) \leq \mathcal{C}_{\frac{\beta\epsilon}{2}}\left(\hat{D}/\hat{B}\right) \leq \mathcal{C}_{\beta\epsilon}(D/B),$$

where we have used (2.5.8) in the first step and Lemma 2.5.1 in the last one. □

Proof of Lemma 2.3.4. We use induction in N. For $N = 1$ the result follows from **(A)**. Suppose that (2.3.7) holds for $|\mathcal{V}| = N$. We wish to establish the induction step, i.e. (2.3.7) for $|\mathcal{V}| = N+1$. We can evaluate $\det \hat{H}_\omega$ using the Schur determinant formula. Namely, for $x \in \mathcal{V}$ let us denote by $\hat{H}_\omega^{(x)}$ the restriction of $\hat{H}_\omega$ to the site x. Then

$$\det \hat{H}_\omega = \det \hat{H}_\omega^{(x)} \det(\hat{H}_\omega/\hat{H}_\omega^{(x)})$$

by Schur's determinant formula. Both determinants on the right-hand side are random, but the first one depends only on randomness associated with $A_\omega(x)$, a fact which we will exploit momentarily. We note now that the Schur complement $\hat{H}_\omega/\hat{H}_\omega^{(x)}$ is by itself also of the form (2.3.5) (with $\mathcal{V}$ replaced by $\mathcal{V} \setminus \{x\}$). Note that the H_0 term in $\hat{H}_\omega/\hat{H}_\omega^{(x)}$ might depend on $A_\omega(x)$, but not on the other random variables $\{A_\omega(y)\}$. By the induction hypothesis, we have

$$\mathbb{P}\left(|\det(\hat{H}_\omega/\hat{H}_\omega^{(x)})| \leq r\right) \leq (2K\alpha)^N \ln^N(r^{-1}) r^\alpha, \quad r \in [0,1]. \tag{2.5.10}$$

Let $S := \{\omega : |\det \hat{H}_\omega| \leq \epsilon\}$, and let

$$F_\omega = |\det \hat{H}_\omega^{(x)}|, \quad G_\omega = |\det(\hat{H}_\omega/\hat{H}_\omega^{(x)})|.$$

We set $Q := \{\omega : \min(F_\omega, G_\omega) \leq \epsilon\}$, then by Assumption **(A)** and the induction hypothesis

$$\mathbb{P}(Q) \leq \left(K + (2K\alpha)^N \ln^N(\epsilon^{-1})\right)\epsilon^\alpha. \tag{2.5.11}$$

On the other hand, we have

$$\chi(S \setminus Q) = \int_\epsilon^1 \chi(sG_\omega \leq \epsilon)\delta(F_\omega - s)ds.$$

Taking expectations on both sides and using (2.5.10), we obtain

$$\begin{aligned}
\mathbb{E}\chi(S \setminus Q) &\leq \mathbb{E}\int_\epsilon^1 ds\, \delta(F_\omega - s)\mathbb{E}\left(\chi(sG_\omega \leq \epsilon) \,\middle|\, A_\omega(x)\right) \\
&\leq (2K\alpha)^N \epsilon^\alpha \mathbb{E}\int_\epsilon^1 \frac{\ln^N(\frac{s}{\epsilon})\delta(F_\omega - s)}{s^\alpha} ds \\
&= (2K\alpha)^N \epsilon^\alpha \mathbb{E}\,\frac{\ln^N(\epsilon^{-1}F_\omega)\chi(1 > F_\omega > \epsilon)}{(F_\omega)^\alpha} \\
&\leq (2K\alpha)^N \epsilon^\alpha \ln^N(\epsilon^{-1})\mathbb{E}\,\frac{\chi(1 > F_\omega > \epsilon)}{(F_\omega)^\alpha}. \qquad (2.5.12)
\end{aligned}$$

Using now **(A)** and the layer cake representation, we get

$$\begin{aligned}
\mathbb{E}\,\frac{\chi(1 > F_\omega > \epsilon)}{(F_\omega)^\alpha} &= \int_1^{\epsilon^{-\alpha}} \mathbb{P}\left((F_\omega)^{-\alpha} \geq t\right) dt \leq K\int_1^{\epsilon^{-\alpha}} \frac{1}{t} dt \\
&= K\alpha \ln(\epsilon^{-1}) \qquad (2.5.13)
\end{aligned}$$

Combination of (2.5.11), (2.5.12), and (2.5.13) yields the induction step.

□

Proof of Lemma 2.4.1. We have

$$\begin{aligned}
&\left|\det(A+J)^{-1}\right| \\
&\qquad = \left|\det(A-a)^{-1}\right| \left|\det(J+a)^{-1}\right| \left|\det\left((A-a)^{-1} + (J+a)^{-1}\right)^{-1}\right|.
\end{aligned}$$

Suppose first that $\mathcal{C}_{16|a|}\left(\left((A-a)^{-1} + (J+a)^{-1}\right)^{-1}\right) = 0$. According to (2)

$$\begin{aligned}
&\mathcal{C}_{16/|a|}\left((A-a)^{-1}\left((A-a)^{-1} + (J+a)^{-1}\right)^{-1}(A-a)^{-1}\right) \\
&\qquad \leq \mathcal{C}_{16/|a|}\left((|a|/2)^{-2}\left((A-a)^{-1} + (J+a)^{-1}\right)^{-1}\right) \\
&\qquad = \mathcal{C}_{4|a|}\left(\left((A-a)^{-1} + (J+a)^{-1}\right)^{-1}\right) = 0,
\end{aligned}$$

where we have used $\|A-a\| \geq |a|/2$. Since

$$\|(A-a)^{-1}\| \leq (|a|-1)^{-1},$$

we can use (2.1.7) to decompose

$$\det (A+J)^{-1} = \det \left((A-a)^{-1} \left((A-a)^{-1} + (J+a)^{-1} \right)^{-1} (A-a)^{-1} \right) \det (H-I),$$

with

$$H = \left((A-a)^{-1} + (J+a)^{-1} \right) (A-a).$$

It follows that $\|H\| \leq 1/2$, and consequently $|\det (H-I)| \geq 2^k$. On the other hand, $|\det(A-a)| \leq (|a|+1)^k$, and we can conclude that

$$\left| \det (A+J)^{-1} \right| \geq \left(2(|a|+1)^2 \right)^{-k} \left| \det \left((A-a)^{-1} + (J+a)^{-1} \right)^{-1} \right|, \tag{2.5.14}$$

whenever $\mathcal{C}_{16|a|} \left(((A-a)^{-1} + (J+a)^{-1})^{-1} \right) = 0$.

On the other hand, if $\mathcal{C}_{16|a|} \left(((A-a)^{-1} + (J+a)^{-1})^{-1} \right) \neq 0$, then

$$\left\| \left((A-a)^{-1} + (J+a)^{-1} \right)^{-1} \right\|^{k-1} \geq \frac{1}{16|a|} \left| \det \left((A-a)^{-1} + (J+a)^{-1} \right)^{-1} \right|.$$

Hence

$$\begin{aligned} &\left\| (A-a)^{-1} \left((A-a)^{-1} + (J+a)^{-1} \right)^{-1} (A-a)^{-1} \right\| \\ &\qquad \geq (|a|+1)^{-2} \left\| \left((A-a)^{-1} + (J+a)^{-1} \right)^{-1} \right\| \\ &\qquad \geq (|a|+1)^{-2} \left\{ \frac{1}{16|a|} \left| \det \left((A-a)^{-1} + (J+a)^{-1} \right)^{-1} \right| \right\}^{1/(k-1)}. \end{aligned}$$

Using (2.1.7), we conclude that

$$\begin{aligned} \left\| (A+J)^{-1} \right\| &\geq (|a|+1)^{-2} \left\{ \frac{1}{16|a|} \left| \det \left((A-a)^{-1} + (J+a)^{-1} \right)^{-1} \right| \right\}^{1/(k-1)} \\ &\qquad - (|a|-1)^{-1} \\ &\geq \frac{1}{2} (|a|+1)^{-2} \left\{ \frac{1}{16|a|} \left| \det \left((A-a)^{-1} + (J+a)^{-1} \right)^{-1} \right| \right\}^{1/(k-1)}, \end{aligned} \tag{2.5.15}$$

if $\mathcal{C}_{16|a|} \left(((A-a)^{-1} + (J+a)^{-1})^{-1} \right) \neq 0$, whenever (2.2.2) holds.

Combining (2.5.14) and (2.5.15) we establish the assertion. □

Chapter 3

Progress report on eigenvalue statistics for alloy-type models

3.1 Introduction

In chapter 2, we proved that if the blocks of a Hamiltonian satisfy a certain criterion, then the $\lambda-$level Wegner estimate will hold. In that work, we applied this criterion by verifying the input condition for a certain quasiparticle model. We now turn to the task of verifying the same input condition for a one-dimensional alloy-type model—that is, a single-particle model with correlated random variables. Before proceeding, we will repeat the key theorem from the previous chapter for reference.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph with degree at most κ, such that the set of vertices (sites) $\mathcal{V}$ is finite with cardinality N. (Here the degree of the graph is the maximum number of edges incident at any vertex. The degree of $\mathbb{Z}^d$, for example, is $2d$.)

Let $(\Omega, \mathbb{P})$ be a probability space. Let $\{V(x) = gA_\omega(x) : \ x \in \mathcal{V}, \omega \in \Omega\}$ be a collection of independent, identically distributed random $k \times k$ Hermitian matrices.

Let $H_\omega(g)$ be a random block operator H_{block} acting on $\ell^2(\mathcal{V}; \mathbb{C}^k)$ (the space of square-summable functions $\psi : \ \mathcal{V} \to \mathbb{C}^k$) as

$$(H_\omega(g)\psi)(x) = (H_0\psi)(x) + gA_\omega(x)\psi(x), \tag{3.1.1}$$

where $g > 0$ is a coupling constant, H_0 is an arbitrary deterministic self-adjoint operator on $\ell^2(\mathcal{V}; \mathbb{C}^k)$, and the matrices $A_\omega(x)$ are independently drawn random $k \times k$ Hermitian matrices. We use the notation $H_\omega(g)$ instead of H_{block} to stress the random nature of this operator as well as the dependence on the parameter g.

