HARVARD COLLEGE

CLASS OF 2017

Bulk Universality of Biregular Bipartite Graphs and Dyson's Brownian Motion for Covariance Matrices

---

Author:
Kevin YANG

Supervisor:
Prof. Horng-Tzer YAU

*Submitted in partial fulfillment of the
requirements for the degree of Bachelor
of Arts with Honors in Mathematics*

March 20, 2017

Introduction

All of our struggles were in vain. 82 levels were too few to give a statistically significant test of the model. As a contribution of the understanding of nuclear physics, random matrix theory was a dismal failure. By 1970 we had decided that it was a beautiful piece of work having nothing to do with physics.

– Freeman Dyson, 1970.

1. Historical Background

In 1928, biologist John Wishart introduced the theory of random matrices in studying large data sets. In [24], Wishart viewed large data sets as so complex that studying the finer details of the corresponding covariance matrices, e.g. with methods of principal component analysis, became computationally intractable problems. To effectively model covariance statistics of large data sets with principal component analysis, Wishart viewed the data as random. Unfortunately, Wishart's perspective did not progress far and was thus left behind.

A few decades later in the 1950s, Nobel prize-winning physicist Eugene Wigner reintroduced random matrices in [23] to study the distribution of energy levels in heavy nuclei. Wigner's motivation for modeling heavy nuclei with large random matrices resembled Wishart's original motivation in statistics: heavy nuclei are too complex to effectively model all precise details. On the other hand, Wigner's ensemble of matrices were either real symmetric or complex Hermitian matrices, as opposed to Wishart's covariance matrices. Wigner discovered important phenomena such as the semicircle law and computed eigenvalue statistics for an important class of matrices known as the Gaussian Orthogonal Ensemble (GOE) and the Gaussian Unitary Ensemble (GUE) for low dimensions. These results became known as *Wigner's surmise*, and Wigner's random real symmetric and complex Hermitian matrices became known as the *Wigner matrix ensemble*.

From here, mathematicians Michel Gaudin and Madan Mehta, along with physicist Freeman Dyson, studied Wigner's work in more detail in [14], [19], and [10]. These three papers extended Wigner's surmise to all dimensions, and moreover [10] introduced the random eigenvalue problem from the perspective of the statistical mechanics of gases. Ultimately, Wigner, Dyson, Gaudin and Mehta made the following loosely stated conjecture concerning *local correlation functions* of eigenvalues, which determine microscopic location statistics.

Conjecture 1.1. (WDGM Conjecture).

The local correlation functions in the bulk of the semicircle distribution (or bulk, for short) for Wigner ensembles are independent of the underlying matrix distribution and depend only on the symmetry class (e.g. real symmetric, complex Hermitian). In particular, the local correlation functions should all converge to the local correlation functions of either the GOE and GUE as the matrix size goes to infinity.

The WDGM conjecture is thus a universality result and may be thought of as a matrix-valued central limit theorem – roughly speaking, for large systems the randomness behaves like a Gaussian family. The first major step towards establishing any precise result along the lines of the WDGM conjecture appeared in [9], in which Dyson defined a stochastic flow for any given Wigner matrix ensemble and computed the corresponding eigenvalue dynamics known as *Dyson's Brownian Motion*, or DBM for short. In many papers such as [11] and [13], Laszlo Erdos, H.T. Yau and many collaborators exploited the relaxation property of DBM and established both short-time stability of eigenvalue statistics under this flow, as well as fast time to relaxation towards Gaussian statistics. Moreover, through analysis of the DBM, Erdos, Yau, and their collaborators have also proven universality of gap statistics, which Wigner viewed as excitation energies. These works answered Dyson's conjecture, which we loosely state as follows.

Conjecture 1.2. (Dyson's Conjecture).

The time to equilibrium of eigenvalue statistics along the DBM flow for matrices of dimension N is roughly $\frac{1}{N}$.

In fact, Dyson's conjecture was inspired by a statistical picture of gas particles reaching equilibrium. We summarize the robust method developed and exploited by Erdos, Yau and their collaborators to answer Dyson's conjecture and also, in part, the WDGM conjecture. The method comes in the following three parts.

- Step 1: Prove a *local law*, a coincidence of eigenvalue distributions with the universal limit on intervals as small as the scale of fluctuations of eigenvalues. This provides a rigidity estimate, which bounds how far eigenvalues can deviate from where the universal limit predicts they are with high probability.
- Step 2: Prove a stability estimate, which shows statistics such as eigenvalue gaps and correlation functions change negligibly in the limit of large N after evolving the initial data through the DBM for time $t = N^{-1+\varepsilon}$. Here, $\varepsilon > 0$ is an arbitrarily small but fixed parameter.
- Step 3: Prove short time to relaxation, which shows the statistics already reach Gaussian behavior after evolving the eigenvalues for time $t > N^{-1+\delta}$. Here, $\delta > 0$ is any fixed positive parameter. In other words, this allows us to compare statistics at time $t = N^{-1+\delta}$ with time $t = +\infty$. This is Dyson's Conjecture.

More recently, as in [3] and [4], universality of eigenvalue statistics have been extended to matrix ensembles beyond the Wigner ensemble, including adjacency matrices of random regular graphs whose entries are sparse and correlated.

2. The Problem for Covariance Matrices

The aforementioned works primarily concern Wigner matrices, whereas the focus of this thesis concerns covariance matrices. With similar ideas, a stochastic eigenvalue flow was computed in the spirit of DBM, e.g. in [12]. Moreover in [12], using ideas from statistical mechanics, ergodic theory of Brownian motions, and techniques used for Wigner matrix ensembles, Erdos, Schlein, and Yau proved the following loosely stated result.

Theorem 2.1. *Suppose the initial data for the covariance matrix DBM is, in some sense, optimal. Then Dyson's Conjecture holds for this initial data.*

The techniques used in [12] address only universality of local correlation functions and not of gap statistics. The initial data assumed in [12] is also restrictive. This implies the results cannot be applied directly to covariance matrix ensembles of interest, e.g. sparse covariance matrices. Assuming the restrictive initial data, however, in [22] universality of local correlation functions was proven for a class of random matrices drawn from a suitably sub-exponential distribution. This paper also follows the robust three-step strategy outlined above. Again, the problem concerning universal gap statistics is omitted from this paper, as it relies on the results in [12] to tackle Dyson's conjecture.

As for sparse covariance matrices, an earlier senior thesis at Harvard [1] establishes a local law for this ensemble. Moreover, a detailed DBM is derived. However, neither short-time stability nor fast convergence to local ergodicity of eigenvalue statistics is addressed.

3. This Thesis and Our Results

In this thesis, we address the issues presented above for covariance matrix ensembles. The following ideas are original to this thesis unless explicitly mentioned as adapted from earlier works.

First, we adapt the methods of [16] and [17] to recover universality of correlation functions in the bulk for a wide class of initial data. This is an improvement of Dyson's Conjecture for covariance matrix ensembles and removes the strong a priori assumptions required in [11] and thus used in [22]. Moreover, in doing so we also derive eigenvalue gap universality. To do this, instead of looking at the covariance matrix ensemble, we look at a linearization whose eigenvalues and eigenvectors are in natural correspondence with those of the corresponding covariance matrix. In particular, we establish gap universality for an ensemble of linearized covariance matrices rather than covariance matrices themselves. Given the nonlinear spectral correspondence between these two ensembles, this will imply no such gap result should exist for covariance matrix ensembles. To this end we also derive the corresponding eigenvalue dynamics for the eigenvalues of the linearized covariance matrix, which is also an original contribution of this thesis. This is the covariance analog to the DBM derived in [9]. With the explicit

eigenvalue dynamics, we may use an important idea from [16] in cutting off interactions on large scales and compare the dynamics to the DBM for Wigner matrices derived in [9].

The core of this thesis, however, is concerned with deriving universality of eigenvalue correlation functions for an ensemble of covariance matrices $X = H^*H$ where the entries of H are both sparse and correlated. In the spirit of the papers [3] and [4], we look at adjacency matrices of biregular bipartite graphs. With this ensemble, we adapt the methods of [4] to prove a local law and we adapt the methods of [3] to short-time stability of local correlation functions in the bulk. This completes Step 1 and Step 2 in the robust method outlined above for bipartite graphs.

In summary, the main result of this thesis is the following informally stated result.

Theorem 3.1. *The three-step strategy holds for local correlation functions for the ensemble given by adjacency matrices of biregular bipartite graphs. Moreover, Dyson's conjecture holds for a wide class of covariance matrix ensembles, including the ensemble of bipartite graphs.*

We conclude a first description of this thesis by noting that we will also address the unanswered questions concerning [1], [12], and [22]. We will not provide all details to fully answer these open questions as this is not the focus of this thesis, i.e. Theorem 3.1. We will, however, comment briefly on the necessary ingredients and methods to provide such answers.

4. Organization of this Thesis

We organize this thesis into three chapters, one for each of the steps in the robust three-step strategy.

In Chapter I, we concern ourselves with deriving a local law for adjacency matrices of biregular bipartite graphs. In this chapter, the local law is phrased in terms of estimates on the entries of the Green's function of the adjacency matrix. We study the adjacency matrices by exploiting the combinatorial structure of the underlying bipartite graph. The most important of these combinatorial methods is the construction of switchings on graphs which generate tractable dynamics of the Green's functions. We also follow standard methods in random matrix theory and exploit the analytic properties of the Green's functions, including the fixed-point theory.

In Chapter II, we begin evolving the ensemble of adjacency matrices under a matrix-valued Ornstein-Uhlenbeck process. We compute the generator of this process as a differential operator with coefficients given in terms of the underlying graph structure. We then exploit this combinatorial data by again appealing to switchings on the underlying graphs. In contrast to Chapter I, we aim to exploit switchings from the perspective of Markov processes. In particular, we view switchings as a jump process on the space of bipartite graphs and use its generator to study the generator of the matrix-valued Ornstein-Uhlenbeck process. This will allow us to obtain estimates showing the stability of local correlation functions under short-time flows.

In Chapter III, we focus on eigenvalue statistics of the evolved potentials *after* evolving for a fixed time-scale. In particular, we prove Dyson's Conjecture for a wide class of potentials. Here, we explicitly compute the underlying eigenvalue process given by the matrix-valued Brownian motion. To recover eigenvalue statistics from this explicit system of equations, we approximate the eigenvalues by cutting off interaction terms to include only those interactions from nearby eigenvalues. This important idea comes from [16] and requires another local law for evolved matrices, which we also prove in this chapter. The cutoff forces the eigenvalues in the bulk to run a classical DBM for Wigner matrices. This allows us to directly compare the bulk eigenvalues of the evolved potential to those of the GOE using ideas from [17]. This concludes the third-step and thus proves a universality result for bipartite graphs.

Acknowledgements

Table of Contents

Chapter III: Convergence of Local Statistics for Dyson's Brownian Motion for Covariance Matrices

Chapter I: Local law for Biregular Bipartite Graphs

I. The Underlying Model and Main Results

1. Biregular Bipartite Graphs

We begin this chapter with a discussion of the underlying matrix model. We will define the random matrix ensemble later from this combinatorial model. We begin by establishing convention and defining the following graph structure. To make this definition, we now fix integers $M, N > 0$.

Definition 1.1. Suppose $\mathscr{V} = \{1_b, 2_b, \ldots, M_b, 1_w, \ldots, N_w\}$ is a set of labeled vertices, and suppose E is a simple graph on $\mathscr{V}$. We say the graph E is *bipartite* with respect to the vertex sets $(\mathscr{V}, V_b, V_w)$ if $\mathscr{V}$ admits the following decomposition:

$$(1.1) \qquad \mathscr{V} = \{1_b, 2_b, \ldots, M_b\} \bigcup \{1_w, 2_w, \ldots, N_w\} =: V_b \cup V_w,$$

such that for any vertices $v_i, v_j \in V_b$ and $v_k, v_\ell \in V_w$, the edges $v_i v_j$ and $v_k v_\ell$ are not contained in E.

In words, a graph E is bipartite if its vertices may be decomposed into two disjoint sets of vertices V_b and V_w. Moreover, both V_b and V_w, as subgraphs of E, contain no edges. In particular, the graph E cannot contain self-loops.

Remark 1.2. For the remainder of this chapter, we will refer to V_b as the set of *black* vertices and V_w as the set of *white* vertices. We will also keep the following assumptions on the size of V_b and V_w for the remainder of this chapter:

$$(1.2) \qquad |V_b| = M, \quad |V_w| = N.$$

We will introduce the precise constraints on M and N later when discussing the associated random matrix ensemble.

Remark 1.3. As consistent with existing conventions in graph theory, for a graph E we let $[E]$ denote the set of vertices $\mathscr{V}$. For the purposes of Definition 1.1, however, the graph E is stipulated on the labeled vertex set $\mathscr{V}$, hence we need to define and fix the vertex set $\mathscr{V}$ first before we define any graph structure. This last point is subtle and will not be crucial to the discussion on the underlying random matrix problem..

We will now introduce a stronger regularity constraint on bipartite graphs. So far, the bipartite structure introduced in Definition 1.1 has only constrained a qualitative graph structure. The goal of the following definition is to introduce a stronger, combinatorial structure. First, we fix integers $d_b, d_w > 0$ and introduce the following notation.

Notation 1.4. *Suppose $v_i \in V_b$ and $v_k \in V_w$. We define $\mathscr{E}_i$ to be the set of edges $e \in E$ such that $v_i \in e$. Similarly, we define $\mathscr{E}_k$ to be the set of edges $e \in E$ such that $v_k \in e$.*

In particular, $[\mathscr{E}_i]$ consists of exactly those vertices $v \in \mathscr{V}$ incident to the vertex v_i, i.e. the edge $v v_i$ is contained in E.

Definition 1.5. Let $\mathscr{V} = V_b \cup V_w$ denote a labeled vertex set. Suppose E is a bipartite graph on $\mathscr{V}$. We say the graph E is (d_b, d_w)-*regular* if the following two conditions hold.
- For each vertex $v_i \in V_b$, the set $\mathscr{E}_i$ has size d_b.
- For each vertex $v_k \in V_w$, the set $\mathscr{E}_k$ has size d_w.

Remark 1.6. For the remainder of this thesis, we will interchange between the terms (d_b, d_w)-regular and *biregular*. Because this latter terminology suppresses the explicit regularity parameters d_b, d_w, we will use it only when these parameters are fixed or known. Moreover, when we say a graph is biregular, we will imply that it is also bipartite.

Notation 1.7. *For a fixed $M, N > 0$ and $d_b, d_w > 0$, we will define $\Omega = \Omega(M, N, d_b, d_w)$ to be the set of (d_b, d_w)-regular graphs on the vertex set $\mathscr{V}$. Again, we will suppress the dependence of Ω on the parameters M, N, d_b, d_w whenever there is no risk of confusion.*
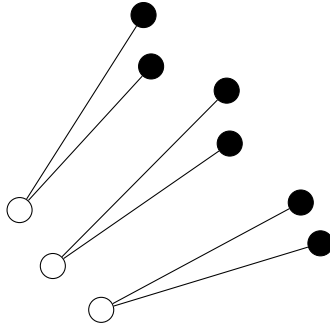
FIGURE 1. A $(1, 2)$-regular graph with $|V_b| = 6$ and $|V_w| = 3$.

In particular, we note the set Ω is finite for any fixed parameters M, N, d_b, d_w.

We end this preliminary discussion on bipartite graphs by recording the following combinatorial identity for biregular graphs which is a direct consequence of Definition 1.1, Definition 1.5, and counting edges:

$$(1.3) \qquad\qquad M d_b = N d_w.$$

Indeed, the LHS counts the number of edges in the graph E by counting vertices in V_b and their neighbors, and the RHS counts vertices in V_w and their neighbors.

2. The Underlying Matrix Model

Before we define the matrix model for biregular graphs, for convenience we first make the following assumption for the remainder of Chapter I. First, we define the following structural parameters:

$$(2.1) \qquad\qquad \alpha := \frac{M}{N}, \quad \gamma := \frac{1}{\alpha}.$$

Assumption 2.1. *For biregular graphs $E \in \Omega$, we assume $\alpha \geqslant 1$, i.e. $M \geqslant N$.*

In particular, by counting edges as in (1.3), Assumption 2.1 implies $d_b \leqslant d_w$. We briefly remark that the regime $M < N$ follows from our discussion under Assumption 2.1 upon a trivial, deterministic relabeling of the vertex set $\mathcal{V}$.

We now introduce the matrix model for biregular graphs, which begins with the following fundamental construction in graph theory.

Definition 2.2. For a biregular graph $E \in \Omega$, we define its *adjacency matrix* $A = (A_{ij})$ to be the $(M + N) \times (M + N)$ matrix whose entries are defined by the edges of E, i.e. for any $i, j \in [[1, M + N]]$, we define

$$(2.2) \qquad\qquad A_{ij} = \begin{cases} 1 & ij \in E \\ 0 & ij \notin E \end{cases},$$

where we interpret ij as the edge consisting of the labeled vertices i, j in $\mathcal{V}$.

Remark 2.3. In the remainder of this chapter, we may interchange between a graph $E \in \Omega$ and its adjacency matrix A when there is no risk of confusion.

In words, the adjacency matrix A is a matrix whose entries are indicator functions for edges in the graph E. We note that an adjacency matrix can be defined for any graph, not necessarily biregular. However, we now exploit the biregular structure in terms of the adjacency matrix A and summarize the results below. In particular, the biregular structure corresponds to spectral information of the adjacency matrix.

Proposition 2.4. *Suppose $E \in \Omega$ is a biregular graph and let A denote its adjacency matrix. Then the following hold:*

- *(I). The adjacency matrix A has the following block representation:*

$$
(2.3) \qquad A = \begin{pmatrix} 0 & H \\ H^* & 0 \end{pmatrix},
$$

  where H is a matrix of size $M \times N$, and the star notation denotes the adjoint of a matrix.

- *(II). The matrix A contains the following (unnormalized) eigenvalue-eigenvector pair:*

$$
(2.4) \qquad \lambda_{\max} = \sqrt{d_b d_w} = d_b \sqrt{\alpha}, \quad \mathbf{e}_{\max} = \begin{pmatrix} 1 \\ 1 \\ \cdots \\ 1 \\ \sqrt{\alpha} \\ \sqrt{\alpha} \\ \cdots \\ \sqrt{\alpha} \end{pmatrix}.
$$

  In particular, the eigenvector $\mathbf{e}_{\max}$ admits a decomposition $\mathbf{e}_{\max} = \begin{pmatrix} \mathbf{e}_b & \mathbf{e}_w \end{pmatrix}^*$, where $\mathbf{e}_b$ is constant and length M, and $\mathbf{e}_w$ is constant and length N.

- *(III). The corresponding eigenspace $V_{\max}$ to $\lambda_{\max}$ is one-dimensional, i.e. $\lambda_{\max}$ is a simple eigenvalue.*

- *(IV). For all other eigenvalues $\lambda \neq \lambda_{\max}$ of the adjacency matrix A, we have the following spectral bound:*

$$
(2.5) \qquad |\lambda| \leqslant C\sqrt{d_w},
$$

  where C is independent of the fundamental parameters M, N, d_b, d_w.

Proof. Statements (I) and (II) follow from the definition of biregular graph and a straightforward calculation. Statements (III) and (IV) are highly nontrivial statements. Statement (III) is a consequence of the Perron-Frobenius Theorem. Statement (IV) follows from a combinatorial interpretation of biregular graphs in terms of trees; for a reference, we cite [8]. □

We now use Proposition 2.4 to construct our matrix ensemble. First, by statements (III) and (IV) in Proposition 2.4, we note the matrix $d_w^{-1/2} A$ has all (but one) eigenvalues bounded by some constant C independent of the fundamental parameters. Moreover, by statement (III) in Proposition 2.4, the following matrix has the same spectrum and eigenspaces as $d_w^{-1/2} A$ except for $V_{\max}$:

$$
(2.6) \qquad X = \begin{pmatrix} 0 & H \\ H^* & 0 \end{pmatrix}, \quad H = d_w^{-1/2}\left(A - \frac{d_b}{N}\mathbf{e}_b \mathbf{e}_w^*\right),.
$$

In particular, X shares the eigenspace $V_{\max}$ but with eigenvalue $\lambda_{\max} = 0$. At least on a heuristic level, because all but one eigenvalue and eigenspace are preserved, the spectral statistics for adjacency matrices of biregular graphs should be insensitive to this deterministic shift.

We may now define our random matrix ensemble as those matrices of the form in (2.6), where A is the adjacency matrix of E. To stipulate a probability measure on this matrix ensemble, we note matrices of the form in (2.6) are in bijective correspondence with biregular graphs $E \in \Omega$, so we may impose a uniform probability measure on matrices (2.6) by imposing the uniform probability measure on the finite set Ω.

Furthermore, with the data in (2.6), we may define two more matrix ensembles of covariance matrices. Precisely, with notation in (2.6), we may define the following two matrices:

$$
(2.7) \qquad X_* := H^*H, \quad X_{*,+} := HH^*.
$$

We briefly remark X_* is a square matrix of dimension $N \times N$, and $X_{*,+}$ is a square matrix of dimension $M \times M$. In particular, under Assumption 2.1, the matrix $X_{*,+}$ is larger, in dimension, than the matrix X_*. This algebraic relation will be important later when discussing spectral properties of the matrices X, X_*, and $X_{*,+}$.

We now formally define our random matrix ensembles as follows for organizational purposes.

Definition 2.5. Retaining the setting of Proposition 2.4, we define $\mathscr{X}$ to be the random matrix ensemble of matrices in the form (2.6), i.e.

$$(2.8) \qquad \mathscr{X} := \left\{ \begin{pmatrix} 0 & H \\ H^* & 0 \end{pmatrix}, \quad H = d_w^{-1/2} \left(A - \frac{d_b}{N} \mathbf{e}_w \mathbf{e}_b^* \right) \right\}.$$

The probability measure on $\mathscr{X}$ is the uniform probability measure induced by the uniform probability measure on Ω. We also define

$$(2.9) \qquad \mathscr{X}_* := \left\{ H^* H : \quad H = d_w^{-1/2} \left(A - \frac{d_b}{N} \mathbf{e}_w \mathbf{e}_b^* \right) \right\},$$

$$(2.10) \qquad \mathscr{X}_{*,+} := \left\{ H H^*, \quad H = d_w^{-1/2} \left(A - \frac{d_b}{N} \mathbf{e}_w \mathbf{e}_b^* \right) \right\}.$$

As with the ensemble $\mathscr{X}$, the probability measures on $\mathscr{X}_*$ and $\mathscr{X}_{*,+}$ are the respective uniform probability measures induced by the uniform probability measure on Ω.

For terminology's sake, we say the ensembles $\mathscr{X}_*$ and $\mathscr{X}_{*,+}$ are the ensembles of *covariance matrices* of dimension N and M, respectively. The ensemble $\mathscr{X}$ is the ensemble of *linearized covariance matrices*.

Before we proceed with discussing the main results of this chapter, we briefly remark here that we will need to understand the correspondence of spectral data, i.e. eigenvalues and eigenvectors, among the three matrix ensembles $\mathscr{X}$, $\mathscr{X}_*$, and $\mathscr{X}_{*,+}$. This discussion will be deferred to a later section in this chapter.

3. Quick Preliminaries on the Stieltjes Transform

In order to state the main results of this chapter, we first recall the following integral transform fundamental throughout the literature in random matrix theory.

Definition 3.1. Suppose $\mu(\mathrm{d}x)$ denotes a probability measure on the real line $\mathbb{R}$. We define the *Stieltjes transform* of $\mu(\mathrm{d}x)$ to be the following complex-valued integral transform

$$(3.1) \qquad s_\mu(z) = \int_{\mathbb{R}} \frac{\mu(\mathrm{d}x)}{x - z}, \quad z \in \mathbb{C}_+.$$

To motivate the use of the Stieltjes transform in terms of random matrices, we first note that the Stieltjes transform admits an inversion formula, implying a sufficiently regular probability measure on the real line is essentially characterized by its Stieltjes transform. For details, we refer to [1] and [2].

Recall, from the introduction, the Marchenko-Pastur law with parameter γ:

$$(3.2) \qquad \varrho_\infty(x) \, \mathrm{d}x := \varrho_{MP}(x) \, \mathrm{d}x = \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{2\pi \gamma x} \mathbf{1}_{x \in [\lambda_-, \lambda_+]} \mathrm{d}x,$$

where we define $\lambda_\pm = (1 \pm \sqrt{\gamma})^2$. We now record the following result, e.g. from [1], which computes the Stieltjes transform of the Marchenko-Pastur law (3.2). We will not include the proof, but briefly remark that it requires nothing more than elementary complex analysis.

Proposition 3.2. *Given the notation above, the Stieltjes transform of the Marchenko-Pastur law in* (3.2) *is given by*

$$(3.3) \qquad m_\infty(z) = \frac{1 - \gamma - z + i\sqrt{(\lambda_+ - z)(z - \lambda_-)}}{2\gamma z}.$$

We also define the following related function, which will be important in stating the main result of this chapter:

$$(3.4) \qquad m_{\infty,+}(z) = \frac{\gamma - 1}{z} + \gamma m_\infty(z),$$
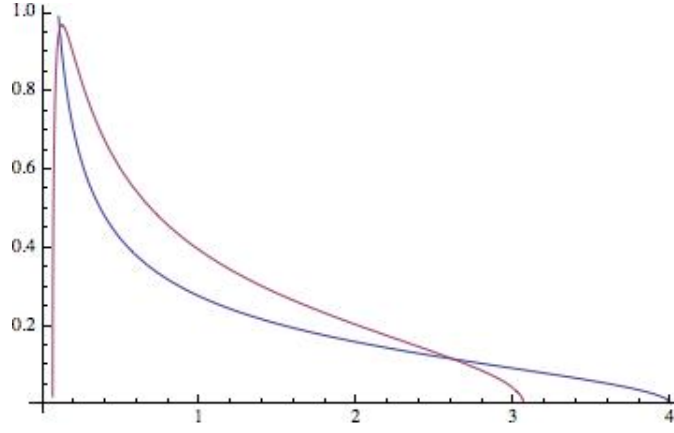
FIGURE 2. The blue curve is the plot of the Marchenko-Pastur Law with $\gamma = 1$. The pink curve is the plot of the Marchenko-Pastur law with $\gamma = 0.75$. We note the singularity at the origin for $\gamma = 1$ only.

which is the Stieltjes transform of the Marchenko-Pastur law ϱ_∞ with an additional weighted point mass at $x = 0$. We now address the Stieltjes transform of the empirical measures of X_* and $X_{*,+}$, respectively. We denote these Stieltjes transforms by s_* and $s_{*,+}$, respectively. By definition, these transforms are given as follows:

$$(3.5) \qquad s_*(z) \;=\; \int_{\mathbb{R}} \frac{1}{x-z} \left(\frac{1}{N} \sum_{i=1}^{N} \delta(x - \lambda_{i,*}) \right) \mathrm{d}x \;=\; \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\lambda_{i,*} - z},$$

$$(3.6) \qquad s_{*,+}(z) \;=\; \int_{\mathbb{R}} \frac{1}{x-z} \left(\frac{1}{M} \sum_{k=1}^{M} \delta(x - \lambda_{k,*,+}) \right) \mathrm{d}x \;=\; \frac{1}{M} \sum_{k=1}^{M} \frac{1}{\lambda_{k,*,+} - z}.$$

Here, the terms $\lambda_{i,*}$ denote the eigenvalues of X_* and the terms $\lambda_{k,*,+}$ denote the eigenvalues of $X_{*,+}$. We will later study how these two spectra correspond to each other.

We proceed by introducing the Green's function of a matrix, which resembles the Green's function of many systems in statistical mechanics; for a reference, we refer to [2].

Definition 3.3. For a real symmetric or complex Hermitian matrix H, we define its *Green's function* to be the following matrix resolvent:

$$(3.7) \qquad G(z; H) \;=\; (H - z)^{-1}, \quad z \in \mathbb{C}_+.$$

Notation 3.4. *For the matrices* $X, X_*,$ *and* $X_{*,+}$, *we define:*

$$(3.8) \qquad G(z) \;:=\; G(z; X), \quad G_*(z) \;:=\; G(z; X_*), \quad G_{*,+}(z; X_{*,+}).$$

We first note the above definition (3.7) of the Green's function makes sense, as the spectrum of H is totally real and $z \notin \mathbb{R}$. Moreover, we now reconcile the Stieltjes transforms $s_*(z)$ and $s_{*,+}(z)$ with the Green's functions of X_* and $X_{*,+}$, respectively, via the following identities:

$$(3.9) \qquad s_*(z) \;=\; \frac{1}{N} \operatorname{Tr} G_*(z), \quad s_{*,+}(z) \;=\; \frac{1}{M} \operatorname{Tr} G_{*,+}(z).$$

We will also define the following *partial* Stieltjes transforms for the linearization X:

$$(3.10) \qquad s_b(z) \;=\; \frac{1}{M} \sum_{i=1}^{M} G_{ii}(z), \quad s_w(z) \;=\; \frac{1}{N} \sum_{k=M+1}^{M+N} G_{kk}(z).$$

Later in this chapter, we will provide a detailed discussion of the properties of a Green's function $G(z; H)$, as well as the Green's functions of covariance matrices $X_*, X_{*,+}$ and their linearization X.

4. The Main Results

We are now in a position to state the main result of this chapter. Recall the fundamental parameters $M, N > 0$ and $d_b, d_w > 0$. In our ensemble, we will take $N \to \infty$. This limit will be taken arbitrarily. We will also stipulate $\alpha \to \alpha_\infty$ for consistency with the existing literature (i.e. [8]), i.e. the ratio α converges to a real number α_∞ as $N \to \infty$. However, this last assumption on $\alpha \to \alpha_\infty$ will be unnecessary for this chapter and necessary only in later chapters. Here we will only need to assume $\alpha = O(1)$.

To state the main theorem, we will first need to define the following subdomain of the complex plane for any fixed $\varepsilon > 0$:

$$(4.1) \qquad U_{\varepsilon,\pm} := \{z = E + i\eta : |E| > \varepsilon, \ \eta > 0\},$$

$$(4.2) \qquad U_\varepsilon := U_{\varepsilon,+} \cup U_{\varepsilon,-}.$$

The repulsion from the origin in the complex plane defined in the domains $U_{\varepsilon,\pm}$ and U_ε are to avoid the singularity of, for example, the transform $m_{\infty,+}$. We will also need to define the following control parameters:

$$(4.3) \qquad D := d_b \wedge \frac{N^2}{d_b^3},$$

$$(4.4) \qquad \Phi(z) := \frac{1}{\sqrt{N\eta}} + \frac{1}{\sqrt{D}},$$

$$(4.5) \qquad F_z(r) = F(r) := \left[\left(1 + \frac{1}{\sqrt{(\lambda_+ - z)(z - \lambda_-)}} \right) r \right] \wedge \sqrt{r}.$$

We may now state the main theorem of this chapter, which compares the Green's functions G, G_*, and $G_{*,+}$ to those of limiting distributions at microscopic scales $\eta = \operatorname{Im} z$. This will be made precise in the statement of the result.