Theorem 3.1.1. *Assume the following:*

For an integer ν, let S be a given set of $2\nu k$ distinct integers. Then the matrix $(A_\omega(x) - a)$ is invertible for all $a \in S$ and all $\omega \in \Omega$. Moreover, there exists an $\alpha > 0$ such that, for any integer $a \in S$, any $\epsilon \in [0, 1]$ and an arbitrary Hermitian $k \times k$ matrix J the following bound holds:

$$\mathbb{P}\left(\left|\det\left((A_\omega(x) - a)^{-1} + (J + a)^{-1}\right)\right| \leq \epsilon\right) \leq K\epsilon^\alpha \tag{3.1.2}$$

Then we have the following eigenvalue statistics result:

I. For any $E \in \mathbb{R}$ the operator $H_\omega(g) - E$ is almost surely invertible.

II. Moreover, there exist $\epsilon_0 > 0$ and $C > 0$ (which depend only on k, λ, α) for which we have

$$\mathbb{P}\left(\mathcal{C}_\epsilon(H_\omega(g) - E) \geq \lambda\right) \leq C\left|\ln(N\epsilon/g)(N\epsilon/g)^\alpha\right|^\lambda \tag{3.1.3}$$

for any $E \in \mathbb{R}$, for any $\epsilon \in [0, \epsilon_0]$ and for all $\lambda \leq \nu$. In the $\lambda = 1$ case we can improve the above bound to

$$\mathbb{P}\left(\mathcal{C}_\epsilon(H_g - E) \geq 1\right) \leq C(N\epsilon/g)^\alpha. \tag{3.1.4}$$

Thus, in order to establish the $\lambda-$level Wegner estimate for a random block operator, we only need to verify that the individual blocks satisfy the condition 3.1.2.

3.2 The alloy-type model

We wish to apply this theorem to an alloy-type model on the one-dimensional lattice $\mathbb{Z}$. We will assume that the Hamiltonian has diagonal blocks of the form A_ω, that these blocks are each of the same form, and that each contain random variables independent of the variables in the other blocks, as described above. The operator $A_\omega + J$ for this case is:

$$H_\omega = A_\omega + J = \left(\begin{array}{ccc|ccc} v & 1 & 0 & 0 & 0 & 0 \\ 1 & v & 1 & 0 & 0 & 0 \\ 0 & 1 & v & 0 & 0 & 0 \\ & & \ddots & & & \\ \hline 0 & 0 & 0 & u_1 & 1 & 0 \\ 0 & 0 & 0 & 1 & u_2 & 1 \\ 0 & 0 & 0 & 0 & 1 & u_3 \\ & & & & & \ddots \end{array}\right) + \begin{pmatrix} b_1 & 0 & 0 & 0 & 0 & 0 & \dots b_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 0 & \\ \vdots & & & & & & \ddots \\ b_2 & 0 & 0 & 0 & 0 & 0 & b_3 \end{pmatrix} \tag{3.2.1}$$

Here we have a block diagonal Hamiltonian in which each block A_ω has an $n \times n$ sub-block depending on v, followed by $(n + m)$ pairwise independent 1×1 sub-blocks depending on the variables $u_1, u_2, \dots, u_{n+m}$. As always, the subdiagonal and superdiagonal entries of A_ω are all

1's, and the other entries are all 0's. The work that follows will not be sensitive to the structure of the matrix J, so we will largely ignore the b_i variables. It is enough to say that the b_i's depend on random variables from other sites not included in the block A_ω. For the precise definition of J, see chapter 2.

The conjecture below represents our main goal.

Conjecture 3.2.1. *Let A_ω and J be as above, and assume that $v, u_1, \ldots, u_{n+m}$ are i.i.d. random variables uniformly distributed on the interval $[-1, 1]$, so that the point $(v, u_1, \ldots, u_{n+m}) = (v, \omega)$ lies in the probability space $[-1, 1] \times \Omega$ where $\Omega := [-1, 1]^{n+m}$. Then:*

$$\mathbb{P}\left(det((A_\omega + a)^{-1} + (J - a)^{-1}) \leq \epsilon\right) \leq K\epsilon^{\alpha} = K\epsilon^{\left(\frac{1}{2}+\delta\right)}$$

for $m \geq 1$.

Note on the exponent: The exponent $\alpha = 1$ would be optimal, but any value of α strictly greater than $\frac{1}{2}$ (hence any value of δ strictly greater than zero) will give a meaningful Minami estimate.

3.3 Outline of the Proof

We will not offer a complete proof of Conjecture 3.2.1, but we will offer several partial results, which together amount to an outline of the proof of this conjecture. Note that the proofs of the lemmas below are complete. (In general, anything in this work not labeled "conjecture" has a complete proof.) For an explanation of why these partial results are insufficient, see the following subsection (and further details in section 3.4 and in the discussion at the end of section 3.5).

3.3.1 Partial Results

The determinant $det\left((A_\omega + a)^{-1} + (J - a)^{-1}\right)$ can be treated as a rational function in v. Let $p(v)$ and $r(v)$, respectively, represent the numerator and denominator of this function, under the assumption that $p(v)$ and $r(v)$ have no common factors, and that the leading coefficient of $p(v)$ is 1. Then we have

$$\frac{p(v)}{r(v)} \leq \epsilon \implies p(v) \leq r(v)\epsilon \implies p(v) \leq r_{max}\epsilon$$

where r_{max} is the maximum value of $r(v)$ on the probability space Ω, which is guaranteed to exist since $r(v)$ is continuous and Ω is compact. Then $p(v)$ is a polynomial of degree n in v with coefficients that depend on the variables $\{u_1, u_2, \ldots u_{n+m}\}$.

Ultimately, we want to prove that $\mathbb{P}(p(v) \leq \epsilon) \leq K\epsilon^{\alpha}$, but as we will see later, this turns out to be problematic if two roots of $p(v)$ are close together. Thus, we will first want to show that when the roots are sufficiently far from each other, the statement above will hold, and then show that the case of closely-lying roots is itself rare.

Define the following two sets:

$$S_\eta = \{\omega \in \Omega \mid \exists \lambda_1, \lambda_2 \in \mathbb{R} : p(\lambda_1) = p(\lambda_2) = 0 \quad \text{AND} \quad |\lambda_1 - \lambda_2| \leq \eta\}$$
$$T_\epsilon = \{\omega \in \Omega \mid p(v) \leq \epsilon\}$$

In this notation, we need to verify the two statements $\mathbb{P}(T_\epsilon \backslash S_\eta) \leq K_1\epsilon^{\alpha}$ and $\mathbb{P}(S_\eta) \leq K_2\epsilon^{\alpha}$ separately. (Here η will turn out to be a power of ϵ.) Note that the probability of the event S_η depends only on the random variables u_i. That is, it does not depend on v.

In fact, we will be able to prove that $\mathbb{P}(S_\eta) \leq K_2\epsilon^{\alpha}$, but we will not be able to prove that $\mathbb{P}(T_\epsilon \backslash S_\eta) \leq K_1\epsilon^{\alpha}$, at least for the desired exponent $\alpha > \frac{1}{2}$. In section 3.4, we offer a proof of this statement for the best exponent we have been able to achieve so far.

3.3.2 Strategy for Proving the Partial Results

We will first summarize our strategy, and explain why the closely-lying roots are a problem, before dedicating one section to each of the two probability statements above.

The method for proving that $\mathbb{P}(T_\epsilon \setminus S_\eta) \leq K_1\epsilon^{\alpha}$ is as follows. We know that the integral $\int_0^1 \frac{1}{x^s}dx$ is finite if and only if $s < 1$. To put this a different way:

Lemma 3.3.1. *If*

$$\int_0^1 \frac{1}{p(v)^s}dv = K_1^* < \infty \tag{3.3.1}$$

then $\mathbb{P}(p(v) \leq \epsilon) \leq K_1\epsilon^s$. *(In fact,* K_1 *will turn out to be proportional to* K_1^**.)*

Proof. First, note that we will clearly have a problem if, for some choice of the u_i variables, $p(v)$ is equal to zero for all v. However, we can essentially rule this out: $p(v) \equiv 0$ if and only if each coefficient of the polynomial is zero. Each of these coefficients is itself a polynomial function in the u_i variables. These polynomials can in fact equal zero, but only on a set of measure zero in Ω, which may be safely ignored.

Of course, if the coefficient functions were themselves identically zero, this reasoning would fail, but that would imply that the matrix from 3.2.1 is singular for all choices of the random variables, and clearly it is not.