Theorem 4.1. *Suppose $\xi = \xi_N$ is a parameter chosen such that the following growth conditions on D and η hold:*

$$(4.6) \qquad \xi \log \xi \gg \log^2 N, \quad |\eta| \gg \frac{\xi^2}{N}, \quad D \gg \eta^2.$$

Then for any fixed $\varepsilon > 0$, we have the following estimates with probability at least $1 - e^{-\xi \log \xi}$, uniformly over all $z = E + i\eta \in U_\varepsilon$ with η satisfying the growth condition in (4.6):

$$(4.7) \qquad \max_i |[G_*(z)]_{ii} - m_\infty(z)| = O\left(F_z(\xi\Phi)\right), \quad \max_{i \neq j} |[G_*(z)]_{ij}| = O\left(\frac{\xi\Phi(z^2)}{z}\right).$$

Similarly, for any fixed $\varepsilon > 0$, we have the following estimates with probability at least $1 - e^{-\xi \log \xi}$, uniformly over all parameters $z = E + i\eta \in U_\varepsilon$ with η satisfying the growth condition in (4.6):

$$(4.8) \qquad \max_i |[G_{*,+}(z)]_{ii} - m_{\infty,+}(z)| = O\left(F_z(\xi\Phi)\right), \quad \max_{i \neq j} |[G_{*,+}(z)]_{ij}| = O\left(\frac{\xi\Phi(z^2)}{z}\right).$$

Conditioning on the estimates (4.7) and (4.8), respectively, uniformly over $z = E + i\eta \in U_\varepsilon$ with η satisfying the growth condition in (4.6), we have

$$(4.9) \qquad |s_*(z) - m_\infty(z)| = O\left(F_z(\xi\Phi)\right), \quad |s_{*,+}(z) - m_\infty(z)| = O\left(F_z(\xi\Phi)\right).$$

Conditioning on the estimates (4.7), uniformly over all $z = E + i\eta$ with η satisfying the growth condition in (4.6), we have

$$(4.10) \qquad \max_{k > M} \left| [G(z)]_{kk} - z m_\infty(z^2) \right| = O\left(z F_{z^2}(\xi\Phi(z^2))\right),$$

$$(4.11) \qquad \max_{M < k < \ell} |[G(z)]_{k\ell}| = O\left(\xi\Phi\right).$$

Moreover, conditioning on the estimates (4.8), for any fixed $\varepsilon > 0$, uniformly over all $z = E + i\eta \in U_\varepsilon$ with η satisfying the growth condition in (4.6), we have

$$(4.12) \qquad \max_{i \leqslant M} \left| [G(z)]_{ii} - z m_{\infty,+}(z^2) \right| = O\left(z F_{z^2}(\xi\Phi(z^2))\right),$$

$$(4.13) \qquad \max_{i < j \leqslant M} |[G(z)]_{ij}| = O(\xi\Phi).$$

Conditioning on (4.7) and (4.8), we have the following estimates uniformly over all $z = E + i\eta$ with η satisfying the growth condition in (4.6):

$$(4.14) \qquad \left|s_b(z) - zm_{\infty,+}(z^2)\right| = O\left(zF_{z^2}(\xi\Phi(z^2))\right),$$

$$(4.15) \qquad \left|s_w(z) - zm_{\infty}(z^2)\right| = O\left(zF_{z^2}(\xi\Phi(z^2))\right).$$

Lastly, the estimates (4.7) and (4.9) hold without the condition $|E| > \varepsilon$ if $\alpha > 1$. The estimates (4.12) and (4.13) hold without the condition $|E| > \varepsilon$ if $\alpha = 1$.

Remark 4.2. We briefly remark here that the estimates obtained in Theorem 4.1 hold at the optimal scale $\eta \gg \xi^2/N$, up to log factors. This is an estimate at the microscopic scale, i.e. the scale at which eigenvalues fluctuate, and is a strong improvement of the scale $N^{-\delta}$ for small $\delta > 0$ obtained in [8]. Moreover, under certain structural conditions the estimates obtained in Theorem 4.1 hold for all energies E; in [8], the repulsion from the origin was necessary.

Remark 4.3. We briefly remark on the repulsion assumption $|E| > \varepsilon$ in Theorem 4.1. The removal of this assumption discussed at the end of the statement of Theorem 4.1 is a direct consequence of studying the dependence of the singularities of the Green's functions and Stieltjes transforms at the origin with respect to the structural parameter α. For example, the presence of a singularity of m_{∞} at the origin occurs exactly when $\alpha = 1$. Moreover, the singularities in the Stieltjes transforms of matrices and the singularities of $m_{\infty,+}$ at the origin cancel each other out, allowing for a regularization at the origin. This will all be discussed in more detail in a later section when studying the Stieltjes transforms and Green's functions in more detail.

Remark 4.4. We last remark that if $\alpha = 1$, the covariance matrices X_* and $X_{*,+}$ are equal in law. This comes from symmetry of the bipartite graph between the two vertex sets V_b and V_w, i.e. the graph statistics are unchanged upon relabeling the graph. This allows us to remove the assumption $|E| > \varepsilon > 0$ for certain estimates in Theorem 4.1 in the regime $\alpha = 1$.

We now discuss important consequences of Theorem 4.1, the first of which is the following result on eigenvector delocalization, i.e. an estimate on the ℓ^{∞}-norm of an eigenvector in terms of its ℓ^2-norm. The proof of this delocalization result will be delegated to a later section after we study in more detail the spectral data of covariance matrices and their linearizations.

Corollary 4.5. (Eigenvector Delocalization).

Assume the setting of Theorem 4.1, and suppose $\mathbf{u}$ *is an eigenvector of* X_* *with eigenvalue* λ. *Then with probability at least* $1 - e^{-\xi\log\xi}$, *we have*

$$(4.16) \qquad \|\mathbf{u}\|_{\ell^{\infty}} = O\left(\frac{\xi}{\sqrt{N}}\|\mathbf{u}\|_{\ell^2}\right).$$

We briefly remark that the eigenvector delocalization fails for the larger covariance matrix $X_{*,+}$. This is because of the singularities in the Green's function at the origin in the complex plane. This will be made clearer when we give the proof of Corollary 4.5.

We conclude this discussion concerning consequences of Theorem 4.1 with the following result concerning estimates on the underlying probability measures deduced from estimates on the Stieltjes transforms as given in Theorem 4.1; these are known as *rigidity estimates*. The proof is quite technical and strays from the scope of this thesis, so we omit it. We briefly remark that it relies heavily upon the Helffer-Sjostrand formula and functional calculus, and beyond these tools, the local law in Theorem 4.1. To state the result, we first introduce the following definition.

Definition 4.6. For each $i \in [[1, N]]$, we define the i-th *classical location*, denoted γ_i, by the following quantile formula:

$$(4.17) \qquad \frac{i}{N} = \int_{-\infty}^{\gamma_i} \varrho_{\infty}(E)\, \mathrm{d}E,$$

where we recall ϱ_{∞} denotes the density function of the Marchenko-Pastur law.

The following consequence of Theorem 4.1 will compare the classical location γ_i to the i-th eigenvalue λ_i of the covariance matrix X_*, where the ordering on the eigenvalues is the increasing order.

Corollary 4.7. *For any fixed $\kappa > 0$ and index $i \in [[\kappa N, (1 - \kappa)N]]$, we have, with probability at least $1 - e^{-\xi \log \xi}$,*

(4.18)
$$|\lambda_i - \gamma_i| \; = \; O\left(\frac{\xi^2}{D^{1/4}}\right).$$

For details of the proof, we refer to Section 5 in [3] and Section 7 in [17].

5. Outline of the Proof of Theorem 4.1

We now give an outline for the derivation of the local law. The proof will roughly consist of the following three steps:

- (I). The first step will be to study the Green's functions G, G_* and $G_{*,+}$ off-the-bat. In particular, we begin by rewriting the Green's function of an arbitrary real symmetric or complex Hermitian matrix in terms of its spectral data. Given this spectral representation, we then aim to understand the correspondence among the spectral data of the covariance matrices $X_*, X_{*,+}$ and the linearization X. This step will also require a preliminary analysis of Green's functions motivated by ideas from [4].

- (II). The second step will be to adapt the methods in [4] to define and study a method of resampling graphs in Ω. The resampling will be generated by local operations on a given graph known as *switchings*, which we will define more precisely in a later section. The local nature of the resampling method will help us derive equations exploiting the probabilistic stability of the Green's function under these switchings.

- (III). The third step will be to derive precisely an approximate self-consistent equation for the diagonal entries of the Green's function and study its stability properties. As in [4], this will help us compare the diagonal of the Green's function to the associated Stieltjes transform. The equation in [4], however, contains a constant leading-order coefficient whereas for covariance matrices the leading-order coefficient is nonconstant. We adapt the methods suitably to handle this nonlinearity.

II. Preliminary Analysis of the Green's Function

We now aim our discussion towards reducing the proof of Theorem 4.1 to proving a local law focusing only on the domain U_ε and for only the covariance matrix ensemble X_*. As a roadmap, we provide a summary of these reductions at the end of this discussion.

6. Resolvent Theory for Real Symmetric Matrices

We begin with a brief overview of resolvent theory for matrices. Before we discuss any results, because the following discussion will hold for Green's functions of a widely general class of matrices, we will not specify which matrix ensemble we are concerned with, i.e. the covariance matrices $X_*, X_{*,+}$ or the linearization X. To this end, we will first introduce notation for when we apply these results to these specific matrix ensembles.

Notation 6.1. *Suppose F is a function of the Green's function or matrix entries of matrices belonging to any one of the matrix ensembles X, X_*, or $X_{*,+}$. Then we establish the notation $F_\star$ to be the function obtained when restricted to the matrix ensemble $X_\star$, where we take $\star$ to be blank or $\star = *$ or $\star = *, +$.*

The first result is the following perturbation estimate on Green's functions resulting from perturbations in matrix entries, known as the *resolvent identity*, whose proof is a standard result in linear algebra.

Lemma 6.2. (Resolvent Identity)
Suppose A and B are invertible matrices. Then we have

$$(6.1) \qquad A^{-1} - B^{-1} \;=\; A^{-1}(B - A)B^{-1}.$$

In particular, if H and $\widetilde{H}$ denote real symmetric or complex Hermitian matrices with Green's functions $G(z)$ and $\widetilde{G}(z)$, respectively, for $z \notin \mathbb{R}$, then

$$(6.2) \qquad G(z) - \widetilde{G}(z) \;=\; \left[G\left(\widetilde{H} - H \right) \widetilde{G} \right](z).$$

Remark 6.3. Typically in this thesis, we will use Lemma 6.2 for $\widetilde{H}$ a tractable perturbation of the matrix entries of H.

An important consequence of Lemma 6.2 that will be used in estimating ℓ^2-averages of Green's functions is the following *Ward Identity*, whose proof follows directly from (6.2) with the choice of Green's functions $G(z; H)$ and $\widetilde{G}(z) = \overline{G(z; H)}$.

Corollary 6.4. (Ward Identity)
Suppose H is a real symmetric matrix of size N with Green's function $G(z)$. Then for any fixed row index $i \in [[1, N]]$,

$$(6.3) \qquad \sum_{k=1}^{N} |G_{ik}(E + i\eta)|^2 \;=\; \frac{\operatorname{Im} G_{ii}(E + i\eta)}{\eta}.$$

In particular, we obtain the following a priori estimate for any matrix index (i, j):

$$(6.4) \qquad |G_{ij}(E + i\eta)| \;\leqslant\; \frac{1}{\eta},$$

and thus for any matrix index (i, j), the function $G_{ij}(z)$ is locally Lipschitz with constant η^{-2}.

The third and final preliminary result we give is the following representation of the Green's function $G(z; H)$ in terms of the spectral data of H. This spectral representation will be indispensable for exploiting the rich spectral correspondence among covariance matrices and their linearizations.

Lemma 6.5. (Spectral Representation)

Suppose H is a real-symmetric or complex-Hermitian matrix with eigenvalue-eigenvector pairs $\{(\lambda_\alpha, \mathbf{u}_\alpha)\}_\alpha$, and let $G(z)$ denote its Green's function. Then for any matrix index (i, j), we have

$$(6.5) \qquad G_{ij}(z) = \sum_{\alpha=1}^{N} \frac{\mathbf{u}_\alpha(i)\overline{\mathbf{u}_\alpha}(j)}{\lambda_\alpha - z},$$

where the overline notation denotes the complex conjugate of the vector entry. In particular, the Green's function is complex Hermitian.

Proof. If H is a diagonal matrix, the result is clear. If H is not diagonal, suppose $\Lambda = UHU^*$ is its diagonalization with a unitary matrix U. Thus, we obtain

$$(6.6) \qquad G(z) = U G_{\mathrm{diag}}(z) U^*,$$

where $G_{\mathrm{diag}}(z)$ is a diagonal matrix with diagonal entries $(\lambda_\alpha - z)^{-1}$. The spectral representation then follows from multiplying the matrices on the RHS, using the construction of the unitary matrix U in terms of the eigenvectors $\{\mathbf{u}_\alpha\}_\alpha$. $\qquad \square$

We now proceed with the fundamental iteration scheme taken from [4] which will allow us to control the Green's function of a real symmetric or complex Hermitian matrix at smaller scales in terms of larger scales. Here, as always in this chapter, the scale will refer to the imaginary part η of the parameter $z = E + i\eta$. The important ingredients will be the resolvent identity in Lemma 6.2 and the Ward identity in Corollary 6.4. First, to state the result, we define the following control parameters for the Green's functions:

$$\Gamma(E + i\eta) = \left[\max_{i,j} |G_{ij}(z)| \right] \vee 1,$$

$$\Gamma^*(E + i\eta) = \sup_{\eta' \geqslant \eta} \Gamma(E + i\eta').$$

Lemma 6.6. For any $z = E + i\eta \in \mathbb{C}_+$, the function $\Gamma(z)$ is locally Lipschitz continuous in η with the following bound on its almost-everywhere derivative:

$$(6.7) \qquad |\partial_\eta \Gamma(z)| \leqslant \frac{\Gamma(z)}{\eta}.$$

In particular, for any $\kappa > 1$ and $z = E + i\eta \in \mathbb{C}_+$, we have

$$(6.8) \qquad \Gamma\left(E + i\frac{\eta}{\kappa}\right) \leqslant \kappa \Gamma(E + i\eta).$$

Proof. To derive (6.8) from (6.7), we first note the Lipschitz bound (6.7) implies

$$(6.9) \qquad \partial_\eta \left(\eta \Gamma(E + i\eta) \right) = \eta \partial_\eta \Gamma(z) + \Gamma(z) \geqslant 0.$$

This implies (6.8) upon taking $\eta_1 = \eta$ and $\eta_2 = \eta/\kappa$; here we used $\eta_2 < \eta_1$. It now suffices to prove the Lipschitz condition and the bound (6.7). To do so, for a small $h > 0$, we use the resolvent identity (6.1) and deduce, for any matrix index (i, j),

$$(6.10) \qquad |G_{ij}(z + ih) - G_{ij}(z)| \leqslant h \sum_{k=1}^{N} |G_{ik}(z + ih)| \, |G_{kj}(z + ih)|$$

$$(6.11) \qquad \leqslant h \left(\sum_{k=1}^{N} |G_{ik}(z + h)|^2 \right)^{1/2} \left(\sum_{k=1}^{N} |G_{kj}(z + h)|^2 \right)^{1/2},$$

where we used the Schwarz inequality in the last line. By the Ward Identity (6.3), and the complex-Hermitian structure of $G(z)$ as obtained in Lemma 6.5, we deduce

$$(6.12) \qquad |G_{ij}(z + ih) - G_{ij}(z)| \leqslant h \sqrt{\frac{\Gamma(z + ih)\Gamma(z)}{(\eta + h)\eta}}.$$

Because the Green's function is continuous, so is the control parameter Γ. Taking a maximum over all matrix indices (i, j), we thus deduce the Lipschitz condition and the derivative estimate (6.7). $\qquad \square$

7. Spectral Theory for Covariance Matrices

We now redirect our focus to Green's functions of covariance matrices. More precisely, our goal in this discussion is to prove a correspondence among the spectral theory of the covariance matrices X_* and $X_{*,+}$ and the linearization X.

Remark 7.1. Before we begin, we note here that the results to be given concerning the aforementioned spectral correspondence hold for a general class of covariance matrices, not just those coming from biregular graphs studied in this thesis. Moreover, the proofs for the results will not appeal to any graph structure underlying the covariance matrices $X_*, X_{*,+}$ or the linearization X. However, for simplicity, we will retain the notation $X_*, X_{*,+}$ and X, and the reader is invited to think exclusively of matrices associated to biregular graphs.

We now record the fundamental result which bridges together the spectral data of the three matrices X_*, $X_{*,+}$ and X. We reemphasize the crucial utility of this result to this thesis.

Proposition 7.2. *Suppose H is a real-valued matrix of size $M \times N$ with $M \geqslant N$, and suppose X is a block matrix of the following form:*

$$(7.1) \qquad X = \begin{pmatrix} 0 & H \\ H^* & 0 \end{pmatrix}.$$

Then the following hold:

- (I). *The spectrum of X admits the following decomposition:*

$$(7.2) \qquad \sigma(X) = \sigma^{1/2}(H^*H) \cup \zeta(X),$$

  where $\sigma^{1/2}(H^*H)$ denotes the pairs of eigenvalues $(\pm\lambda)$ such that $(\pm\lambda)^2$ is an eigenvalue of H^*H. Here, $\zeta(X)$ denotes the set of eigenvalues not in $\sigma^{1/2}(H^*H)$, all of which are 0.
- (II). *The spectrum of HH^* admits the following decomposition:*

$$(7.3) \qquad \sigma(HH^*) = \sigma(H^*H) \cup \zeta^2(X),$$

  where $\zeta^2(X)$ denotes the set of eigenvalues not in $\sigma(H^*H)$, all of which are 0.
- (III). *Suppose $\lambda^2 \in \sigma(H^*H)$ is associated to the following ℓ^2-normalized eigenvectors:*

$$(7.4) \qquad \mathbf{v}_* \leftrightsquigarrow H^*H, \quad \mathbf{v}_{*,+} \leftrightsquigarrow HH^*.$$

  Then $\pm\lambda$ is associated to the following ℓ^2-normalized eigenvector pair of X:

$$(7.5) \qquad \pm\lambda \leftrightsquigarrow \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{v}_{*,+} \\ \pm\mathbf{v}_* \end{pmatrix}.$$

- (IV). *Conversely, any eigenvalue pair $\pm\lambda \in \sigma^{1/2}(H^*H)$ is associated to the following ℓ^2-normalized eigenvector pair of X:*

$$(7.6) \qquad \pm\lambda \leftrightsquigarrow \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{v}_{*,+} \\ \pm\mathbf{v}_* \end{pmatrix},$$

  where $\mathbf{v}_{*,+}$ is an ℓ^2-normalized eigenvector of HH^* with eigenvalue λ^2 and $\mathbf{v}_*$ is an ℓ^2-normalized eigenvector of H^*H with eigenvalue λ^2.
- (V). *Suppose $\lambda = 0 \in \zeta^2(X)$ is associated to the ℓ^2-normalized eigenvector $\mathbf{v}_\lambda$ of HH^*. Then for some $\lambda' = 0 \in \zeta(X)$, the corresponding ℓ^2-normalized eigenvector is given by*

$$(7.7) \qquad \lambda' \leftrightsquigarrow \begin{pmatrix} \mathbf{v}_\lambda \\ 0 \end{pmatrix}.$$

- (VI). *Conversely, suppose $\lambda \in \zeta(X)$. Then $\lambda = 0$ is associated to the following ℓ^2-normalized eigenvector of X:*

(7.8)
$$\lambda \;\longleftrightarrow\; \begin{pmatrix} \mathbf{v}_\lambda \\ 0 \end{pmatrix},$$

where $\mathbf{v}_\lambda$ is an ℓ^2-normalized eigenvector of HH^* with eigenvalue $\lambda' = 0$.

Before we proceed with the proof of Proposition 7.2, we briefly explain its meaning in words. Proposition 7.2, effectively, decomposes the spectrum of the linearization X into the spectrum of H^*H and a trivial set of eigenvalues. This, in turn, may be realized as the spectrum of the larger covariance matrix HH^*. In particular, the only nonzero and interesting eigenvalues among these three matrices are those eigenvalues coming from the smaller covariance matrix H^*H.

We also briefly remark for each set of eigenvalues in the decomposition of $\sigma(X)$, Proposition 7.2 gives a correspondence between the eigenvectors as well. Essentially, the eigenvectors of the linearization X are composed of the eigenvectors corresponding to the same eigenvalue of the covariance matrices H^*H and HH^*.

Both of these ingredients, i.e. for eigenvalues and eigenvectors, are indispensable for studying the corresponding Green's functions $G_*, G_{*,+}$ and G. The main tool, on the side of Green's functions, for exploiting Proposition 7.2 will be the spectral representation of the Green's function (6.5) in Lemma 6.5. This discussion will take place after the proof of Proposition 7.2.

Proof. (of Proposition 7.2). We organize the proof of Proposition 7.2 by grouping together its statements. In the following we do not provide all explicit details as they follow from elementary linear algebra.

- Proof of Statements (I) – (II): This is a consequence of the SVD (singular value decomposition) of the matrix H and dimension-counting.
- Proof of Statements (III) – (VI): This is a consequence of direct calculation and dimension-counting.

$\square$

We are now in a position to deduce eigenvector delocalization, which we recall bounds, in high probability, the ℓ^∞-norm an eigenvector of X_* in terms of its ℓ^2-norm.

Proof. (of Eigenvector Delocalization).

First, we note in the case $\mathbf{u} \in \mathrm{Span}(\mathbf{e}_b)$, the result is true trivially. Moreover, by Proposition 7.2, it suffices to prove the claim for eigenvectors of the linearization X, replacing the ℓ^∞-norm by a supremum over indices $k \in [[M, M+N]]$.

We now take for granted $|zm_\infty(z^2)| = O(1)$ uniformly for $z = E + i\eta \in \mathbb{C}_+$; this follows from an elementary analysis of the Stieltjes transform we will discuss shortly. This allows us to obtain the following string of inequalities with probability at least $1 - e^{-\xi \log \xi}$ and any index $k \in [[M, M+N]]$:

(7.9)
$$|\mathbf{u}(k)|^2 \;\leqslant\; \sum_{\mathbf{v}_\beta \neq \mathbf{u}} \frac{\eta^2 |\mathbf{v}_\beta(k)|^2}{(\lambda_\beta - \lambda)^2 + \eta^2}$$

(7.10)
$$= \; \eta \, \mathrm{Im}[G(\lambda + i\eta)]_{kk}$$

(7.11)
$$\leqslant \; \eta \, |zm_\infty(z^2)| + O(\eta\sqrt{\xi\Phi})$$

(7.12)
$$\leqslant \; 2\eta,$$

where we used the local law for the linearization X to estimate the second line. This completes the derivation of the eigenvector delocalization. $\square$

We now use Proposition 7.2 and the spectral representation (6.5) to deduce the following relation among the Green's functions $G(z), G_*(z)$, and $G_{*,+}(z)$ for any $z \in \mathbb{C}_+$.

Lemma 7.3. *Suppose X is a block matrix of the form* (7.1), *and suppose $i, j \in [[1, M + N]]$ are indices chosen such that either $i, j \leqslant M$ or $i, j > M$. Then, for any $z = E + i\eta \in \mathbb{C}_+$, we have*

$$
(7.13) \qquad G_{ij}(z) = \begin{cases} zG_{*,+}(z^2) & i, j \leqslant M \\ zG_*(z^2) & i, j > M \end{cases}.
$$

Proof. For simplicity, we suppose X is real symmetric as the proof for complex Hermitian matrices is similar. First suppose $i, j \leqslant M$. By the spectral representation in (6.5) and Proposition 7.2, we obtain

$$
(7.14) \qquad G_{ij}(z) = \sum_{\alpha} \frac{\mathbf{u}_\alpha(i)\mathbf{u}_\alpha(j)}{\lambda_\alpha - z} = \sum_{\lambda \in \sigma(HH^*)} \frac{1}{2} \left(\frac{\mathbf{u}_\alpha(i)\mathbf{u}_\alpha(j)}{\sqrt{\lambda_\alpha} - z} + \frac{\mathbf{u}_\alpha(i)\mathbf{u}_\alpha(j)}{-\sqrt{\lambda_\alpha} - z} \right),
$$

where the last equality holds by abuse of notation for eigenvectors of the covariance matrix HH^* versus the linearization X. This completes the derivation for the case $i, j \leqslant M$. The proof for the case $i, j > M$ follows by the exact same calculation, but instead taking a summation over $\sigma(H^*H)$ and noting the eigenvector terms $\mathbf{u}_\alpha(i)\mathbf{u}_\alpha(j)$ vanish for $\lambda_\alpha \in \zeta(X)$ by Statements (V) and (VI) in Proposition 7.2. $\qquad \square$

We now proceed with extending and improving upon the relations given in Lemma 7.3 for the Stieltjes transforms. In particular, upon averaging over eigenvector indices for the Stieltjes transforms s, s_*, and $s_{*,+}$, we may forget about eigenvectors and instead focus solely on the correspondence of the spectra among X, X_*, and $X_{*,+}$. This will give us an extension of Lemma 7.3. The proof of this following result is a direct consequence of Proposition 7.2 (statements (V) and (VI)) and Lemma 7.3.

Corollary 7.4. *Suppose X is a block matrix of the form* (7.1), *and define the following Stieltjes transforms (as done previously in this chapter) for any $z = E + i\eta \in \mathbb{C}_+$:*

$$
(7.15) \quad s_b(z) = \frac{1}{M} \sum_{i=1}^{M} G_{ii}(z), \quad s_w(z) = \frac{1}{N} \sum_{k=M+1}^{M+N} G_{kk}(z), \quad s_{*,+}(z) = \frac{1}{M} \operatorname{Tr} G_{*,+}(z), \quad s_*(z) = \frac{1}{N} \operatorname{Tr} G_*(z).
$$

Then we have the following relations:

$$
(7.16) \qquad\qquad s_b(z) = zs_{*,+}(z^2), \quad s_w(z) = zs_*(z^2),
$$

$$
(7.17) \qquad\qquad s_b(z) = \frac{\gamma - 1}{z} + \gamma s_w(z).
$$

We now remark that Lemma 7.3 and Corollary 7.4 allow us to go between the Stieltjes transforms s, s_*, and $s_{*,+}$, and also between the Green's function entries themselves. This will be important in reducing the proof of the local laws for the linearization X to the local laws for the covariance matrices X_* and $X_{*,+}$. Moreover, this will help us compute the partial Stieltjes transforms s_b and s_w of the linearization X using the Stieltjes transforms of X_* and $X_{*,+}$. We will make this precise shortly, beginning with the following result which will serve as the first reduction of the proof of the local laws.

Lemma 7.5. *Assuming the setting of Theorem 4.1, then the following two estimates are equivalent:*

- (I). *For any fixed $\varepsilon > 0$, we have with probability at least $1 - e^{-\xi \log \xi}$, uniformly over $z = E + i\eta \in U_\varepsilon$ with $\eta \gg \xi^2/N$,*

$$
(7.18) \qquad \max_i |[G_*(z)]_{ii} - m_\infty(z)| = O\left(F_z(\xi\Phi)\right), \quad \max_{i \neq j} |[G_*(z)]_{ij}| = O(\xi\Phi).
$$

- (II). *For any fixed $\varepsilon > 0$, we have with probability at least $1 - e^{-\xi \log \xi}$, uniformly over $z = E + i\eta \in U_\varepsilon$ with $\eta \gg \xi^2/N$,*

$$
(7.19) \qquad \max_{k > M} |[G(z)]_{kk} - zm_\infty(z^2)| = O\left(zF_{z^2}(\xi\Phi(z^2))\right),
$$

$$
(7.20) \qquad \max_{M < k < \ell} |[G(z)]_{k\ell}| = O\left(z\xi\Phi(z^2)\right).
$$

Similarly, the above equivalence holds replacing G_* with $G_{*,+}$ and taking the maximums over $i \leqslant M$ and $i, j \leqslant M$.

Proof. This follows immediately from Lemma 7.3. $\qquad \square$

We recall the Marchenko-Pastur density function with parameter $\gamma \leqslant 1$ and its Stieltjes transform as follows:

$$(8.1) \qquad \varrho_\infty(E)\,\mathrm{d}E \;=\; \frac{\sqrt{(\lambda_+ - E)(E - \lambda_-)}}{2\pi\gamma E}\,\mathbf{1}_{E\in[\lambda_-,\lambda_+]}\,\mathrm{d}E,$$

$$(8.2) \qquad m_\infty(z) \;=\; \frac{1 - \gamma - z + i\sqrt{(\lambda_+ - z)(z - \lambda_-)}}{2\gamma z},$$

where $\lambda_\pm = (1 \pm \sqrt{\gamma})^2$. By a straightforward calculation, we may deduce the following *self-consistent* equation for the Stieltjes transform (for details, see [1]):

$$(8.3) \qquad m_\infty(z) \;=\; \frac{1}{1 - \gamma - z\gamma m_\infty(z) - z}.$$

We also recall the following auxiliary integral transform:

$$(8.4) \qquad m_{\infty,+}(z) \;=\; \frac{\gamma - 1}{z} + \gamma m_\infty(z),$$

which may be realized as the Stieltjes transform of the scaled Marchenko-Pastur law $\gamma\varrho_\infty$ with an additional term given by uniformly weighted point masses all concentrated at $E = 0$. Precisely, we have

$$(8.5) \qquad m_{\infty,+}(z) \;=\; \int_{\mathbb{R}} \frac{1}{x - z}\left(\gamma\varrho_\infty(x)\mathrm{d}x \;+\; \frac{1}{M}\sum_{\lambda\in\zeta^2(X)}\delta(x - \lambda)\right).$$

With this representation of the auxiliary transform $m_{\infty,+}$ and Proposition 7.2, we may deduce the next reduction of the proof of the local law which follows immediately by Lemma 7.3 and the definition of the auxiliary transform $m_{\infty,+}$.