Rewrite the integral using the layer cake representation.

$$\int_0^1 \frac{1}{p(v)^s} dx = \int_{M^s}^{\infty} m\left\{v \in [0,1] : \frac{1}{p(v)^s} \geq t\right\} dt \tag{3.3.2}$$

where M is the maximum value of $p(v)$ on $[0,1]$, and $m(\cdot)$ represents Lebesgue measure on $\mathbb{R}$. It is easy to show that, for any given polynomial $p(v)$ (with fixed coefficients), the preimage under p of the interval $[0,\epsilon]$ has measure $K_1\epsilon^\gamma$, for all sufficiently small $\epsilon > 0$ and for some γ which depends only on the polynomial. Note that this is not an inequality—we actually have the equation:

$$m\left(\{v \in [0,1] : p(v) \leq \epsilon\}\right) = K_1\epsilon^\gamma$$

(In fact, γ will simply be the reciprocal of the highest multiplicity of any of the roots of p, which means that with probability 1, we have $\gamma = 1$.) We may modify this statement slightly to get:

$$m\left(\{v \in [0,1] : p(v)^s \leq \epsilon\}\right) = K_1\epsilon^{\gamma/s} \tag{3.3.3}$$

This last equation may be further rewritten:

$$m\left(\left\{v \in [0,1] : p(v)^s \leq \frac{1}{t}\right\}\right) = K_1\left(\frac{1}{t}\right)^{\gamma/s} \tag{3.3.4}$$

$$m\left(\left\{v \in [0,1] : \frac{1}{p(v)^s} \geq t\right\}\right) = K_1\left(\frac{1}{t}\right)^{\gamma/s} \tag{3.3.5}$$

Substituting this expression into 3.3.2 will show that that integral converges if and only if $\gamma > s$. (Note that the interval of integration here is $[M^s, \infty)$, so we want an exponent greater than one, in contrast to the original integral, where we needed an exponent less than one.) Specifically, the integral would be:

$$\begin{aligned}
\int_0^1 \frac{1}{p(v)^s} dx &= \int_{M^s}^{\infty} K_1\left(\frac{1}{t}\right)^{\gamma/s} dt \\
&= \left[K_1 \frac{t^{1-\gamma/s}}{1-\gamma/s}\right]_{M^s}^{\infty} \\
&= -K_1 \frac{M^{s-\gamma}}{1-\gamma/s} \\
&= K_1 \frac{M^{s-\gamma}}{\gamma/s-1} \\
&= K_1^*
\end{aligned}$$

If the integral above converges, then we have:

$$\mathbb{P}(p(v) \leq \epsilon) \leq K_1 \epsilon^{\gamma} < K_1 \epsilon^{s}$$

Thus, if the integral from Equation 3.3.2 (or equivalently the integral from Equation 3.3.1) converges to K_1^* for some value of s, then the probability statement $\mathbb{P}(p(v) \leq \epsilon) \leq K_1 \epsilon^s$ holds for the same s. Here $K_1 = K_1^* \left(\frac{\gamma/s-1}{M^{s-\gamma}}\right)$. □

This means that we can verify the input condition of Theorem 3.1.1 simply by checking that a certain integral converges. The potential problem here is that the integral can diverge when two or more of the roots of $p(v)$ are equal, and it can become arbitrarily large when these roots are nearly equal. This is why we need to first exclude a set in probability space (in Ω) of the form S_η. We must now show that both $\mathbb{P}(T_\epsilon \setminus S_\eta)$ and $\mathbb{P}(S_\eta)$ are acceptably small.

3.4 Bound on $\mathbb{P}(T_\epsilon \setminus S_\eta)$

If $p(v)$ has no multiple roots, then the integral from Equation 3.3.1 can be broken into a finite number of integrals over subintervals which each contain only one root. On each subinterval, the integral is at worst of the form below. Here we assume that $p(v)$ has roots at 0 and at η, and that the other roots $r_1, \ldots, r_j$ are far away. (That is, they have absolute value at least d, for some $d > 0$ which does not depend on η.)

$$\begin{aligned}
\int_0^{\eta/2} \frac{1}{p(v)^s} dv &= \int_0^{\eta/2} \frac{1}{c_1(v-r_1)^s \ldots (v-r_j)^s v^s (v-\eta)^s} dv \\
&\leq C \int_0^{\eta/2} \frac{1}{v^s(v-\eta)^s} dv \\
&\leq C \left(\frac{2^s}{\eta^s}\right) \int_0^{\eta/2} \frac{1}{v^s} dv \\
&= C \left(\frac{2^s}{\eta^s}\right) \left(\frac{\eta^{(1-s)}}{(1-s)2^{(1-s)}}\right) \\
&= C \left(\frac{\eta^{(1-2s)}}{(1-s)2^{(1-2s)}}\right) \qquad (3.4.1)
\end{aligned}$$

Here C is a constant that depends only on d and s. Note that we must have $s < 1$ and $\eta > 0$. This gives the following lemma:

Lemma 3.4.1. *If the coefficients of $p(v)$, other than the constant term, are fixed, and the nearest two roots of $p(v)$ are separated by a distance of at least η, and the other roots are separated from these and from each other by a distance of at least d, then for $s < 1$ and $\eta > 0$,*

$$\int_0^1 \frac{1}{p(v)^s} dv \leq L\eta^{(1-2s)} \tag{3.4.2}$$

where the constant L is independent of η.

The problem with this calculation is that we assume the other roots are "far away." In fact, the definition of S_η does not guarantee this. (We could, for example, have roots at 0, η, and 2η.) Thus, we will need to do a separate calculation for the case in which there are three or more roots in a small interval. The proof of the next lemma is essentially the same as the proof of Lemma 3.4.1 above.

Lemma 3.4.2. *If the coefficients of $p(v)$, other than the constant term, are fixed, and no k roots of $p(v)$ are in a single interval of width less than η, then for $s < 1$ and $\eta > 0$,*

$$\int_0^1 \frac{1}{p(v)^s} dv \leq L\eta^{(1-ks)} \tag{3.4.3}$$

where the constant L is independent of η.

This immediately implies another lemma.

Lemma 3.4.3. $\mathbb{P}(T_\epsilon \setminus S_\eta) \leq K_1\epsilon^s \leq L\epsilon^s\eta^{1-ks}$ *where L is independent of ϵ and η.*

The problem here is that the exponent on ϵ is too small, especially for large k values. Before we can prove the conjecture, we will need to tighten the bound on $\mathbb{P}(T_\epsilon \setminus S_\eta)$ in this section.

3.5 Bound on $\mathbb{P}(S_\eta)$

We now turn to the task of proving that two roots of $p(v)$ will rarely be closer together than η. We will extend this at the end of the section to consider the probability that several roots are close together.

3.5.1 Strategy

Consider the graph of $p(v)$ in the plane. The polynomial p can have two roots within distance η of each other if and only if its graph has an extremum with a function value near zero. (Hence the $v-$axis nearly touches an extremum.) We want to control the probability of this event.

The first n coefficients of the polynomial $p(v)$ determine the shape of the curve, while the constant term $c_{n+1}(u_1, \ldots, u_{n+m})$ determines the vertical position of the curve. If c_{n+1} depended on a variable independent of the variables that determine the other coefficients, then we could argue that very few values of c_{n+1} would cause an extremum to be near zero, and thus S_η would be appropriately small. The trouble, of course, is that c_{n+1} and the other coefficients all depend on the same $m + n$ random variables, so that changing any one of these variables alone will affect both the shape and the vertical position.

However, there is another way to decouple the constant term c_{n+1} from the other coefficients. Suppose the shape of the curve is fixed, and therefore the first n coefficients are fixed. This gives us $n - 1$ polynomial equations in $n + m$ variables. (Recall that the leading coefficient c_1 was fixed already.) Intuitively, one would expect that the solution set for this system would have Hausdorff dimension $m + 1$. Within this solution set, the last coefficient function c_{n+1} will still depend on the position of ω, but the shape of the curve will not. This means that we can change c_{n+1} without affecting the shape, at least a bit. It should then follow that S_η is small in some (hopefully useful) sense.

3.5.2 Outline of the proof

To make this argument work, we need to establish two things. (1) The solution set for the system of polynomial equations is usually appropriately large, in a sense to be made precise later, and (2) the dependence of c_{n+1} on the u_i variables is acceptable, in the sense that c_{n+1} cannot lie in the same small interval for too many values of the random variables. These two goals will occupy the remainder of this chapter and the next. The results are stated below. The proof of the first is in chapter 4, and the proof of the second is in the last section of this chapter.

Theorem 4.2.1 (Size of the solution set) Let $n \geq 2$ and $m \geq 2$ be integers. Let $\{p_i(x)\}_{i=1}^n$ be a collection of polynomial equations of degree at most $n(n + m)$ on a compact set $\Omega \subset \mathbb{R}^{n+m}$. For a point $\omega \in \mathbb{R}^{n+m}$, define $Q_\omega := \{x \in \mathbb{R}^{n+m} : p_i(x) = p_i(\omega)\, for\, all\, i \leq n\}$. Then the following statements hold:

(1) Q_ω has Hausdorff dimension at least m for almost all values of ω.

(2) $\mu_{m+n}\{\omega \in \Omega : \mu_m(Q_\omega \cap \Omega) \leq \epsilon\} \leq C_1 \epsilon^{1-\delta_0}$
where $\delta_0 > 0$ may be chosen to be arbitrarily small and may depend on m, but not on ϵ, C_1 depends only on n, δ_0 and Ω, and μ represents the Hausdorff measure of the appropriate dimension.

Note that for simplicity of notation, Theorem 4.2.1 and its proof use n polynomial equations in $n + m$ variables, whereas in our case we actually have the somewhat better system of $n - 1$ equations in $n + m$ variables. This means that in our case, the restriction on m is actually $m \geq 1$.

Theorem 4.2.1 establishes that the solution set is usually not too small. The next lemma shows that, within this solution set, the constant term c_{n+1} from the polynomial $p(v)$ is "well-behaved" in

the sense that c_{n+1} is in any given small interval only with small probability.

Lemma 3.5.1. (Dependence of c_{n+1} on u_i)
Assume the same conditions as in Theorem 4.2.1. Let a be any real number. Let $\omega_0 \in \Omega$ be fixed, and let $\mathbb{P}_{\omega_0}$ denote the relative probability of an event given that ω is in the solution set generated by ω_0. Then the inequality $\mathbb{P}_{\omega_0}(a \leq c_{n+1} \leq a + \epsilon) \leq \hat{L}\epsilon^{1/2+\delta_0}$ holds for all ω_0 outside a set of measure $\tilde{L}\epsilon^{1/2+\delta_0}$.

Lemma 3.5.1 allows us to exclude c_{n+1} from any interval of width ϵ, with probability $1 - \epsilon^{1/2+\delta_0}$. It is not hard to show that if we move c_{n+1} by a distance ϵ, we can guarantee that no two roots of $p(v)$ are closer together than $\epsilon^{1/2}$. Thus, we have this lemma:

Lemma 3.5.2. *The probability that the polynomial $p(v)$ has at least two roots in an interval of width η is bounded above by a constant multiple of $\eta^{1+\delta_0}$.*

Equivalently, there are at least two roots in an interval of width $\epsilon^{1/2-\delta_1}$ with probability at most a constant multiple of $\epsilon^{1/2+\delta_1}$.

Now, assume that there are at most two closely-lying roots of $p(v)$. Then we may apply Lemma 3.4.3, with $k = 2$, and get this result:

$$\begin{aligned}\mathbb{P}(T_\epsilon \setminus S_\eta) \leq K\epsilon^s &= L_2\epsilon^s\eta^{1-2s} \\ &= L_2\epsilon^s\left(\epsilon^{1/2-\delta_1}\right)^{1-2s} \\ &= L_2\epsilon^{s+\frac{1}{2}-s-\delta_1+2s\delta_1} \\ &= L_2\epsilon^{\frac{1}{2}+(2s-1)\delta_1}\end{aligned}$$

Provided $s > 1/2$, this probability is acceptably small.

More than two nearby roots. The above calculation works only for the case $k = 2$. For the other cases, we will need to modify the technique. We begin with the following lemma.

Lemma 3.5.3. *If all coefficients of the polynomial $p(v)$ other than the constant term are fixed, then $p(v)$ cannot have more than two distinct roots in an arbitrarily small interval, unless $p(v)$ is a constant function.*

Proof. If the first n coefficients are fixed, then the shape of the curve $y = p(v)$ is fixed, and only its vertical position can be changed. Let $\beta_1 \leq \beta_2 \leq \cdots \leq \beta_\ell$ be the values of v at which $p(v)$ has local extrema. (Note that $\ell \leq n - 1$.) Now let $d_p := min\{|\beta_i - \beta_j|\}$. d_p is independent of the vertical position of the curve, so in the context of this lemma, it is fixed.

Now suppose $\eta < d_p$. An interval of width η can contain at most one local extremum, and therefore it can contain at most two distinct roots of $p(v)$, by the mean value theorem. Hence, if we make η

sufficiently small, we can guarantee that there will be at most two roots of $p(v)$ in any interval of width η, regardless of vertical spacing. □

Lemma 3.5.3 does allow for the existence of more than two roots in an interval of width η if they are not distinct, but this case is easily avoided. A double or higher-order root may exist only if the function $p(v)$ has a critical point v_c such that $p(v_c) = 0$. Changing the constant term by an arbitrarily small amount will be sufficient to avoid this. Thus, outside a finite set of c_{n+1} values, it is safe to assume that all roots are simple, and then Lemma 3.5.3 shows that there can be at most two roots in a sufficiently small interval, even counting multiplicities.

Lemma 3.5.3 by itself does not solve our problem, because η depends d_p, which depends on the choice of $\omega \in \Omega$. However, this lemma does clarify one matter: it shows that d_p depends only on the first n coefficients of $p(v)$, whereas the vertical position of the curve depends only on c_{n+1}. This hints at a sort of statistical independence between the two, which we will exploit shortly.

Consider the case $k = 3$. That is, suppose there are three roots of $p(v)$ in an interval of width η, and all the others roots are separated from each other by a distance of at least d, where d is fixed. By the mean value theorem, this can happen only if there are at least two critical points of $p(v)$ in the $\eta-$interval, but note that critical points of $p(v)$ are roots of $p'(v)$. If we apply Lemma 3.5.2 to $p'(v)$ instead of $p(v)$, then we can see that this happens with probability at most a multiple of η.

This is not all we have to work with. The fact that there are three roots in the $\eta-$interval implies not only that there are two critical points in this interval, but also that the function values of these critical points are close to zero. To see this, observe that if $p'(v)$ is treated as a function of all the variables $v, u_1, \ldots, u_{n+m}$, then this function is a polynomial with fixed coefficients on a compact space, and thus its absolute value is bounded by a constant. Hence the slope of $p(v)$ cannot be too large, so in an interval of width η, the range of $p(v)$ will have width at most a multiple of η.

Thus, if we are to have three roots in this small interval, we must have both (1) two critical points of $p(v)$ in the interval, and (2) function values at those critical points which are less than a multiple of η. Lemma 3.5.3 shows that these two events are statistically independent. Lemma 3.5.2 shows that the first event occurs with probability η, and Lemma 3.5.1 shows that the second event occurs with probability $\eta^{1/2+\delta_0}$. Thus, the probability that three roots of $p(v)$ lie in the same interval of width η is bounded above by a constant multiple of $\eta^{3/2+\delta_0}$.

Since we are allowed to discard a set in probability space of size $\epsilon^{1/2+\delta_0}$, we may choose $\eta = \epsilon^{1/3+\delta_0}$ for this case. (This is not a sharp bound, but it is good enough.)

Now we may repeat the calculation based on Lemma 3.4.3, with $k = 3$.

$$\begin{aligned}\mathbb{P}(T_\epsilon \setminus S_\eta) \leq K\epsilon^s &= L_2\epsilon^s\eta^{1-3s} \\ &= L_2\epsilon^s\left(\epsilon^{1/3+\delta_0}\right)^{1-3s} \\ &= L_2\epsilon^{s+\frac{1}{3}-s+\delta_0-3s\delta_0} \\ &= L_2\epsilon^{\frac{1}{3}+(1-3s)\delta_0}\end{aligned}$$

Note: Unfortunately, the exponent in the above calculation is too small—it would need to be strictly greater than $\frac{1}{2}$ to b useful. In order to complete the proof of the desired theorem, we would need to find a tighter bound. Below we describe the analogous situation for $k > 3$.

Of course, k can be greater than 3 as well, but by now the technique for handling this situation is clear: if $k = 4$, for example, we know that there must be at least three roots of $p'(v)$ in the $\eta-$interval, and therefore there must be at least two roots of $p''(v)$ in the same interval. Thus, we have three events that must occur in order to have four roots of $p(v)$ in the $\eta-$interval: (1) there must be at least two roots of $p''(v)$ in this interval, (2) the $p'(v)$ values must be near zero, and (3) the $p(v)$ values must be near zero. All three events will be statistically independent, and their respective probabilities will be $\eta^{1/2+\delta_0}$, η, and η. Thus, the probability of all three events is proportional to at most $\eta^{5/2+\delta_0}$.

More generally, the probability that there are k roots of $p(v)$ in an interval of width η is bounded above by a constant multiple of $\eta^{\frac{2k-3}{2}+\delta_0}$. Thus, we may choose $\eta = \epsilon^{\frac{1}{2k-3}+\delta_0}$.

Thus, if we are allowed to discard a set in probability space of size $\eta = \epsilon^{1/2+\delta_0}$, then the worst-case scenario is that $p(v)$ has roots:

$$v = 0, \eta, \eta^{1/3}, \eta^{1/5}, \eta^{1/7}, \ldots, \eta^{\frac{1}{2n-3}} \tag{3.5.1}$$

Note: Once again, the integral calculation from Equation 3.4.1 will not work in this setting. We will need to either rule out this worst-case scenario, or find some tighter bounds on the calculation from Equation 3.4.1.

3.6 Proof of Lemma 3.5.1

Lemma 3.5.1 (Dependence of c_{n+1} on u_i)
Let a be any real number. Let $\omega_0 \in \Omega$ be fixed, and let $\mathbb{P}_{\omega_0}$ denote the relative probability of an event given that ω is in the solution set generated by ω_0. Then the inequality $\mathbb{P}_{\omega_0}(a \leq c_{n+1} \leq a + \epsilon) \leq \hat{L}\epsilon^{1/2+\delta}$ holds for all ω_0 outside a set of measure $\tilde{L}\epsilon^{1/2+\delta}$.

Proof. Intuitively, this lemma works if we can guarantee that moving ω some distance in the solution set will cause c_{n+1} to move a comparable distance on the real line. Our solution set may not leave much room for movement, so before even fixing a point ω_0, we need to exclude certain regions of probability space *a priori*. In particular, we wish to avoid any point at which any first partial derivative of c_{n+1} equals zero. It is near these points that c_{n+1} might change too slowly with respect to some of its variables.

Fortunately, this is not difficult to do. The function $c_{n+1}(u_n, \ldots, u_{n+m-1})$ is linear in each of its variables (with coefficients depending on the others). The derivative of this function with respect to any one variable will of course still be linear with respect to each variable.

We want to show that these derivatives are less than ϵ on a set of size at most $L_1\epsilon^{1-\delta}$. Then we could *a priori* exclude a set in probability space of size proportional to $\epsilon^{1/2+\delta}$, and outside this set we would be guaranteed that the absolute value of each derivative is reasonably large.

Lemma 3.6.1. *Let the function $f(u_1, u_2, \ldots, u_{n+m})$ be linear with respect to each of its variables, with coefficients possibly depending on the other variables. Then*

$$\mu\{\omega \in \Omega : |f(x)| \leq \epsilon\} \leq L_1\epsilon^{1-\delta} \tag{3.6.1}$$

Proof. First, we will use a change of variables to rewrite the function f in a more convenient form. Since f is linear with respect to each variable, we may write:

$$\begin{aligned} f(u_1, u_2, \ldots, u_{n+m}) &= q_1(u_2, \ldots, u_{n+m})u_1 + q_2(u_2, \ldots, u_{n+m}) \\ &= \left(u_1 + \frac{q_2(u_2, \ldots, u_{n+m})}{q_1(u_2, \ldots, u_{n+m})}\right) q_1(u_2, \ldots, u_{n+m}) \end{aligned}$$

Let the new variable w_1 denote the quantity inside the parentheses above. It is easy to see that $\mathbb{P}(w_1 \leq \epsilon) \leq D_1\epsilon$, since the variable u_1 is uniformly distributed and independent of q_1 and q_2. Furthermore, this inequality holds for the same D_1 regardless of the values of the $u_2, \ldots, u_{n+m}$ variables. This leaves us with a new formula for f:

$$f(w_1, u_2, \ldots, u_{n+m}) = w_1 q_1(u_2, \ldots, u_{n+m})$$

We may iterate this process to replace each variable u_i with a new variable w_i, each satisfying $\mathbb{P}(w_i \leq \epsilon) \leq D_i\epsilon$. Now we have:

$$f(w_1, u_2, \ldots, u_{n+m}) = w_1 w_2 \cdots w_{n+m}$$

Note that the w_i variables are not independent, but since the D_i coefficients are constant, the w_i's are "independent enough." That is, the probability that any two variables w_i, w_j are simultaneously less than ϵ is bounded above by a constant multiple of ϵ^2.