Lemma 8.1. *Assuming the setting of Theorem 4.1, the following estimates are equivalent for any $z \in \mathbb{C}_+$:*

$$(8.6) \qquad \left|s_b(z) - zm_{\infty,+}(z^2)\right| \;=\; O\left(zF_{z^2}(\xi\Phi(z^2))\right),$$

$$(8.7) \qquad \left|s_w(z) - zm_\infty(z^2)\right| \;=\; O\left(zF_{z^2}(\xi\Phi(z^2))\right).$$

Similarly, the following estimates are equivalent for any $z \in \mathbb{C}_+$:

$$(8.8) \qquad \left|s_*(z) - m_\infty(z)\right| \;=\; O\left(F_z(\xi\Phi)\right),$$

$$(8.9) \qquad \left|s_{*,+}(z) - m_{\infty,+}(z)\right| \;=\; O\left(F_z(\xi\Phi)\right).$$

Proof. The singularities in $s_b(z)$ and $zm_{\infty,+}(z^2)$ cancel each other. Precisely, by Corollary 7.4 we have the formula

$$(8.10) \qquad s_b(z) - zm_{\infty,+}(z^2) \;=\; \gamma\left(s_w(z) - zm_\infty(z^2)\right).$$

This allows us to deduce the equivalence of the first two estimates. The equivalence of the last two estimates follows via an analogous cancellation of the singularities of $s_{*,+}$ and $m_{\infty,+}$; this is a consequence of Proposition 7.2 and the definition of the Stieltjes transform of the empirical measure of a real symmetric matrix. $\qquad\square$

Remark 8.2. Before we proceed, we note that Lemma 8.1 improves upon Lemma 7.5 in that it removes the restriction $|E| > \varepsilon$ on the energy. This is a consequence of the averaging phenomenon in defining the Stieltjes transform. In particular, we may only remove the energy assumption after averaging away the dependence on the eigenvectors.

We now study the modified Stieltjes transform $zm_\infty(z^2)$; this is the integral transform used for comparison for the linearization X. To begin, we first construct a probability measure on the real line whose density is induced by the Marchenko-Pastur density upon the change of variables $E \mapsto \sqrt{E}$. Precisely, we define the following density for the *linearized Marchenko-Pastur Law*:

$$(8.11) \qquad \varrho(E) \;=\; \begin{cases} \frac{\gamma}{(1+\gamma)\pi|E|}\sqrt{(\lambda_+ - E^2)(E^2 - \lambda_-)} & E^2 \in [\lambda_-, \lambda_+] \\ 0 & E^2 \notin [\lambda_-, \lambda_+] \end{cases}.$$
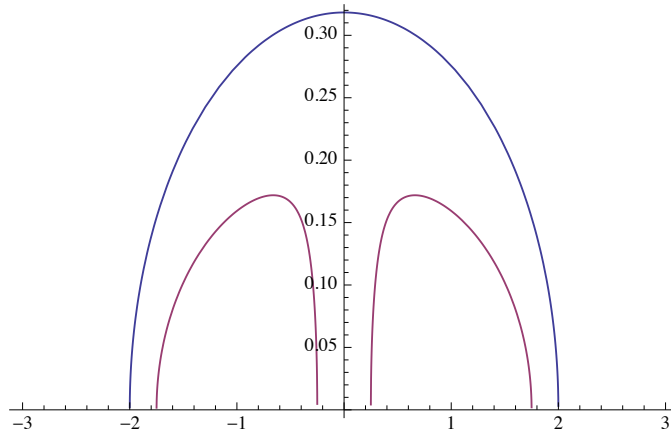
FIGURE 3. The blue curve is the plot of the linearized Marchenko-Pastur Law with $\lambda = 1$; this is the famous Wigner semicircle law. The pink curve is the plot of the linearized Marchenko-Pastur law with $\lambda = 1.25$.

In particular, the density ϱ is related to the Marchenko-Pastur density with parameter γ by the following change of variables relation:

$$(8.12) \qquad \varrho(E^2)\mathrm{d}E^2 \; = \; \varrho_\infty(E)\mathrm{d}E.$$

Similarly, we may define the probability measure with additional point masses at $E = 0$, i.e.

$$(8.13) \qquad \varrho_+(E) \; = \; (\gamma - 1)\delta(E) + \gamma\varrho(E).$$

Lastly, we define $m(z)$ to be Stieltjes transform of ϱ, and we define $m_+(z)$ to be the Stieltjes transform of ϱ_+. With these definitions, by essentially the proof of Corollary 7.4, we deduce the following identities:

$$(8.14) \qquad m(z) \; = \; zm_\infty(z^2), \quad m_+(z) \; = \; zm_{\infty,+}(z^2).$$

Thus, the modified transforms $zm_\infty(z^2)$ and $zm_{\infty,+}(z^2)$ may be realized as honest Stieltjes transforms. It now suffices to study the Stieltjes transforms $m(z)$ and $m_+(z)$. We limit our analysis to the Stieltjes transform m, as we note the identity

$$(8.15) \qquad m_{\infty,+}(z) \; = \; \frac{\gamma - 1}{z} + \gamma m(z),$$

which follows by definition of the augmented measure ϱ_+. The first result we want to prove in studying the Stieltjes transform is the following $O(1)$ estimate.

Lemma 8.3. *Uniformly over $z \in \mathbb{C}_+$, we have*

$$(8.16) \qquad m(z) \; = \; O(1).$$

We proceed by considering the two regimes $\gamma = 1$ and $\gamma < 1$, which we refer to as the *square regime* and *rectangular regime*, respectively.

Square Regime. If $\gamma = 1$, we rewrite the density $\varrho(E)$ as

$$(8.17) \qquad \varrho_{\gamma=1}(E) \; = \; \frac{\sqrt{4 - E^2}}{2\pi}\mathbf{1}_{E\in[-2,2]},$$

which is the well-studied semicircle density, whose Stieltjes transform is given by

$$(8.18) \qquad m(z) \; = \; \frac{-z + \sqrt{z^2 - 4}}{2}.$$

For a reference on the semicircle law and its Stieltjes transform, we cite [2], [3], [4], [11], [13], and [17]. We note the branch of the square root is taken so that $\sqrt{z^2 - 4} \sim z$ for large z, in which case the bound (8.16) follows immediately.

Rectangular Regime. Fix constants $\Lambda > 0$ and $\varepsilon > 0$ to be determined. Suppose $|E| \in [\varepsilon, \Lambda]$. By the representation $m(z) = zm_\infty(z^2)$ and the explicit formula for $m_\infty(z)$ as given in (8.2), the bound (8.16) follows immediately in this energy regime, where the implied constant in (8.16) may be taken independent of η.

Suppose now that $|E| > \Lambda$. Again by the representation $m(z) = zm_\infty(z^2)$, we have

$$(8.19) \qquad |m(z)| = O\left(-z^2 + i\sqrt{(\lambda_+ - z^2)(z^2 - \lambda_-)}\right)$$

$$(8.20) \qquad = O\left(-z^2 + \sqrt{(z^2 - \lambda_+)(z^2 - \lambda_-)}\right)$$

$$(8.21) \qquad = O(1),$$

since the square root is, again, chosen so that $\sqrt{z^4 + O(z^2)} \sim z^2$ for large z.

Lastly, suppose $|E| < \varepsilon$. By definition of $m(z)$ as the Stieltjes transform of ϱ, we obtain

$$(8.22) \qquad |m(z)| = \left| \int_{E^2 \in [\lambda_\pm]} \frac{\gamma}{(1+\gamma)\pi|E|(E-\varepsilon)} \sqrt{(\lambda_+ - E^2)(E^2 - \lambda_-)} \, dE \right| = O\left(\frac{1}{\sqrt{\lambda_- - \varepsilon}}\right),$$

where the implied constant depends only on fixed data $\gamma, \lambda_\pm$. Choosing $\varepsilon = \sqrt{\lambda_-}/100 > 0$, we obtain the desired bound. We note this choice of ε is positive if and only if $\alpha > 1$. This completes the proof of Lemma 8.3. $\qquad \square$

We are now in a position to make our final reduction of the proof of the local laws in Theorem 4.1. This reduction exploits the first reduction in Lemma 7.5, allowing us to focus on the covariance matrices X_* and $X_{*,+}$. The reduction will depend on the following result, for which we need to define the following spectral domain.

$$(8.23) \qquad \mathbf{D}_{N,\delta,\xi} = \{z = E + i\eta : |E| \leqslant N^\delta, \, \xi^2/N \leqslant \eta \leqslant N\}.$$

Proposition 8.4. *Suppose $\xi, \zeta > 0$ and $D \gg \xi^2$. If, for a fixed $z \in \mathbf{D}_{N,\delta,\xi}$, we have*

$$(8.24) \qquad \mathbb{P}\left(\Gamma_\star^*(z) = O(1)\right) \geqslant 1 - e^{-\zeta},$$

then, with probability at least $1 - e^{-(\xi \log \xi) \wedge \zeta + O(\log N)}$, we have

$$(8.25) \qquad \max_i |[G_\star(z)]_{ii} - m(z)| = O(F_z(\xi\Phi(z))), \quad \max_{i \neq j} |[G_\star(z)]_{ij}| = O\left(\frac{\xi\Phi(z^2)}{z}\right).$$

Here, $\star$ can take the values $\star = *$ and $\star = *, +$.

Compared to the local laws for X_* and $X_{*,+}$ in Theorem 4.1, the statement of Proposition 8.4 requires only a pointwise estimate instead of a uniform estimate. To go from a pointwise estimate to a uniform estimate, we appeal to an adaptation of a standard argument in random matrix theory using the Lipschitz continuity of the Green's function entries. Moreover, since the estimate holds with high probability at each point, we will need to bound from below the probability of the intersection of these events – by our notion of high probability, this intersection of high probability events also occurs with high probability. For a reference on this method, we appeal to [4]. Lastly, the statement of Proposition 8.4 requires an a priori estimate on the Green's function entries.

We now use Proposition 8.4 to derive the final reductions of the proof of Theorem 4.1. The first reduction is to use Proposition 8.4 to derive the local law for the Green's functions G_* and $G_{*,+}$, i.e. the estimates on the covariance matrices in Theorem 4.1. The last reductions are removing the repulsion assumption $|E| > \varepsilon$ on the energy whenever suitable. These two procedures will follow the same argument, essentially, which is a method of scaling using Lemma 6.6. We begin the with the first reduction.

Lemma 8.5. *Suppose Proposition 8.4 holds. Then the estimates (4.7) and (4.8) in Theorem 4.1 hold.*

Proof. We first note that, by the same argument as given in Lemma 8.3, uniformly on the set U_ε we have the estimates

$$(8.26) \qquad m_\infty(z), m_{\infty,+}(z) \leqslant C,$$

for some constant $C = O(1)$. Moreover, we note that it suffices to restrict from U_ε to the domain $\mathbf{D}_{N,\delta,\xi}$, as by Proposition 2.4, the spectrum of X_* and $X_{*,+}$ is bounded by $O(1)$.

Fix an energy E_0 in $\mathbf{D}_{N,\delta,\xi} \cap U_\varepsilon$, i.e. $z_0 := E_0 + iN \in \mathbf{D}_{N,\delta,\xi} \cap U_\varepsilon$. For any $\eta \in [\xi^2/N, N]$, fix $z = E_0 + i\eta$ and define the integer K by

$$(8.27) \qquad\qquad K = \min\{k : N/2^k \leqslant \eta\}.$$

First, we note $K = O(\log N)$ by the constraints on η given by $\mathbf{D}_{N,\delta,\xi}$. We now define $z_k = E_0 + iN/2^k$ for $k = 0, \dots, K$.

By the Ward identity, we have the following deterministic bound:

$$(8.28) \qquad\qquad \Gamma_\star^*(z_1) \leqslant \frac{1}{N}$$

Thus, by Proposition 8.4, we deduce the local laws at $z = E_0 + iN$. We now proceed inductively; suppose the local law holds at some $k \in [[0, N]]$ with probability at least $1 - e^{(\xi \log \xi) \wedge \zeta + O(\log N)}$. Then, on this event, we know

$$(8.29) \qquad\qquad \Gamma_\star^*(z_k) = O(1),$$

where the implied constant is independent of k by the comparison given in the local law, and the uniform $O(1)$ estimate on $m_\infty, m_{\infty,+}$ on U_ε. By Lemma 6.6, on this event we obtain the following with probability at least $1 - e^{(\xi \log \xi) \wedge \zeta + O(\log N)}$:

$$(8.30) \qquad\qquad \Gamma_\star^*(z_{k+1}) \leqslant 2\Gamma_\star^*(z_k) = O(1).$$

By Proposition 8.4, we obtain the local laws at z_{k+1} with probability at least $1 - e^{\xi \log \xi}$. Continuing inductively, we deduce the local law at the parameter z.

To obtain a local law uniformly over U_ε, we now discretize $U_\varepsilon \cap \mathbf{D}_{N,\delta,\xi}$ into a lattice Λ of mesh bounded by N^{-4}. If we define $A(z)$ to be the event that the local laws hold at the point z, then we have by a standard probability bound

$$(8.31) \qquad\qquad \mathbb{P}\left(\bigcap_{z \in \Lambda} A(z)\right) \geqslant 1 - |\Lambda| e^{-\xi \log \xi} \geqslant 1 - e^{-\xi \log \xi},$$

where in the last lower bound we abuse notation and adapt the parameter ξ such that the growth bound $\xi \log \xi \gg \log N$ still holds. In the last lower bound, we also use the simple bound $|\Lambda| = N^{O(1)}$. This completes the proof of Lemma 8.5. $\qquad\square$

We now proceed to the final reduction of the proof of Theorem 4.1. This reduction is concerned with extending the local law for the Green's function of the linearization X beyond the domain U_ε and to energies $|E| < \varepsilon$. The strategy of this extension mimics the proof of Lemma 8.5, instead now using the regularity of the Stieltjes transform m near the origin established in Lemma 8.3. Moreover, by Lemma 8.1, it suffices to extend the estimate beyond U_ε for the partial Stieltjes transform s_w. Lastly, this extension may be performed for the Green's function G_* of the smaller covariance matrix X_* if $\alpha > 1$. This is summarized in the following lemma.

Lemma 8.6. *Suppose Proposition 8.4 holds. Then the estimates* (4.10) *and* (4.11) *hold without the assumption* $|E| > \varepsilon$. *Consequently, the estimates* (4.14) *and* (4.15) *hold without the assumption* $|E| > \varepsilon$. *Moreover, if* $\alpha > 1$, *then the estimate* (4.7) *holds without the assumption* $|E| > \varepsilon$.

Proof. We first assume $\alpha \geqslant 1$, i.e. the general regime. The first statement follows via the same inductive argument in Lemma 8.5 once we have the regularity of the Stieltjes transform uniformly over the upper-half plane given in Lemma 8.3. In particular, we may begin the iteration scheme over dyadic scales over energies $|E_0| < \varepsilon$ as well.

For the rectangular regime $\alpha > 1$, we again use the same inductive argument in Lemma 8.5 upon noting $m_\infty = O(1)$ on $z \in \mathbb{C}_+$. This follows by a similar argument given in the proof of Lemma 8.3 concerning the Stieltjes transform m of the linearization.

Lastly, we deduce the estimate (4.15) holds without the energy assumption by averaging over indices $k \in [[M+1, M+N]]$. We deduce the estimate (4.14) without the energy assumption by Lemma 8.1. $\qquad\square$

8.1. **Recap of Reductions.** We now summarize the reductions achieved in this discussion as follows.

- We begin by focusing on the domain U_ε, i.e. some repulsion estimate on the energy E. We first derive the equivalence of entry-wise local laws for the Green's functions of the covariance matrices G_* and $G_{*,+}$ and the Green's function of the linearization G. This is the content of Lemma 7.5. This allows us to focus solely on the covariance matrices whenever working in the domain U_ε.
- As a brief aside, we prove a strong equivalence for the local laws concerning the Stieltjes transforms s_* and $s_{*,+}$ and an equivalence for the local laws concerning the partial Stieltjes transforms s_b and s_w. The strength in this equivalence is removing the repulsion assumption in U_ε. This is the content of Lemma 8.1, and it will be important in extending the local laws outside U_ε whenever appropriate.
- We now assume an a priori estimate, both on the parameter z in constructing the domain $\mathbf{D}_{N,\delta,\xi}$, and on the Green's function entries. From this, we derive the local laws for G_* and $G_{*,+}$. By the first point, this gives the entry-wise local laws for the Green's function G. This is the content of Lemma 8.5.
- We now extend the estimate for the lower-right entries of the Green's function G beyond the domain U_ε, i.e. removing the repulsion assumption on the energy. This is the content of Lemma 8.6. This requires an a priori estimate on the Stieltjes transform m uniform over the upper-half plane. This point is taken care of in Lemma 8.3. By the same method, we also remove the repulsion assumption on the energy for the Green's function G_* in the regime $\alpha > 1$.