Now we have f in the same form as the function from Section 4.3.4. The calculations from that section may be repeated here, with the same result. □

Thus, we know that each of the first partial derivatives of f with respect to u_i is less than ϵ on a set of size at most $L_1\epsilon^{1-\delta}$. Equivalently, we may discard a set in probability space of measure $\epsilon^{1/2+\delta}$, and outside this set we know that the absolute value of each derivative is at least

$$\begin{aligned} L_2\epsilon^{\left(\frac{1/2+\delta}{1-\delta}\right)} &= L_2\epsilon^{(1/2+\delta)(1+\delta+\delta^2+\delta^3+\dots)} \\ &\geq L_2\epsilon^{(1/2+\delta)(1+2\delta)} \\ &= L_2\epsilon^{\left(\frac{1}{2}+\delta+\delta+2\delta^2\right)} \\ &\geq L_2\epsilon^{\left(\frac{1}{2}+4\delta\right)} \end{aligned} \tag{3.6.2}$$

Fix some point ω_0 outside the throwaway set described above. This point determines a solution set I_n. Now suppose that on the solution set I_n, some coordinate u_i is free to take any value in an interval of width d. Then by Equation 3.6.2 above, the corresponding c_{n+1} values will have a range of width at least

$$L_2\, d\, \epsilon^{1/2+4\delta} \tag{3.6.3}$$

We now need to prove that d is in fact sufficiently large—that is, we need to know how far the variables u_i are able to move within the solution set I_n determined by ω_0.

According to Theorem 4.2.1, the $m-$dimensional Hausdorff measure of I_n is at least $L_3\epsilon^{1/2+\delta}$, for ω_0 values outside an acceptably small set. We will now show that, for a solution set of this size, some of the coordinates are able to move a reasonably large distance, in a sense to be clarified below.

First, we need a certain technical lemma. As counterintuitive as it may seem, in order to prove that some of the coordinates are able to move a reasonably large distance, we must first show that none of them can move too far. That is, we need to determine how long a certain type of curve in the solution set can be. Specifically, take the original n equations defining the solution set, and add more equations of the form $u_i = constant$ until the solution set for the new system is one-dimensional. Essentially, we have fixed enough variables that the solution set becomes a curve.

Lemma 3.6.2. *Let $\hat{C} \subset I_n$ be an algebraic curve in a compact set $\Omega \subset R^{n+m}$, defined by $n+m-1$ or more polynomial equations of degree at most $n(n+m)$, and let $\langle u_1(t), u_2(t), \dots, u_{n+1}(t)\rangle$ be a parametrization of this curve. Then for each single variable function $u_i(t)$ one of the following is true.*
(1) The function $u_i(t)$ has at most N local extrema, where N depends only on n, m and the diameter of Ω.
(2) The function $u_i(t)$ is constant on some component of the curve, and nonzero elsewhere.

Proof. Note that we assume the algebraic variety $\hat{C}$ is in fact a curve: that is, it has Hausdorff dimension 1. This may in some cases require more equations than variables.

We want to know how many times $u_i'(t) = 0$. First, let $X(t)$ be the wedge product of the n gradient vectors $\nabla p_i(r(t))$. By the definition of the curve $r(t)$, the tangent vector $r'(t)$ is perpendicular to all of these gradient vectors, and thus $r'(t)$ is parallel to $X(t)$ for any t. If we are to have $u_i'(t) = 0$, then we must have $X_i(t) = 0$ as well. For any particular t, the equation $X_i(t) = 0$ is a polynomial equation in the u_j variables of degree at most $n(n+m)(n-1)$.

Now, to have a point on $\hat{C}$ where $u_i'(t) = 0$, it must be true that this point satisfies the original $n+m-1$ polynomial equations, as well as the new equation $X_i(\omega) = 0$. Thus we have a set of $n+m$ polynomial equations in $n+m$ variables. Let us call the solution set for this new system $\hat{C}^*$. This intersection could be one-dimensional or higher, but if it is zero-dimensional, then by Bézout's theorem, the number of solutions is bounded above by a finite number N which depends only on the degrees of the polynomials. Thus, we have a bound on the number of solutions, provided that the number of solutions is actually finite.

If the number of solutions is infinite, then it is an entire component of the curve $\hat{C}$. In that case, $u_i'(t) = 0$ everywhere on this component, so the coordinate function will be constant there, and the derivative will not be zero on any other components of the curve.

Hence, if the set $\hat{C}^*$ is finite, then the single-variable function $u_i(t)$ has at most N local extrema, as desired. □

From this lemma, it follows that there are at most nN points at which any one of the single variable functions $u_i(t)$ has a local extremum. For values of t in an interval (a, b) that does not contain any of these points, each of the functions $u_i(t)$ is monotone. Now if Ω has diameter d_Ω and dimension $n+m$, a curve in Ω with monotone components can have arc length at most $d_\Omega(n+m)$. It follows that the length of the curve $r(t)$ overall (not just on (a, b)) is at most $d_\Omega(n+m)nN$.

In the case that the single-variable function $u_i'(t)$ equals zero everywhere on a component of the curve $\hat{C}$, it follows that u_i is constant on that component, and monotone elsewhere. In this case, the reasoning from the previous paragraph still works. Thus, we have proved the following lemma.

Lemma 3.6.3. *Let $\hat{C}$ be an algebraic curve in a compact set $\Omega \subset \mathbb{R}^{n+m}$, where the $n+m-1$ polynomials defining $\hat{C}$ each have degree at most $n(n+m)$. Let d_Ω be the diameter of the set Ω. Then the curve $\hat{C}$ has length bounded above by a constant ℓ_{max} which depends only on n, m and d_Ω.*

We may now represent the measure of I_n in terms of the measures of its coordinate projections. For each i, Let S_i be the projection of I_n onto the u_i axis.

For illustrative purposes, we first consider the simple case in which I_n has Hausdorff dimension 2. (That is, the case $m = 3$.) Assume, without loss of generality, that the largest coordinate projection of I_n is S_1, the second largest is S_2, and so on. Suppose we fix the value of u_2 within S_2. This gives a curve $\hat{C}$ in I_n which can have length at most ℓ_{max}. In fact, we can say more. Since S_1 is the largest coordinate projection, the entire curve $\hat{C}$ is contained in an $(n+m)-$dimensional

cube of side length $\mu_1(S_1)$, in the same way that the more general curve from Lemma 3.6.3 was contained in a cube of side length d_Ω. From the proof of that lemma, it follows that the curve $\hat{C}$ in this new context has length at most

$$\mu_1(S_1)(n+m)nN = \frac{\mu_1(S_1)\ell_{max}}{d_\Omega}$$

This is an upper bound on any curve $\hat{C}$ in I_n, for any fixed value of u_2. We can then integrate this upper bound over the u_2 values to get the area of I_n. However, we cannot simply integrate with respect to u_2 itself, since this would not take into account the curvature of the surface I_n. Instead, we would integrate with respect to arc length along the curve in I_n on which u_1 is fixed. By the same reasoning as above, this curve can have length at most

$$\mu_1(S_2)(n+m)nN = \frac{\mu_1(S_2)\ell_{max}}{d_\Omega}$$

Thus the area of I_n can be at most $\mu_1(S_1)\mu_1(S_2)\left((n+m)nN\right)^2$. Similarly, for the more general case that I_n has dimension m, we have this bound on the measure of I_n.

$$\mu_m(I_n) \leq \mu_1(S_1)\ldots\mu_1(S_m)\left((n+m)nN\right)^m$$

Since each of the one-dimensional measures also satisfies $\mu_1(S_i) \leq 2$, by definition of the set Ω, we have:

$$\begin{aligned}\mu_1(S_1) &\geq \frac{\mu_m(I_n)}{\mu_1(S_2)\ldots\mu_1(S_m)\ \left((n+m)nN\right)^m} \\ &\geq \frac{\mu_m(I_n)}{2^{m-1}\ \left((n+m)nN\right)^m}\end{aligned} \tag{3.6.4}$$

Outside a set of size $\epsilon^{1/2+\delta}$, this becomes:

$$\begin{aligned}\mu_1(S_1) &\geq \frac{L_4\epsilon^{1/2+\delta+\delta_0}}{2^{m-2}\,\ell_{max}} \\ &\geq \frac{L_4\epsilon^{1/2+2\delta}}{2^{m-1}\ \left((n+m)nN\right)^m}\end{aligned} \tag{3.6.5}$$

provided $\delta > \delta_0$. (Recall that δ_0 comes from Theorem 4.2.1, and may be arbitrarily small.) Similar inequalities hold for any other projection S_i.

Thus we have:

$$\mu_1(S_i) \geq L_5 \epsilon^{1/2+2\delta} \tag{3.6.6}$$

where S_i may be any of the projections, for $i \leq m$, and L_5 may depend on n and m, but not on ϵ.

All of the S_i sets have at least this measure, but clearly only one projection could actually be this small, and only if the others had maximal measure. The largest set S_i must have measure at least:

$$\mu_1(S_i) \geq L_5^{1/m} \epsilon^{\frac{1+4\delta}{2m}} \tag{3.6.7}$$

Similarly, the largest two sets must each have measure at least

$$\mu_1(S_i) \geq L_5^{\frac{1}{m-1}} \epsilon^{\frac{1+4\delta}{2(m-1)}} \tag{3.6.8}$$

This, finally, is the distance d that we needed in equation 3.6.3. If a coordinate u_i is free to take any value in a set of the size given by equation 3.6.8, then c_{n+1} will have a range in $\mathbb{R}$ of measure at least

$$L_2 L_5^{\frac{1}{m-1}} \epsilon^{\frac{1+4\delta}{2(m-1)}} \epsilon^{1/2+4\delta} = L_6 \epsilon^{\left(\frac{1+4\delta}{2(m-1)} + \frac{1}{2} + 4\delta\right)} \tag{3.6.9}$$

Now we need to know the probability that $a \leq c_{n+1} \leq a + \epsilon$ given that u_i is in a set as described above. This probability is simply ϵ divided by the last expression from equation 3.6.9, which is:

$$\left(\frac{1}{L_6}\right) \epsilon^{\left(\frac{1}{2} - \frac{1+4\delta}{2(m-1)} - 4\delta\right)} \tag{3.6.10}$$

This probability by itself is too large, but since we have two different u_i coordinates which are statistically independent of each other, the probability with respect to both these u_i that $a \leq c_{n+1} \leq a + \epsilon$ is given by the square of the above expression:

$$\left(\frac{1}{L_6}\right)^2 \epsilon^{\left(\frac{1}{2} - \frac{1+4\delta}{2(m-1)} - 4\delta\right)(2)} = \left(\frac{1}{L_6}\right)^2 \epsilon^{\left(1 - \frac{1+4\delta}{(m-1)} - 8\delta\right)} \tag{3.6.11}$$

For $m \geq 2$, and sufficiently small δ, the exponent on ϵ will be strictly greater than $\frac{1}{2}$, as needed. This completes the proof of Lemma 3.5.1. □

Chapter 4

Solutions to systems of polynomial equations

4.1 Background

A familiar result from linear algebra states that a system of n linear homogeneous equations in n variables, which can be written as a matrix equation $Ax = 0$ for an $n \times n$ matrix A, has a unique solution, provided that the columns of A are linearly independent. Furthermore, a system of n linear homogeneous equations in $n + m$ variables (an underdetermined system), which may be described by a matrix equation $Ax = 0$ for an $n \times (n+m)$ matrix A, has a solution set isomorphic to $\mathbb{R}^m$, again provided that the columns of A are linearly independent.

It is natural to ask whether these two statements could be extended to systems of polynomial equations of degree greater than one. In asking these questions, we will relax the conditions somewhat, and will no longer demand that the right hand side of each equation (the term independent of any of the variables) is equal to zero.

The first question is answered by Bezout's theorem. According to this theorem, the solution set for a system of n polynomial equations in n variables is either infinite, or is finite with a number of solutions bounded above by the product of the degrees of the polynomials. [30]

The second question concerns underdetermined systems of polynomial equations. We would like to show that a system of n polynomial equations in $n + m$ variables "normally" has a solution set of Hausdorff dimension m, and that the Hausdorff measure of this set is "usually not too small" in a sense that obviously will need to be made more precise later.

NOTATION: In what follows, the word "surface" refers to an algebraic variety of arbitrary dimension. The word "dimension" refers to Hausdorff dimension, and finally, the term "k–dimensional measure of S" (denoted by $\mu(S)$ or $\mu_k(S)$) refers to the k–dimensional Hausdorff measure on the appropriate surface. (Hausdorff measure is defined differently by different authors. Here we

use the form that is normalized so that the $k-$dimensional Hausdorff measure is equal to the $k-$dimensional Lebesgue measure for subsets of $\mathbb{R}^k$.)

4.2 Size of the Solution Set

Theorem 4.2.1. *Let $n \geq 2$ be an integer. Let $\{p_i(x)\}_{i=1}^n$ be a collection of polynomials of degree at most $n(n+m)$ on a compact set $\Omega \subset \mathbb{R}^{n+m}$. For a point $\omega \in \Omega$, define $Q_\omega := \{x \in \mathbb{R}^{n+m} : p_i(x) = p_i(\omega)\, for\, all\, i \leq n\}$. Then the following statements hold:*

(1) Q_ω has Hausdorff dimension at least m for almost all values of ω.

(2) $\mu_{m+n}\{\omega \in \Omega : \mu_m(Q_\omega \cap \Omega) \leq \epsilon\} \leq C_1\epsilon^{1-\delta_0}$
where $\delta_0 > 0$ may be chosen to be arbitrarily small and is independent of ϵ, C_1 depends only on n, δ_0 and Ω, and μ represents the Hausdorff measure of the appropriate dimension.

Note 1: Q_ω may or may not lie completely within Ω, but the measure in part (2) refers to the intersection of the set Q_ω with Ω, not the set Q_ω itself.

Note 2: Part (1) above is technically a special case of part (2), since a set of dimension less than m would have $m-$dimensional measure zero.

Note 3: It is possible for a system of polynomial equations to have an empty solution set, but the setup of this theorem guarantees that this will not happen. Thus, this theorem is not a result on systems of polynomial equations in general, but rather a result on systems of polynomial equations that have nonempty solution sets.

4.3 Proof

4.3.1 Strategy

Let $\{S_i\}_{i=1}^n$ be the collection of level surfaces defined by $S_i := \{x \in \mathbb{R}^{n+m} : p_i(x) = p_i(\omega)\}$. Next, let $I_k := S_1 \cap S_2 \cap \cdots \cap S_k$, for each k. In this notation, $Q_\omega = I_n$.

We want to show that the final intersection I_n has measure at least ϵ in most cases. Recall that the surfaces S_i are not fixed—they depend on the choice of the point $\omega \in \Omega \subset \mathbb{R}^{n+m}$. The key idea of this proof is to find a relationship between the measure of I_i and the measure of I_{i+1}, which will then allow an iterative estimate of the measure of the final intersection I_n.

The strategy will be as follows. Suppose we consider the first surface S_1 to be momentarily fixed, but allow ω to change within this surface, thus changing the choice of S_2. Then we can use a variant of Fubini's Theorem: if the "bad" values of ω (values that give an I_n with measure less than or equal to ϵ) occupy only a set of $(n+m-1)-$dimensional measure at most ϵ on S_1 for each choice of S_1, then they occupy a set of $(n+m)-$dimensional measure at most $C_2\epsilon$ in Ω as well. (It is crucial here that Ω is compact. Otherwise this result would not necessarily hold, because we could have a thin but arbitrarily tall cylinder of bad points, for example.)

Note: Compactness can be problematic as well as helpful. It is possible that even though the solution set I_n is reasonably large, as we will show below, it lies mostly outside Ω. We will return to this point at the end of the proof.

Thus, we must show that the set of "bad" values of ω on S_1 has small $(n+m-1)-$dimensional measure, as described above, and this task will occupy the remainder of the proof.

We will begin by considering the $(n+m-2)-$dimensional measure of the intersection I_2, and later we will intersect this with the other S_i surfaces one at a time, to perform an iterative calculation of the size of the measure of the final solution set I_n. It is of course possible that the intersection I_2 has infinite measure, which would be a good thing, since we want the final intersection I_n to be reasonably large. Here though, we will consider the worst-case scenario in which all of the intersections I_i have finite measure, in an attempt to determine how small I_n could be.

We will argue that it suffices to consider cases in which the set I_{i+1} defined above has dimension $d-1$, where d is the dimension of I_i. In fact, it is entirely possible that I_{i+1} has dimension d. (For example, we could have $I_{i+1} = I_i$.) However, for our purposes this would actually be better—that is, it would give a larger solution set. Once again, we consider the worst-case scenario. It is also possible that I_{i+1} has dimension smaller than $d-1$, but we do not actually need to consider this case separately: instead, we can simply treat a set of dimension smaller than $d-1$ as a set of $(d-1)-$dimensional measure zero. We will see later that this happens with probability zero.

4.3.2 Relationship Between Measures

We would like to have a connection between the $d-$dimensional measure of the intersection I_i, and the $(d+1)-$dimensional measure of its interior. (This will be crucial to our iterative calculation.) The simplest connection would be the isoperimetric inequality. Unfortunately, this inequality does not hold in general for a $d-$dimensional algebraic variety embedded in a $(d+1)-$dimensional algebraic variety. (Imagine that the $(d+1)-$dimensional algebraic variety had an hourglass shape, for example.)

However, we will be able to prove a weaker result which will suffice for our purposes.

Lemma 4.3.1. *Let V be an algebraic variety of dimension $d+1$ in a compact set $\Omega \subset R^{n+m}$, and let $p(\omega)$ be a polynomial on Ω. Let $C(\omega)$ be the contour given by the intersection of V with the set where $p(\omega) = a$ for some constant a. Let D_ϵ be the subset of V defined by the following:*

$$D_\epsilon = \{\omega \in V : \mu_d(C(\omega)) \leq \epsilon\}$$

Then $\mu_{d+1}(D_\epsilon) \leq B\epsilon$. Here B depends only on the set Ω and the degrees of the polynomials.

The idea of this lemma is that we have taken the union of all contours with "length" ($(d-1)-$dimensional measure) less than or equal to ϵ, and this union has "area" ($d-$dimensional measure) at most a multiple of ϵ. This area can be interpreted as a probability: the probability that a point $\omega \in V$ is on a contour of measure less than or equal to ϵ is itself less than or equal to a multiple of ϵ.

Proof. Each contour is determined by a level set of the polynomial $p(\omega)$, and has length at most ϵ. The intuition behind this proof is that if we could define some coordinate system on V in which one coordinate is perpendicular to the contours, and the other $d-1$ coordinates follow the contours, then integrating with respect to these new coordinates would give a $d-$dimensional measure that is proportional to ϵ.

Define a "gradient curve" as follows. A curve $r(t)$ in V is said to be a gradient curve if, for any t, the orthogonal projection of $\nabla p(r(t))$ onto the tangent plane to V at $r(t)$ is parallel to the tangent vector $r'(t)$ of the curve.

Each point on $D_\epsilon \subset V$ is on exactly one contour of length at most ϵ, and on exactly one gradient curve. If we could show that the gradient curves all have length less than or equal to some maximum value B, which does not depend on ϵ, then it would follow that the measure of D_ϵ is at most $B\epsilon$.

To find this maximum length, note that we can actually write an equation to describe a gradient curve $r(t)$. Recall that V is an algebraic variety defined by some set $\{q_i(x) = q_i(\omega)\}_{i=1}^k$ of polynomial equations. The tangent space $T(r(t))$ to the surface V at the point $r(t)$ is defined as the space of all vectors perpendicular to each of the gradients $\nabla q_i(r(t))$.