We reemphasize that the underlying mechanism for establishing these reductions is given by Proposition 7.2, which relates all Green's functions and Stieltjes transforms by Lemma 7.3 and Corollary 7.4. Extending the estimates past U_ε requires additional regularity of the limiting Stieltjes transform.

To recap, to prove Theorem 4.1, it now suffices to prove Proposition 8.4. This will be the focus for the remainder of this chapter.

III. Switchings on Graphs

We now focus on constructing a method of resampling graphs in Ω. We will construct the resamplings using tools in graph theory known as *switchings*. The resampling method via switchings is discussed in detail for the ensemble of d-regular graphs in [3] and [4]. Our methods will mimic the methods used in those papers, adjusted appropriately for biregular graphs.

9. Switchings on Biregular Graphs

We begin by introducing notation necessary to define switchings on biregular graphs. Switchings will be local operations on the biregular graphs, so we will establish notation for vertices and edges containing said vertices as follows.

Notation 9.1. *A generic vertex in V_b (resp. V_w) will be denoted by v_b (resp. v_w).*

For a fixed graph $E \in \Omega$, we will denote the edges in E containing v_b by $\{e_{b,\mu}\}_{\mu=1}^{d_b}$. Moreover, for a fixed edge $e_{b,\mu}$ containing v_b, we will denote the neighboring vertex by $v_{b,\mu}$, so that $e_{b,\mu} = v_b v_{b,\mu}$.

Similarly, the edges in E containing v_w will be denoted by $\{e_{w,\nu}\}_{\nu=1}^{d_w}$. For a fixed edge $e_{w,\nu}$ containing v_w, we will denote the neighboring vertex by $v_{w,\nu}$.

For a fixed vertex $v_b \in V_b$, we establish the notation for the set of edges not containing v_b:

$$(9.1) \qquad U_{v_b} := \{\text{edges } e \in E : v_b \notin e\}.$$

Similarly for a fixed vertex $v_w \in V_w$, we define the following set of edges not containing v_w:

$$(9.2) \qquad U_{v_w} := \{\text{edges } e \in E : v_w \notin e\}.$$
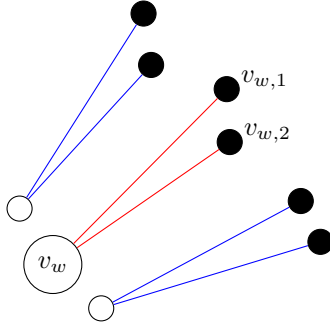
FIGURE 4. We illustrate the established notation for local data in this graph. Here having fixed a white vertex v_w, the red edges denote the defined edges $e_{w,1}$ and $e_{w,2}$, whose black vertices are defined as $v_{w,1}$ and $v_{w,2}$ respectively. The blue edges denote the set U_{v_w}.

Remark 9.2. Before we proceed, we note that the notation established for vertices and the corresponding edges and neighbors suppresses the vertex v_b from the explicit notation. We do this without the risk of confusion among edges and neighbors of different vertices of the same color (black or white).

We may now begin to define a switching on a generic graph $E \in \Omega$. To this end, we fix a black vertex $v_b \in V_b$ and an edge $e_{v,\mu}$ for some $\mu \in [[1, d_b]]$. We define the following space of subgraphs of E:

$$(9.3) \qquad \mathbf{S}_{v_b,\mu,E} := \{S \subset E : S = \{e_{b,\mu}, p_{b,\mu}, q_{b,\mu}\}, \ p_{b,\mu} \neq q_{b,\mu} \in U_{v_b}\}.$$
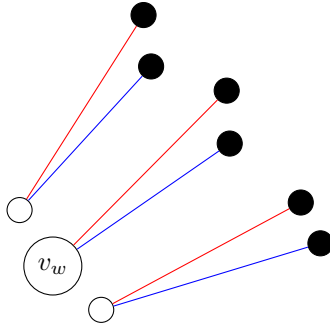
FIGURE 5. We illustrate the indicator functions defined in (9.6) through (9.8) with the fixed *white* vertex v_w. Here, $S_{w,1}$ denotes the subgraph defined by the red edges and $S_{w,2}$ denotes the subgraph defined by the blue edges. For both graphs, $I(S_{w,1}) = I(S_{w,2}) = 1$. However, $J(\mathbf{S}_{v_w}, \nu) = 0$ for both edges $\nu = 1$ and $\nu = 2$, where the product in (9.7) is defined over the two graphs $S_{w,1}$ and $S_{w,2}$. Thus, the set W is empty.
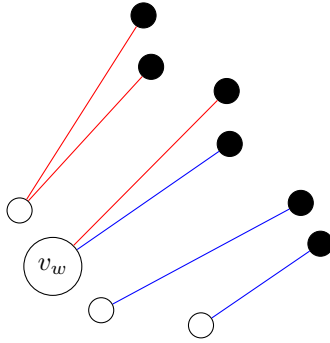
FIGURE 6. Again we fix a *white* vertex v_w. We let $S_{w,1}$ denote the subgraph defined by the red edges and we let $S_{w,2}$ denote the subgraph defined by the blue edges. We note $I(S_{w,1}) = 0$ while $I(S_{w,2}) = 1$. Moreover, we know $J(\mathbf{S}_{v_w}, \nu) = 1$ for $\nu = 1$ and $\nu = 2$, as the two subgraphs share only the vertex v_w. Thus, the set W contains the edge label $\nu = 2$.

In words, the set $\mathbf{S}_{v_b,\mu,E}$ is the set of graphs consisting of the edges $e_{b,\mu}$ and any two distinct edges $p_{b,\mu}$ and $q_{b,\mu}$, neither of which contains the vertex v_b. Similarly, we may define for a fixed white vertex $v_w \in V_w$ and edge $e_{w,\nu}$, for some $\nu \in [[1, d_w]]$, the same set of graphs:

$$(9.4) \qquad \mathbf{S}_{v_w,\mu,E} := \{S \subset E : S = \{e_{w,\nu}, p_{w,\nu}, q_{w,\nu}\} : p_{w,\nu} \neq q_{w,\nu} \in U_{v_w}\}.$$

Notation 9.3. *A generic graph in $\mathbf{S}_{v_b,\mu,E}$ will be denoted by $S_{b,\mu}$. A generic graph in $\mathbf{S}_{v_w,\nu,E}$ will be denoted by $S_{w,\nu}$.*

The set $\mathbf{S}_{v_b,\mu,E}$ contains the edge-local data along which switchings on graphs will be defined. To make this precise, we need to introduce the following indicator functions. First, we define the following configuration vectors for fixed vertices $v_b \in V_b$ and $v_w \in V_w$:

$$(9.5) \qquad \mathbf{S}_{v_b} := (S_{b,\mu})_{\mu=1}^{d_b}, \quad \mathbf{S}_{v_w} := (S_{w,\nu})_{\nu=1}^{d_w}.$$

With this notation, we define the following indicator functions that detect graph properties in $S_{b,\mu}$ and $S_{w,\nu}$.

$$(9.6) \qquad I(S_{b,\mu}) = \mathbf{1}\left(|[S_{b,\mu}]| = 6\right),$$

$$(9.7) \qquad J(\mathbf{S}_{v_b}, \mu) = \prod_{\mu' \neq \mu} \mathbf{1}\left([S_{b,\mu}] \cap [S_{b,\mu'}] = \{v_b\}\right),$$

$$(9.8) \qquad W(\mathbf{S}_{v_b}) = \{\mu : I(S_{b,\mu})J(\mathbf{S}_{v_b}, \mu) = 1\}.$$

For white vertices $v_w \in V_w$, the functions I, J and W retain the same definition upon replacing b with w and μ with ν.

We now define the augmented probability spaces $\widetilde{\Omega}$ which will make the switchings systematic from the perspective of Markovian dynamics. For a fixed black vertex $v_b \in V_b$ and a fixed white vertex $v_w \in V_w$, we define the following augmented space:

$$(9.9) \qquad \widetilde{\Omega} = \left\{ (E, \mathbf{S}_{v_b}, \mathbf{S}_{v_w}), \quad E \in \Omega, \ \mathbf{S}_{v_b} \in \prod_{\mu=1}^{d_b} \mathbf{S}_{v_b, \mu, E}, \ \mathbf{S}_{v_w} \in \prod_{\nu=1}^{d_w} \mathbf{S}_{v_w, \nu, E} \right\}$$

$$(9.10) \qquad \approx \Omega \times \prod_{\mu=1}^{d_b} \mathbf{S}_{v_b, \mu, E} \times \prod_{\nu=1}^{d_w} \mathbf{S}_{v_w, \nu, E}.$$

We now clarify the definition of $\widetilde{\Omega}$, in particular the definition in terms of a product space is an abuse of notation. Precisely, $\widetilde{\Omega}$ is the space of triples $(E, \mathbf{S}_{v_b}, \mathbf{S}_{v_w})$ such that E is a biregular graph in Ω, and $\mathbf{S}_{v_b}$ and $\mathbf{S}_{v_w}$ are configuration vectors. In particular, the product space representation for $\widetilde{\Omega}$ given above is ill-defined as the second and third factors in the product depend on the first factor Ω, hence the approximation. However, the precise set-theoretic definition given is well-defined.

9.1. **Switchings of Biregular Graphs.** We now precisely define switchings by defining dynamics on $\widetilde{\Omega}$. To this end we define switchings on configuration vectors $\mathbf{S}_{v_b}$ and $\mathbf{S}_{v_w}$; we first focus on the configuration vectors for black vertices.

Fix a label μ and consider a component $S_{b,\mu}$ of a uniformly sampled configuration vector $\mathbf{S}_{v_b}$. Precisely, the components of $\mathbf{S}_{v_b}$ are sampled jointly uniformly and independently from $\mathbf{S}_{v_b, \mu, E}$, where $E \in \Omega$ is uniform over all μ and sampled uniformly. We now define the following map:

$$(9.11) \qquad T_b : \prod_{\mu=1}^{d_b} \mathbf{S}_{v_b, \mu, E} \longrightarrow \prod_{\mu=1}^{d_b} \mathbf{S}_{v_b, \mu, E'}$$

where $E' \in \Omega$ is possibly different from E. The map is given as follows: for any μ, we define the map $T_{b,\mu}$

$$(9.12) \qquad T_{b,\mu}(S_{b,\mu}) = \begin{cases} S_{b,\mu} & \mu \notin W(\mathbf{S}_{v_b}) \\ (S_{b,\mu}, s_{b,\mu}) & \mu \in W(\mathbf{S}_{v_b}) \end{cases}.$$

We define the graph $(S_{b,\mu}, s_{b,\mu})$ as follows; this is where we now introduce randomness into the dynamics T_b. Suppose $\mu \in W(\mathbf{S}_{v_b})$, in which case $S_{b,\mu}$ is 1-regular and bipartite with respect to the vertex sets $([S_{v_b}], V_1, V_2)$. Consider the set of 1-regular bipartite graphs with respect to the vertex set $([S_{b,\mu}], V_1, V_2)$. In words, this is the set of 1-regular graphs on $[S_{b,\mu}]$ such that, upon replacing $S_{b,\mu}$ with any such graph, the global graph E remains biregular. We now define $(S_{b,\mu}, s_{b,\mu})$ to be drawn from this set uniformly at random conditioning on the event $(S_{b,\mu}, s_{b,\mu}) \neq S_{b,\mu}$. Lastly, we define the following global dynamics:

$$(9.13) \qquad T_b = \prod_{\mu=1}^{d_b} T_{b,\mu},$$

where the product is taken as composition. We note this product is independent of the order of composition; this is a consequence of the definition of the functions I, J and W. For white vertices $v_w \in V_w$, we define the map T_w by replacing all black indices b and white indices w.

Remark 9.4. We briefly remark here that $s_{b,\mu}$ may be thought of as a fair Bernoulli random variable as there exist only two possible graphs on $[S_{b,\mu}]$ distinct from $S_{b,\mu}$ that preserve the bipartite structure of the global graph E. This perspective was motivated by [4] in the discussion of d-regular graphs, but is unimportant to the discussion at hand.

We note that the maps T_b and T_w define maps on $\widetilde{\Omega}$, because we are allowed to change the underlying graph E when varying over the space $\widetilde{\Omega}$; this is the utility of the almost-product representation of $\widetilde{\Omega}$. This allows us to finally define switchings of a biregular graph.

Definition 9.5. For a fixed black vertex $v_b \in V_b$ and a fixed label $\mu \in [[1, d_b]]$, the *local switching* at v_b along μ is the map $T_{b,\mu}$. The *global switching* is the map T_b.
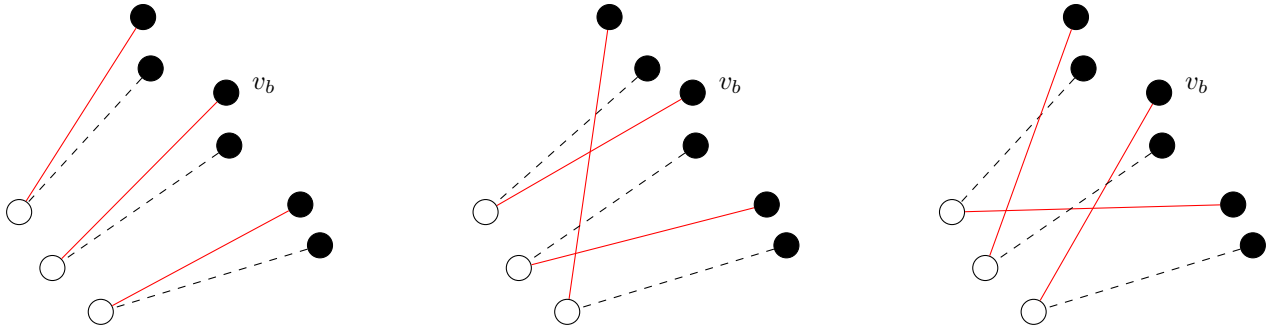
FIGURE 7. We illustrate a local switching at v_b along the red edge. Here, $S_{b,\mu}$ denotes the subgraph defined by the red edges on the far LHS graph. We note the red edge is contained in the set W, so we may switch to either the middle graph or the RHS graph.

Similarly, for a fixed white vertex $v_w \in V_w$ and a fixed label $\nu \in [[1, d_w]]$, the *local switching* at v_w along ν is the map $T_{w,\nu}$. The *global switching* at v_w is the map T_w.

Remark 9.6. We note our construction, technically, implies the mappings $T_{b,\mu}$ and $T_{w,\nu}$ are random mappings on the augmented space $\widetilde{\Omega}$. Via this construction, we obtain a probability measure $\widetilde{\Omega}$ induced by the uniform measure and a uniform sampling of switchings. To obtain an honest mapping on the original space Ω, we may instead construct deterministic mappings by averaging over the random switchings. For precise details, we cite [4].

To even be able to exploit switchings in our analysis of biregular graphs and their Green's functions, we need the following lemma which establishes the switchings do not change the underlying probability measures on $\widetilde{\Omega}$ and Ω. To make sense of this, we impose the uniform measure on $\widetilde{\Omega}$. To show this coincides with our uniform model on Ω amounts to showing the uniform probability measure on Ω is the pushforward of the uniform probability measure on $\widetilde{\Omega}$ under the projection $\widetilde{\Omega} \to \Omega$, which follows from a direct inspection. Thus, we must show the uniform probability measure on $\widetilde{\Omega}$ is preserved by switchings. This is the content of the following lemma.