Thus we have:

$$\begin{aligned} r'(t) = M(t)[\nabla p(r(t)) - \nabla q_1(r(t))\,(\nabla p(r(t)) \cdot \nabla q_1(r(t))) \\ \cdots - \nabla q_n(r(t))\,(\nabla p(r(t)) \cdot \nabla q_n(r(t)))] \end{aligned} \tag{4.3.1}$$

where $M(t)$ is some real-valued function. Also, without loss of generality, we will choose the parametrization for $r(t)$ such that $0 \leq t \leq 1$ for the part of the curve that lies in Ω.

The significance of equation 4.3.1 is that it implies that the components of $r'(t)$ are polynomials in the u_i variables, with degrees that cannot exceed $N = (n(n+m))^2$. We will return to this formula in a moment.

The length of the curve $r(t)$ may be calculated as follows:

$$\begin{aligned}
\ell(r) &= \int_0^1 |r'(t)|dt \\
&= \int_0^1 \sqrt{r'(t) \cdot r'(t)}dt \\
&\leq \sqrt{\int_0^1 r'(t) \cdot r'(t)dt} \\
&= \sqrt{\int_0^1 M(t)[\nabla p - \nabla q_1 (\nabla p \cdot \nabla q_1) \cdots - \nabla q_n (\nabla p \cdot \nabla q_n)] \cdot r'(t)dt}
\end{aligned} \tag{4.3.2}$$

Here we have suppressed the dependence of the various gradients on the point $r(t)$ to simplify notation. Note also that we used Jensen's inequality to get line three above. More specifically, Jensen's inequality guarantees that if $b - a = 1$, then:

$$\left(\int_a^b f(x)dx\right)^2 \leq \int_a^b (f(x))^2 dx$$

If we take the square root of both sides of this inequality, and define $g(x) = (f(x))^2$, then this gives:

$$\int_a^b \sqrt{g(x)}dx \leq \sqrt{\int_a^b g(x)dx}$$

It is this last inequality that justifies line three from Equation 4.3.2.

For the next step, we will need to know that each of the dot products $(\nabla q_i \cdot \nabla p)$ has absolute value bounded above by some constant R. This is guaranteed since these expressions are polynomial functions on a compact set.

Also, we would like to remove the dependence on $M(t)$. First, note that the only restrictions on the curve $r(t)$ are that its derivative is parallel to $\nabla p(r(t))$ at any point, and that the parametrization takes t values in the unit interval. Within these restrictions, we are free to choose $M(t) = M$, a constant.

$$
\begin{aligned}
&\int_0^1 M[\nabla p - \nabla q_1 (\nabla p \cdot \nabla q_1) \cdots - \nabla q_n (\nabla p \cdot \nabla q_n)] \cdot r'(t)dt \\
&\leq \left| \int_0^1 M[\nabla p - \nabla q_1 (\nabla p \cdot \nabla q_1) \cdots - \nabla q_n (\nabla p \cdot \nabla q_n)] \cdot r'(t)dt \right| \\
&\leq M \left| \int_0^1 \nabla p \cdot r'(t)dt \right| + M \left| \int_0^1 (\nabla q_1 \cdot \nabla p) \nabla q_1 \cdot r'(t)dt \right| \\
&\qquad \cdots + M \left| \int_0^1 (\nabla q_n \cdot \nabla p) \nabla q_n \cdot r'(t)dt \right| \\
&\leq M \left| \int_0^1 \nabla p \cdot r'(t)dt \right| + R\, M \left| \int_0^1 \nabla q_1 \cdot r'(t)dt \right| \\
&\qquad \cdots + R\, M \left| \int_0^1 \nabla q_n \cdot r'(t)dt \right| \\
&= M\, |p(r(1)) - p(r(0))| + R\, M\, |q_1(r(1)) - q_1(r(0))| \\
&\qquad \cdots + R\, M\, |q_n(r(1)) - q_n(r(0))|
\end{aligned}
$$

The last line follows from a multivariable form of the fundamental theorem of calculus. Since the p and q_i functions are polynomials in the u_i variables, and since Ω is compact, the absolute values from the last line above each have a uniform upper bound on Ω. It follows immediately that the length of the curve is bounded above by some constant B which depends only on the degrees of the polynomials and on the set Ω. □

The point of this section is that we have a relationship between the measure of a contour on V and the probability that ω lies on a contour of that size or smaller. Specifically, Lemma 4.3.1 implies the following:

$$\mu_d(C) \geq \hat{K} \mu_{d+1}(D_C) \tag{4.3.3}$$

where C is the contour through some point in V and D_C is the set of all $\omega \in V$ whose contours have measure less than or equal to the measure of C. Equation 4.3.3 may be viewed as a rough analog of the isoperimetric inequality, in the sense that it gives an upper bound on a certain $(d+1)-$dimensional measure (analogous to area) in terms of a $d-$dimensional area (analogous to perimeter). This inequality implies:

$$\frac{\mu_d(C)}{\mu_{d+1}(V)} \geq \hat{K} \frac{\mu_{d+1}(D_C)}{\mu_{d+1}(V)}$$

Note that the right-hand side above is simply a ratio of two areas, and hence if we define $r := m(D_C)/m(v)$, the ratio r will be a random variable that is uniformly distributed on the interval $[0, 1]$.

4.3.3 Iterative Calculation of Arc Length

Suppose in the worst case scenario that all the intersections I_i have finite measure. As above, we have:

$$\frac{\mu(I_n)}{\mu(I_{n-1})} \geq K_n \frac{\mu(D_n)}{\mu(I_{n-1})}$$

but this may be rewritten as:

$$\begin{aligned}\frac{\mu(I_n)}{\mu(I_0)} &\geq K_n\left(\frac{\mu(D_n)}{\mu(I_{n-1})}\right) K_{n-1}\left(\frac{\mu(D_{n-1})}{\mu(I_{n-2})}\right)\ldots K_1\left(\frac{\mu(D_1)}{\mu(I_0)}\right)\\ &= K_n(r_n)\,K_{n-1}(r_{n-1})\ldots K_1(r_1)\\ &= \tilde{K} r_n r_{n-1}\ldots r_1\end{aligned}$$

Here the symbol I_0 is used to represent the "zeroth intersection" which is the probability space Ω itself. This inequality may be rewritten as:

$$\begin{aligned}\mu(I_n) &\geq \tilde{K}\,(r_n r_{n-1}\ldots r_1)\,\mu(I_0)\\ &= \tilde{K}\,(r_n r_{n-1}\ldots r_1)\,\mu(\Omega)\end{aligned}$$

4.3.4 Probability Calculation

We wish to estimate the probability that $\mu(I_n) < \epsilon$ —that is, the probability that the above product is less than epsilon. That statement may be rewritten as:

$$\begin{aligned}\mu(I_n) \leq \epsilon &\Longrightarrow \tilde{K}\,(r_n r_{n-1}\ldots r_1)\,\mu(\Omega) \leq \epsilon\\ &\Longrightarrow r_n \leq \frac{\epsilon}{\tilde{K}\,(r_{n-1}\ldots r_1)\,\mu(\Omega)}\end{aligned}$$

Since each of the random variables r_i is uniformly distributed on the unit interval $[0, 1]$, the probability of this event will be given by the integral:

$$\begin{aligned}
&\mathbb{P}\left(r_n \leq \frac{\epsilon}{\tilde{K}\left(r_{n-1}\dots r_1\right)\mu(\Omega)}\right) \\
&\leq \frac{\epsilon}{\tilde{K}\mu(\Omega)}\int_0^1\dots\int_0^1 \frac{1}{r_{n-1}\dots r_1}dr_1\dots dr_{n-1} \\
&= \frac{\epsilon}{\tilde{K}\mu(\Omega)}\left(\int_0^1 \frac{1}{r_1}dr_1\right)\dots\left(\int_0^1 \frac{1}{r_{n-1}}dr_{n-1}\right)
\end{aligned}$$

A priori this is not good, because the integrals above all diverge. However, the problem can be circumvented. The integrals diverge only because of the infinite tails near 0, where r_n grows very large, but since in reality r_n should be at most 1, we can cut off the tails and salvage the integral. The detailed calculations are below.

For only two variables, the probability that $r_1 r_2 \leq \epsilon$ given that $0 \leq r_1 \leq 1$ and $0 \leq r_2 \leq 1$ is:

$$\begin{aligned}
&\int_0^\epsilon 1\,dr_2 + \int_\epsilon^1 \left(\frac{\epsilon}{r_2}\right)dr_2 \\
&= \epsilon + \epsilon\left(ln(1) - ln(\epsilon)\right) \\
&= \epsilon + \epsilon\, ln\left(\frac{1}{\epsilon}\right) \\
&\leq \epsilon + \epsilon\left(\frac{1}{\epsilon}\right)^{\delta_1} \\
&= \epsilon + \epsilon^{1-\delta_1} \\
&\leq 2\epsilon^{1-\delta_1}
\end{aligned}$$

where $\delta_1 > 0$ may be arbitrarily small, and the inequality holds for all $\epsilon < \epsilon_0$, for some ϵ_0 that depends on δ_1.

We may now use induction to prove that a similar result holds for more than two variables. In particular, suppose we have already verified that $\mathbb{P}\left(r_1 r_2 \dots r_{n-1} \leq \epsilon\right) \leq \tilde{K}_{n-1}\epsilon^{1-\delta_2}$. To incorporate the last variable r_n and prove that

$$\mathbb{P}\left(r_1 r_2 \dots r_{n-1} r_n \leq \epsilon\right) \leq \tilde{K}_n \epsilon^{1-\delta_3}$$

we will let $q_n = r_1 r_2 \dots r_{n-1}$ and note that q_n is a random variable on $[0,1]$ with a known distribution. Thus, we may use a similar calculation to the above to find $\mathbb{P}\left(q_n r_n \leq \epsilon\right)$.

In the integral, we may replace q_n with $x^{\frac{1}{1-\delta_2}}$ for a random variable x that is uniformly distributed on $[0,1]$. The probability that $q_n r_n \leq \epsilon$ (equivalently, $r_n \leq \frac{\epsilon}{q_n}$) given that $q_n, r_n \in [0,1]$ is:

$$\begin{aligned}
\int_0^\epsilon 1\,dr_1 + \int_\epsilon^1 \left(\frac{\epsilon}{x^{\frac{1}{1-\delta_2}}}\right) dx &= \int_0^\epsilon 1\,dr_1 + \int_\epsilon^1 \left(\epsilon x^{-\frac{1}{1-\delta_2}}\right) dx \\
&= \epsilon + \epsilon \left[\frac{x^{1-\frac{1}{1-\delta_2}}}{1-\frac{1}{1-\delta_2}}\right]_\epsilon^1 \\
&= \epsilon + \epsilon \left[\frac{x^{\frac{-\delta_2}{1-\delta_2}}}{\frac{-\delta_2}{1-\delta_2}}\right]_\epsilon^1 \\
&= \epsilon + \epsilon \left(\frac{1-\delta_2}{-\delta_2}\right) \left[x^{\frac{-\delta_2}{1-\delta_2}}\right]_\epsilon^1 \\
&= \epsilon + \epsilon \left(\frac{1-\delta_2}{-\delta_2}\right) \left(1-\epsilon^{\frac{-\delta_2}{1-\delta_2}}\right)
\end{aligned}$$

If we let $\delta_3 := \frac{-\delta_2}{1-\delta_2}$, then this becomes:

$$\begin{aligned}
\epsilon + \epsilon \left(\frac{1-\delta_2}{-\delta_2}\right) \left(1-\epsilon^{\frac{-\delta_2}{1-\delta_2}}\right) &= \epsilon + \epsilon \left(-\frac{1}{\delta_3}\right) \left(1-\epsilon^{-\delta_3}\right) \\
&= \left(1-\frac{1}{\delta_3}\right)\epsilon + \left(\frac{1}{\delta_3}\right)\epsilon^{1-\delta_3} \\
&\leq \left(\frac{1}{\delta_3}\right)\epsilon^{1-\delta_3}
\end{aligned}$$

where the last inequality holds because $\left(1-\frac{1}{\delta_3}\right)$ is negative for sufficiently small δ_3.

The precise value of this integral is unimportant. The only important part is the dependence on ϵ. On that matter, we may conclude that the probability of the event in question is at most proportional to $\epsilon^{1-\delta_3}$, for some δ_3 which may be arbitrarily small. (This δ_3 is the δ_0 value from the statement of Theorem 4.2.1.) Note that δ_3 does not depend on ϵ, but the coefficient of ϵ does depend on δ_3. In applications this will mean that it is important to keep δ_3 fixed.

Thus, with probability $1 - B\epsilon^{1-\delta_3}$ for some constant B, the solution set will have measure at least ϵ.

Note 1: Suppose the algebraic varieties $I_2, \ldots, I_n$ are not all of finite measure, as assumed above. This is possible only if the first several $I_2, \ldots, I_k$ have infinite measure and the rest have finite measure. If this happens, we can evaluate a truncated version of the integral above, with similar results. (Here $k = n$ is a real possibility, and in that case there is nothing to prove: the measure of the set I_n will be infinite, which is far better than necessary.)

Note 2: So far we have established that the solution set Q_ω is not too small for most values of ω. The remaining problem is that for some ω, the solution set may be sufficiently large, but lie mostly outside Ω. We may sidestep this issue by expanding Ω. Specifically, we define $\bar{\Omega}$ to be the subset of $\mathbb{R}^{n+m}$ that includes Ω, and any point that shares a common solution set with a point of Ω. In some cases, this set may not be compact, since some solution sets may have infinite measure.

However, we can correct this problem by cutting off all solution sets at some distance d from the boundary of Ω, with d chosen so that each solution set will still have measure at least ϵ_0. This third probability space may be called Ω_0, and we may now conclude that Theorem 4.2.1 holds on this space. Clearly, though, if the statements from that theorem hold for all ω in some set that contains Ω, then they hold for all ω in Ω itself. Thus, while the solution sets that are cut off by the boundary of Ω may be a problem at intermediate steps of the proof, the final conclusions will still hold on the original probability space.

This completes the proof of Theorem 4.2.1.

Chapter 5

Conclusion

In summary, we have proved a criterion under which operators of a certain block matrix form are guaranteed to satisfy Wegner and Minami estimates, as well as higher-order estimates of the same form. We have demonstrated the utility of this criterion by applying it to a Hamiltonian representing a quasiparticle model arising in the theory of dirty superconductors.

We also proved two more abstract theorems, which may be of interest in their own right. The first is a result from linear algebra which gives a lower bound on the number of small eigenvalues of a matrix A based on properties of certain sub-matrices of A^{-1}. The second is a result from algebraic geometry which asserts that, subject to certain conditions, the solution set of a system of n polynomial equations in $n+m$ variables is almost always $m-$dimensional, and its $m-$dimensional measure satisfies a probabilistic lower bound.

We have also presented a progress report on an attempt to verify the Wegner and Minami estimates for a class of alloy-type models, using the same criterion mentioned above, as well as the theorem on solution sets.

As always, this work leaves plenty of room for improvement. It would be desirable, for example, to prove the estimates for a more general class of alloy-type models. Nonetheless, the examples given here should suffice to demonstrate the potential of this new method of determining eigenvalue statistics for various types of random block operators.

Bibliography

[1] M. Aizenman and G. M. Graf, Localization bounds for an electron gas, J. Phys. A 31 (1998), 6783-6806.

[2] F. Almgren, Optimal Isoperimetric Inequalities, Indiana University Mathematics Journal, Vol. 35, No. 3 (1986).

[3] P. W. Anderson, Absence of diffusion in certain random lattices. *Phys. Rev.* **109** (1958), 1492–1505.

[4] J. Bellissard, P. Hislop, A. Klein, and G. Stolz, Random Schrödinger operators: universal localization, correlations, and interactions. www.birs.ca/workshops/2009/09w5116/report09w5119.pdf

[5] J. Bellissard, P. Hislop, and G. Stolz, Correlation estimates in the Anderson model. *J. Stat. Phys.* **129** (2007), 649–662.

[6] J. Bourgain, An approach to Wegner's estimate using subharmonicity. *J. Stat. Phys.* **134** (2009), 969–978.

[7] Z. Cao and A. Elgart, The weak localization for the alloy-type Anderson model on a cubic lattice. *J. Stat. Phys.* **148** (2012), 1006–1039.

[8] R. Carmona and J. Lacroix, Spectral theory of random Schrödinger operators. Boston: Birkh"auser (1990).

[9] J.-M. Combes, F. Germinet, and A. Klein, Generalized eigenvalue-counting estimates for the Anderson model. *J. Stat. Phys.* **135** (2009), 201–216.

[10] J.-M. Combes, F. Germinet, and A. Klein, Poisson statistics for eigenvalues of continuum random Schrödinger operators. *Anal. PDE* **3** (2010), 49–80.

[11] A. Elgart and A. Klein, Ground state energy of trimmed discrete Schrödinger operators and localization for trimmed Anderson models. *J. Spectr. Theory* **4** (2014), 391–413.

[12] A. Elgart and D. Schmidt, Eigenvalue Counting Inequalities, with Applications, To appear in the Journal of Spectral Theory.

[13] A. Elgart, M. Shamis and S. Sodin, Localisation for non-monotone Schrödinger operators. *J. Eur. Math. Soc.* **16** (2014), 909–924.

[14] S. Fishman, Y. Krivolapov, and A. Soffer, Perturbation Theory for the Nonlinear Schrödinger Equation with a random potential. *Nonlinearity* **22** (2009), 2861.

[15] M. Gebert and P. Müller, Localization for random block operators. *Oper. Theory Adv. Appl.* **232** (2013), 229–246.

[16] G.-M. Graf and A. Vaghi, A remark on the estimate of a determinant by Minami. *Lett. Math. Phys.* **79** (2007), 17–22.

[17] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[18] W. Kirsch, Wegner estimates and Anderson localization for alloy-type potentials. *Math. Z.* **221** (1996), 507–512.

[19] W. Kirsch, An invitation to random Schrödinger operators, Random Schrödinger operators, Panor. Synthèses, vol. 25, Soc. Math. France, 2008, With an appendix by Frédéric Klopp, pp. 1–119.

[20] W. Kirsch, B. Metzger and P. Müller, Random block operators. *J. Stat. Phys.* **143** (2011), 1035–1054.

[21] F. Klopp, Inverse tunneling estimates and applications to the study of spectral statistics of random operators on the real line. *J. Reine Angew. Math.* **690** (2014), 79–113.

[22] K. Leonhardt, N. Peyerimhoff, M. Tautenhahn, and I. Veselić, Wegner estimate and localization for alloy-type models with sign-changing exponentially decaying single-site potentials. http://arxiv.org/abs/1309.0109 arXiv:1309.0109

[23] R. de la Madrid, Quantum mechanics in rigged Hilbert space language, PhD thesis, Depertamento de Fisica Teorica Facultad de Ciencias, Universidad de Valladolid, 2001.

[24] N. Minami, Local fluctuation of the spectrum of a multidimensional Anderson tight binding model. *Commun. Math. Phys.* **177** (1996), 709–725.

[25] S. Molchanov, The local structure of the spectrum of the one-dimensional Schrödinger operator. *Commun. Math. Phys.* **78** (1981), 429–446.

[26] H. L. Royden, Real analysis, Third Edition. New York: Macmillan (1988).

[27] P. Stollmann, Caught by disorder: bound states in random media. Progress in Mathematical Physics. 20. Boston: Birkhäuser (2001).

[28] G. Stolz, An introduction to the mathematics of Anderson localization. Entropy and the quantum II, Contemp. Math. **552**, pp. 71–108, Amer. Math. Soc., Providence, RI (2011).

[29] M. Tautenhahn and I. Veselić, Minami's Estimate: Beyond Rank One Perturbation and Monotonicity. *Ann. Henri Poincaré* **15** (2014), 737–754.

[30] W. Vogel, Lectures on results on Bezout's theorem. Springer-Verlag, New York (1984).

[31] F. Wegner, Bounds on the density of states in disordered systems. *Z. Phys. B* **44** (1981), 9–15.

[32] *The Schur complement and its applications.* Edited by Fuzhen Zhang. Numerical Methods and Algorithms, 4. Springer-Verlag, New York, 2005.