Lemma 9.7. *The uniform probability measure on $\widetilde{\Omega}$ is invariant under the dynamics T_b and T_w. In particular, for any function $f : \widetilde{\Omega} \to \mathbb{R}$, we have*

(9.14)
$$\mathbb{E}_{\mathrm{Unif}} f(T_w(E)) \;=\; \mathbb{E}_{\mathrm{Unif}} f(E) \;=\; \mathbb{E}_{\mathrm{Unif}} f(T_b(E)).$$

Proof. It suffices to note that T_b and T_w define bijections on $\widetilde{\Omega}$, which is a finite set. $\qquad\square$

Remark 9.8. We briefly remark that Lemma 9.7 may be upgraded to a stronger result; in particular, the dynamics T_b and T_w are reversible with respect to the uniform probability measure on $\widetilde{\Omega}$. The reversibility amounts to deriving the involution property of T_w and T_b. This, however, will not be important in our analysis of Green's functions, so we omit this discussion. For details, however, we cite Lemma 6.1 in [4].

10. Switchings on Adjacency Matrices

We now redirect our focus to the construction of switchings on biregular graphs, now from the perspective of linear perturbations on a matrix space rather than combinatorial operations on graphs. The upshot to the matrix perspective is a method of computing stability of the Green's functions after switchings. To this end, suppose $E \in \Omega$ is a biregular graph with adjacency matrix A. We will fix the following notation.

Notation 10.1. *For an edge $e = ij$ on the vertex set $\mathscr{V}$, we let Δ_{ij} denote the adjacency matrix of the graph on $\mathscr{V}$ consisting only of the edge e. In particular, Δ_{ij} is the matrix whose entries are given by*

(10.1)
$$(\Delta_{ij})_{k\ell} \;=\; \delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk}.$$

23

In the context of switchings on biregular graphs, the matrices Δ_{ij} are perturbations of adjacency matrices. This is made precise in the following definition.

Definition 10.2. Fix a black vertex $v_b \in V_b$ and a label $\mu \in [[1, d_b]]$. A *local* switching of A, denoted $T_{b,\mu}$, at v_b along μ is given by the following formula:

$$(10.2) \qquad T_{b,\mu}(A) \;=\; A - S_{b,\mu} + (S_{b,\mu}, s_{b,\mu}),$$

where $S_{b,\mu}$ is a component of a random, uniformly sampled configuration vector $\mathbf{S}_{v_b}$. A *global* switching of A, denoted T_b, is the composition of the random mappings $T_{b,\mu}$.

Similarly, we may define local switchings and global switchings of adjacency matrices for white vertices by replacing the black subscript b with the white subscript w, and replacing the label μ with ν.

Clearly, a local or global switching of an adjacency matrix is the adjacency matrix corresponding to a local or global switching of the underlying graph. To realize the matrices Δ_{ij} as perturbations, we will rewrite the formula defining $T_{b,\mu}$ as follows. As usual, we carry out the discussion for black vertices $v_b \in V_b$, though the details for white vertices v_b follow analogously.

First, we recall the following notation for a component $S_{b,\mu} \in \mathbf{S}_{v_b,\mu,E}$ of a configuration vector $\mathbf{S}_{v_b}$:

$$(10.3) \qquad S_{b,\mu} \;:=\; \{e_{b,\mu}, p_{b,\mu}, q_{b,\mu}\},$$

subject to the constraint that $S_{b,\mu}$ contains three distinct edges.

Notation 10.3. *We will denote the vertices of $p_{b,\mu}$ by $a_{b,p,\mu} \in V_b$ and $a_{w,p,\mu} \in V_w$. Similarly, we will denote the vertices of $q_{b,\mu}$ by $a_{b,q,\mu} \in V_b$ and $a_{w,q,\mu} \in V_w$.*

With this notation, we may rewrite the random mapping $T_{b,\mu}$ as follows:

$$(10.4) \qquad T_{b,\mu}(A) \;=\; A - \left(\Delta_{v_b,v_{b,\mu}} + \Delta_{a_{b,p,\mu}a_{w,p,\mu}} + \Delta_{a_{b,q,\mu}a_{w,q,\mu}}\right) + \left(\Delta_{v_b,x} + \Delta_{a_{b,p,\mu},y} + \Delta_{a_{b,q,\mu},z}\right),$$

where we recall $e_{b,\mu} = v_b v_{b,\mu}$. Here, the variables x, y, z are three distinct vertices sampled from the set of white vertices $\{v_{b,\mu}, a_{w,p,\mu}, a_{w,q,\mu}\}$ conditioning on the following constraint on ordered triples:

$$(10.5) \qquad (x, y, z) \;\neq\; (v_{b,\mu}, a_{w,p,\mu}, a_{w,q,\mu}).$$

The representation (10.4) of the random mapping $T_{b,\mu}$ gives an interpretation of switchings analytically as perturbations on a linear space. This representation will be important in studying the Green's functions of adjacency matrices upon switchings.

To exploit the perturbation representation (10.4), however, we must first understand the randomness of the vertices of the randomly sampled components $S_{b,\mu}$. This will be the remaining focus of our discussion on constructing switchings on biregular graphs.

11. Probability Estimates on Vertices

We begin by defining the following probabilistic notion which will be helpful in recording estimates on the distribution of vertices of randomly sampled components $S_{b,\mu}$.

Definition 11.1. Suppose S is a finite set and X is an S-valued random variable. We say (the distribution of) X is *approximately uniform* if the following bound on total variation holds:

$$(11.1) \qquad \sum_{s \in S} \left| \mathbb{P}\left(X = s\right) - \frac{1}{|S|} \right| \;\leqslant\; O\left(\frac{1}{\sqrt{d_w D}}\right).$$

We now introduce the following σ-algebras on $\widetilde{\Omega}$. These σ-algebras will allow us to focus on edge-local features of graphs $E \in \Omega$ upon conditioning on the global data E.

Definition 11.2. For a fixed label $\mu \in [[0, d_b]]$, we define the following σ-algebras:

$$\text{(11.2)} \qquad \mathscr{F}_\mu := \sigma\left(E, (S_{b,1}, s_{b,1}), \ldots, (S_{b,\mu}, s_{b,\mu})\right),$$

$$\text{(11.3)} \qquad \mathscr{G}_\mu := \sigma\left(E\left(S_{b,\mu'}, s_{b,\mu'}\right)_{\mu' \neq \mu}\right).$$

We similarly define the σ-algebras $\mathscr{F}_\nu$ and $\mathscr{G}_\nu$ for $\nu \in [[1, d_w]]$ for white vertices.

We briefly clarify the meaning of the σ-algebras $\mathscr{F}_\mu$ and $\mathscr{G}_\mu$, in particular remarking the notation $(S_{b,\mu'}, s_{b,\mu'})$ denotes conditioning on the switching data at v_b along μ', and similarly for $(S_{w,\nu'}, s_{w,\nu'})$.

The last piece of probabilistic data we introduce is the following notation, which will allow us to compare i.i.d. switchings on biregular graphs.

Notation 11.3. *Suppose X is a random variable on the graph data $E, \{(S_{b,\mu}, s_{b,\mu})\}_\mu, \{(S_{w,\nu}, s_{w,\nu})\}_\nu$. Then $\widetilde{X}$ denotes a random variable on the variables $\widetilde{E}, \{(\widetilde{S}_{b,\mu}, \widetilde{s}_{b,\mu})_\mu\}, \{(\widetilde{S}_{w,\nu}, \widetilde{s}_{w,\nu})\}$, where the tildes on the graph data denote i.i.d. resamplings.*

Notation 11.4. *For notational simplicity, by p_μ, q_μ or p_ν, q_ν, we will refer to either $p_{b,\mu}, q_{b,\mu}$ or $p_{w,\nu}, q_{w,\nu}$, respectively, whenever the discussion applies to both situations.*

Having established this probabilistic data, we now focus on obtaining an estimate on the distribution of the pair of edges (p_μ, q_μ), and similarly for (p_ν, q_ν). As with all results concerning switchings from here on, details of proofs resemble those of Section 6 in [4], so we omit details whenever redundant.

Lemma 11.5. *Conditioned on $\mathscr{G}_\mu$, the pair (p_μ, q_μ) is approximately uniform, i.e., for any bounded symmetric function F, we have*

$$\text{(11.4)} \qquad \mathbb{E}_{\mathscr{G}_\mu} F(p_\mu, q_\mu) = \frac{1}{(Nd_w)^2} \sum_{p,q \in E} F(p, q) + O\left(\frac{1}{N}\|F\|_\infty\right).$$

Similarly, for any bounded function F, we have

$$\text{(11.5)} \qquad \mathbb{E}_{\mathscr{G}_\mu, q_\mu} F(p_\mu) = \frac{1}{Nd_w} \sum_{p \in E} F(p) + O\left(\frac{1}{N}\|F\|_\infty\right).$$

Proof. Assume we resample about $v_b \in V_b$; the case for $v_w \in V_w$ follows analogously. By definition, we have

$$\text{(11.6)} \qquad \mathbb{E}_{\mathscr{G}_\mu} F(p_\mu, q_\mu) = \frac{1}{(Nd_w - d_w)(Nd_w - d_w - 1)} \sum_{p \in E_{v_b}} F(p),$$

where E_{v_b} is the set of edges in E that are not incident to v_b. Then, (11.4) follows from the following estimate

$$\text{(11.7)} \qquad \frac{1}{(Nd_w - d_w)(Nd_w - d_w - 1)} = \frac{1}{(Nd_w)^2} + O\left(\frac{1}{N^3 d_w^2}\right)$$

as well as the estimate $|E_{v_b}| \leqslant (Nd_w)^2$, and lastly the estimate $|E_{v_b}^C| \leqslant Nd_w$. This last upper bound follows combinatorially; for details, see the proof of Lemma 6.2 in [4]. The estimate (11.5) follows from a similar argument. □

Because an edge is uniquely determined by its vertices in the graph, we automatically deduce from Lemma 11.5 the following approximately uniform estimate for resampled vertices as well.

Corollary 11.6. *Conditioned on $\mathscr{G}_\mu$, the pair $(p_\mu(b), q_\mu(b))$ (resp. $(p_\mu(w), q_\mu(w))$) is approximately uniform.*
Similarly, conditioned on $\mathscr{G}_\mu$ and $q_\mu(b)$ (resp. $q_\mu(w)$), the random variable $p_\mu(b)$ (resp. $p_\mu(w)$) is approximately uniform.

Proof. This follows immediately upon applying Lemma 11.5 to the function $F(p_\mu, q_\mu) = f(p_\mu(b), q_\mu(b))$. □

To fully exploit the resampling dynamics, we need a lower bound on the probability that a local switching $S_{b,\mu}, s_{b,\mu}$ around a vertex $v_b \in V_b$ does not leave the graph fixed. In particular, we need an estimate for the probability of the event $\mu \in W(\mathbf{S}_{v_b})$ where here μ is fixed and the set W is viewed as random. As discussed in [4], to provide an estimate, the naive approach to estimating this probability conditioning on $\mathscr{G}_\mu$ fails in an exceptional set. Precisely, suppose the μ-th neighbor $v_{b,\mu}$ of v_b

lives in $S_{b,\mu'}$ for some $\mu' \neq \mu$. In this case, almost surely, we have $[S_{b,\mu}] \cap [S_{b,\mu'}]$ is nontrivial. It turns out this is the only obstruction, so we aim to show that $v_{b,\mu} \in [S_{b,\mu'}]$ occurs with low probability for any $\mu' \neq \mu$.

Formally, we define the following indicator random variable which detects this exceptional set:

$$h(\mathbf{S}_{v_b}, \mu) \; = \; \prod_{\mu' \neq \mu} \mathbf{1}\left(v_{b,\mu} \in S_{b,\mu'}\right). \tag{11.8}$$

Thus, the estimates we need are given in the following result.

Lemma 11.7. *For any neighbor index μ, we have*

$$\mathbb{P}_{\mathscr{G}_\mu}\left[I(S_{b,\mu})J(\mathbf{S}_{v_b}, \mu) \; = \; h(\mathbf{S}_{v_b}, \mu)\right] \; \geqslant \; 1 - O\left(\frac{d_b}{N}\right). \tag{11.9}$$

Moreover, we have

$$\mathbb{P}_{\mathscr{F}_0}\left[h(\mathbf{S}_{v_b}, \mu) = 1\right] \; \geqslant \; 1 - O\left(\frac{d_b}{N}\right). \tag{11.10}$$

Proof. We first note that (11.9) follows immediately conditioning on $h = 0$. In particular, the first lower bound (11.9) follows from a combinatorial analysis of the underlying graph using the following union bound:

$$\mathbb{P}_{\mathscr{G}_\mu, h=1}\left[I(S_{b,\mu})J(\mathbf{S}_{v_b}, \mu) = 0\right] \; \leqslant \; \mathbb{P}_{\mathscr{G}_\mu}\left[I(S_{b,\mu}) = 0\right]$$
$$+ \; \mathbb{P}_{\mathscr{G}_\mu, h=1}\left[J(\mathbf{S}_{v_b}, \mu) = 0\right]. \tag{11.11}$$

Similarly, (11.10) follows from the union bound

$$\mathbb{P}_{\mathscr{F}_0}\left[h(\mathbf{S}_{v_b}, \mu) = 0\right] \; \leqslant \; \sum_{\mu' \neq \mu} \mathbb{P}_{\mathscr{F}_0}\left[v_{b,\mu} \in [S_{\mu'}]\right]. \tag{11.12}$$

For details, we refer back to [4]. $\qquad\square$

We conclude this section with an estimate that compares independent resamplings. Recall that $\widetilde{W}, W$ are i.i.d. copies of the random variable $W(\mathbf{S}_{v_b})$. The following result bounds the fluctuation in $W(\mathbf{S}_{v_b})$ from independent resamplings.

Lemma 11.8. *Almost surely, we know*

$$\#\left(W \Delta \widetilde{W}\right) \; = \; O(1), \tag{11.13}$$

where the implied constant is independent of N. Moreover, we also have

$$\mathbb{P}_{\mathscr{G}_\mu}\left[W \Delta \widetilde{W} \neq \emptyset\right] \; \leqslant \; O\left(\frac{d_b}{N}\right). \tag{11.14}$$

The proof follows the argument concerning Lemma 6.3 in [4] almost identically, so we omit it. We now present the final estimate on adjacency matrices comparing switched matrices upon i.i.d. switchings in the sense of matrix perturbations. This will allow us to perform and control resamplings of biregular graphs, in particular using the resolvent perturbation identity.

Lemma 11.9. *Under the setting of the resampling dynamics, we have*

$$\widetilde{A} - A \; = \; T_{b,\mu}(A) - \widetilde{T}_{b,\mu}(\widetilde{A}) \tag{11.15}$$

with probability at least $1 - O(d_b/N)$. Almost surely, we have

$$\widetilde{A} - A \; = \; \sum_{x,y=1}^{O(1)} \Delta_{xy}, \tag{11.16}$$

such that either, conditioning on $\mathscr{G}_\mu$, the random indices x, y are approximately uniform in the corresponding set V_b or V_w or, conditioning on $\mathscr{G}_\mu, p_\mu, \widetilde{p}_\mu$, at least one of the random indices x, y is approximately uniform in the appropriate vertex set.

Lastly, the statement remains true upon switching instead at a white vertex v_w along an edge label ν.

Proof. The result follows from unfolding Lemma 11.8 and the following deterministic identity:

$$\widetilde{A} - A = \mathbf{1}_{\mu \in \widetilde{W}} \left[\widetilde{T}_{b,\mu}(E) - E \right] - \mathbf{1}_{\mu \in W} \left[T_{b,\mu}(A) - A \right]$$

(11.17)
$$+ \sum_{\mu' \in \widetilde{W} \Delta W} \pm \left[T_{b,\mu}(A) - A \right],$$

where the sign corresponds to which of the random sets W or $\widetilde{W}$ contains the indexing label μ'. $\qquad \square$

IV. Green's Function Analysis

12. Switchings on Green's Functions

We first discuss several preliminary estimates which will ultimately help us deduce a self-consistent equation describing the Green's function G of the linearized covariance matrix X. Before we begin with the results, we will assume that all graphs E are *switched*, so that for example the vertices of E are approximately uniformly drawn.

The first of these results is a deterministic estimate controlling Green's functions upon i.i.d. switchings of any fixed biregular graph $E \in \Omega$. To state this estimate, we now introduce the following control parameter for any fixed edge label $\mu \in [[1, d_b]]$ or $\nu \in [[1, d_w]]$:

$$(12.1) \qquad \Gamma_\mu = \Gamma_\mu(z) := \|\Gamma(z)\|_{L^\infty(\mathscr{G}_\mu)},$$

where the notation $L^\infty(\mathscr{G}_\mu)$ in the norm denotes the L^∞-norm conditioning on the σ-algebra $\mathscr{G}_\mu$.

Lemma 12.1. *For any fixed indices $i, j \in [[1, M + N]]$ and any label $\mu \in [[1, d_b]]$, we have*

$$(12.2) \qquad G_{ij} - \widetilde{G}_{ij} = O(d_w^{-1/2} \Gamma_\mu \Gamma).$$

Moreover, suppose x, y are random variables such that, conditioned on $\mathscr{G}_\mu$ and x, the random variable y is approximately uniform. Then we have

$$(12.3) \qquad \mathbb{E}_{\mathscr{G}_\mu} |G_{xy}|^2 = O(\Gamma_\mu^4 \Phi^2).$$

The estimate also holds for any fixed label $\nu \in [[1, d_w]]$ of a white vertex.

Proof. By the resolvent identity, we have

$$G_{ij} - \widetilde{G}_{ij} = \sum_{k,\ell=1}^{M+N} G_{ik}(\widetilde{X} - X)_{k\ell} \widetilde{G}_{\ell j}$$

$$(12.4) \qquad = d_w^{-1/2} \sum_{k=1}^{M} \sum_{\ell=M+1}^{M+N} G_{ik}(\widetilde{A} - A)_{k\ell} \widetilde{G}_{\ell j} + d_w^{-1/2} \sum_{k=M+1}^{M+N} \sum_{\ell=1}^{M} G_{ik} \left(\widetilde{A} - A\right)_{\ell k} \widetilde{G}_{\ell j}.$$

The estimate (12.1) now follows from the almost sure estimate in Lemma 11.9. The proof of (12.3) follows from (12.1) and straightforward calculations; for details, see Lemma 3.9 in [4]. $\qquad \square$

As an immediate corollary, we deduce the following estimate which follows from (12.1) by taking suprema.

Corollary 12.2. *In the setting of Lemma 12.1, we have*

$$(12.5) \qquad \Gamma = \Gamma_\mu + O\left(d_w^{-1/2} \Gamma_\mu \Gamma\right).$$

Having established preliminary, deterministic estimates, we now aim to exploit the approximately uniform structure of the vertices and edges of a switched biregular graph $E \in \Omega$. This will help us deduce stability of the Green's function of a switched graph, though these results will only hold with high probability and are thus non-deterministic. To this end, we introduce the following notion of high-probability, which will closely resemble other notions of high-probability events used in this thesis.

Definition 12.3. Fix a parameter $t = t_N \gg \log N$, and a probability space Ω. We say an event $\Xi \subset \Omega$ holds with *t-high probability*, or *t*-HP for short, if

$$(12.6) \qquad \mathbb{P}\left(\Xi^C\right) \leqslant e^{-t + O(\log N)}.$$

As suggested in the statement of Theorem 4.1, we will take the parameter $t = (\xi \log \xi) \wedge \zeta$. Before we proceed with the Green's function estimates, however, we now illustrate the utility in the notion of t-HP via the following result for t-HP events, whose proof follows from a simple union bound.

Lemma 12.4. *Suppose $\Xi_1, \ldots, \Xi_{O(N)}$ is a collection of events that hold with t-HP. Then the union of these events $\Xi = \cup_n \Xi_n$ holds with t-HP.*

Lastly, we record the following translation of t-HP events in terms of a t-HP bound on conditional expectations. This result follows from a conditional Markov inequality.

Lemma 12.5. *Suppose $\Xi \subseteq \Omega$ is an event that holds with t-HP and let $\mathscr{F}$ denote a σ-algebra on Ω which Ξ is measurable with respect to. Then for any $k = O_{N \to \infty}(1)$, we have with t-HP*

$$(12.7) \qquad \mathbb{E}_{\mathscr{F}} \left(\mathbf{1}_{\Xi^C} \right) \leqslant N^{-k}.$$

Proof. Indeed, we have the following sequence of inequalities:

$$(12.8) \qquad \mathbb{P} \left(\mathbb{E}_{\mathscr{F}} \left(\mathbf{1}_{\Xi^C} \right) > N^{-k} \right) \leqslant N^k \, \mathbb{E} \, \mathbb{E}_{\mathscr{F}} \left(\mathbf{1}_{\Xi^C} \right) \leqslant N^k e^{-t + O(\log N)}.$$

$\square$

We now state the main stability estimate for the Green's function, which, upon taking expectation, provides a precise $o(1)$ estimate on the entries of the Green's function $G(z)$ of the linearization. This result is one of the two important results regarding switchings on Green's functions; it is a bipartite version of Lemma 5.2 in [4]. However, for the ensemble of biregular graphs, the combinatorial constraints on the graph introduce structural differences, which we will address in the proof of this result.

Lemma 12.6. *Fix a vertex $i = v_b \in V_b$ and an edge label $\mu \in [[1, d_b]]$. Suppose $z = E + i\eta \in \mathbb{C}_+$ satisfies the following constraints for a fixed $\varepsilon > 0$:*

$$(12.9) \qquad |E| > \varepsilon, \quad \eta \gtrsim \frac{1}{N}.$$

Suppose further that $\Gamma = O(1)$ holds with t-HP. Then for all fixed indices j, ℓ, r we have

$$\mathbb{E}_{\mathscr{F}_0} \left(G_{v_b,\mu j} - \frac{1}{N} \sum_{k=M+1}^{M+N} G_{kj} \right) = -\mathbb{E}_{\mathscr{F}_0} \left(d_w^{-1/2} s_w G_{ij} \right) + O(d_w^{-1/2} \Phi),$$

$$\mathbb{E}_{\mathscr{F}_0} \left[G_{\ell r} \left(G_{v_b,\mu j} - \frac{1}{N} \sum_{k=M+1}^{M+N} G_{kj} \right) \right] = -\mathbb{E}_{\mathscr{F}_0} \left(G_{\ell r} d_w^{-1/2} s_w G_{ij} \right) + O(d_w^{-1/2} \Phi).$$

Similarly, fix a vertex $k = v_w \in V_w$ and $\mu \in [[1, d_w]]$, and suppose $z \in \mathbb{C}_+$ satisfies the constraints (12.9). Further suppose $\Gamma = O(1)$ with t-HP. Then for all fixed indices j, ℓ, r, we have

$$\mathbb{E}_{\mathscr{F}_0} \left(G_{v_w,\mu j} - \frac{1}{M} \sum_{i=1}^{M} G_{ij} \right) = -\mathbb{E}_{\mathscr{F}_0} \left(d_w^{-1/2} s_b G_{kj} \right) + O(d_w^{-1/2} \Phi),$$

$$\mathbb{E}_{\mathscr{F}_0} \left[G_{\ell r} \left(G_{v_w,\mu j} - \frac{1}{M} \sum_{i=1}^{M} G_{ij} \right) \right] = -\mathbb{E}_{\mathscr{F}_0} \left(G_{\ell r} d_w^{-1/2} s_b G_{kj} \right) + O(d_w^{-1/2} \Phi).$$

Proof. We prove the first estimate only for $v_b \in V_b$; the proof of the second estimate and the estimates for $v_w \in V_w$ are analogous. In expectation, we first note

$$\mathbb{E}_{\mathscr{F}_0} G_{v_b,\mu j} = \mathbb{E}_{\mathscr{F}_0} \widetilde{G}_{\widetilde{v}_b,\mu j}.$$

Moreover, because $G(z)$ is independent of the random variable $\widetilde{v}_{b,\mu}$, and because, conditioned on $\mathscr{G}_\mu$, the random variable $\widetilde{v}_{b,\mu} \in V_w$ is approximately uniform, we also know

$$\mathbb{E}_{\mathscr{F}_0} \left(\frac{1}{N} \sum_{k=M+1}^{M+N} G_{kj} \right) \;=\; \mathbb{E}_{\mathscr{F}_0} G_{\widetilde{v}_{b,\mu}j} + O(d_b^{-1/2}\Phi).$$

Thus, it suffices to compute

$$-\mathbb{E}_{\mathscr{F}_0} \left(G_{\widetilde{v}_{b,\mu}j} - \widetilde{G}_{\widetilde{v}_{b,\mu}j} \right).$$

By the resolvent identity, we have the following equation holding in expectation:

$$(12.10) \qquad \mathbb{E}_{\mathscr{F}_0} \left(-G_{\widetilde{v}_{b,\mu}j} - \widetilde{G}_{\widetilde{v}_{b,\mu}j} \right) \;=\; \mathbb{E}_{\mathscr{F}_0} \left(\sum_{k,\ell} G_{\widetilde{v}_{b,\mu}k}(\widetilde{X} - X)_{k\ell} \widetilde{G}_{\ell j} \right).$$

Unfolding the high-probability equation (11.15) in Lemma 11.9, we have, with probability at least $1 - O(d_w^{-1/2}\Phi)$ conditioned on $\mathscr{F}_0$,

$$(12.11) \qquad \widetilde{X} - X \;=\; d_w^{-1/2} \left(\Delta_{v_b\widetilde{v}_{b,\mu}} - \Delta_{v_b v_{b,\mu}} + \Sigma_b \right) \;+\; d_w^{-1/2} \left(\Delta_{\widetilde{v}_{b,\mu}v_b} - \Delta_{v_{b,\mu}v_b} + \Sigma_b^* \right).$$

Here, we recall that $v_{b,\mu}$ (resp. $\widetilde{v}_{b,\mu}$) is the vertex adjacent to v_b in $S_{v_b,\mu}$ (resp. $\widetilde{S}_{v_b,\mu}$) *after* resampling. Also, Σ_b is the matrix given by a sum of terms $\pm\Delta_{xy}$ where one of the following two conditions holds:

- Conditioned on $\mathscr{G}_\mu, \widetilde{p}_\mu$, the random variable x is approximately uniform, or;
- Conditioned on $\mathscr{G}_\mu$, the random variable y is approximately uniform.

Thus, upon unfolding the RHS of (12.10), we see one term is given by, in expectation,

$$\mathbb{E}_{\mathscr{F}_0} \left[d_w^{-1/2} G_{\widetilde{v}_{b,\mu}\widetilde{v}_{b,\mu}} \widetilde{G}_{ij} \right] \;=\; \mathbb{E}_{\mathscr{F}_0} \left[d_w^{-1/2} \underline{s} \widetilde{G}_{ij} \right] + O(d_w^{-1/2}\Phi)$$
$$=\; \mathbb{E}_{\mathscr{F}_0} \left[d_w^{-1/2} \underline{s} G_{ij} \right] + O(d_w^{-1/2}\Phi),$$

where the first equality holds because $\widetilde{v}_{\beta,\mu}$ is approximately uniform by Corollary 11.6 and the second holds since for any fixed indices i,j, we have $G_{ij} \sim \widetilde{G}_{ij}$ conditioned on $\mathscr{G}_\mu$. In particular, we have

$$\mathbb{E}_{\mathscr{F}_0} \left(G_{\widetilde{v}_{b,\mu}\widetilde{v}_{b,\mu}} \widetilde{G}_{ij} \right) \;=\; \mathbb{E}_{\mathscr{F}_0} \left(G_{\widetilde{v}_{b,\mu}\widetilde{v}_{b,\mu}} G_{ij} \right) \;+\; \mathbb{E}_{\mathscr{F}_0} \left[G_{\widetilde{v}_{b,\mu}\widetilde{v}_{b,\mu}} \left(\widetilde{G}_{ij} - G_{ij} \right) \right]$$
$$=\; \mathbb{E}_{\mathscr{F}_0} \left(G_{\widetilde{v}_{b,\mu}\widetilde{v}_{b,\mu}} G_{ij} \right) + O\left(\frac{1}{\sqrt{D}} \right),$$

where the second equality holds by Lemma 12.1. Thus, it suffices to bound the remaining terms in (12.10). By (12.11), it suffices to estimate the expectation of terms $G_{\widetilde{v}_{b,\mu}x} \widetilde{G}_{yj}$. By the second result in Lemma 12.1 and the Schwarz inequality, in the case where y is approximately uniform conditioning on $\mathscr{G}_\mu$, we have, with high probability,

$$\mathbb{E}_{\mathscr{F}_0} G_{\widetilde{v}_{b,\mu}x} \widetilde{G}_{yj} \;\leqslant\; \mathbb{E}_{\mathscr{F}_0} |\widetilde{G}_{yj}|^2 + O(D^{-1/2}) \;\leqslant\; O(\Phi).$$

Thus, by the assumption $\Gamma = O(1)$, Lemma 12.6 follows after accumulating the finitely many events all holding with probability at least $1 - O(d^{-1/2}\Phi)$. $\qquad\square$

13. Concentration Estimates

We briefly remark here that the result in Lemma 12.6 is an estimate that holds in expectation – a more precise result is seemingly unobtainable as it is impossible to control the resampled neighbor of a fixed vertex $v_b \in V_b$ or $v_w \in V_w$. Our current aim is to show that it suffices to prove stability in expectation – this is the goal of concentration, i.e. to obtain an estimate on the fluctuation of the Green's function about its expectation conditioning on $\mathscr{F}_0$. We will make this discussion more precise as follows, but we remark that the proofs of the following results on concentration are exactly those in Section 4 of [4]. We will briefly remark on the idea of the proofs, however.

The discussion on concentration estimates begins with the following definition ubiquitous throughout probability theory.

Definition 13.1. Suppose X is an L^1-random variable, and suppose $\sigma(\cdot)$ is a σ-algebra which X is measurable with respect to. We define the $\sigma(\cdot)$-*fluctuation* of X to be the following centered random variable:

(13.1)
$$X_{\sigma(\cdot)} \ := \ X - \mathbb{E}_{\sigma(\cdot)}\, X.$$

In words, the fluctuation of a random variable with respect to some σ-algebra detects the concentration of X towards its expectation. Our current goal will thus be to control the fluctuation of polynomials in the Green's function entries. This will allow us to effectively use Lemma 12.6 in studying the stability of Green's functions upon switchings of the underlying graph and adjacency matrix. The main result towards this goal is the following concentration estimate taken from [4] in Section 4 adapted for biregular graphs.

Proposition 13.2. *Suppose that $z = E + i\eta \in \mathbf{D}_{N,\delta,\xi} \cap U_\varepsilon$, that $\zeta > 0$, and that $\Gamma = O(1)$ with probability at least $1 - e^{-\zeta}$. Fix $k = O(1)$ and pairs of indices $I_k = \{(i_1, j_1), \dots, (i_k, j_k)\}$. Define the random variable*

$$X_{I_k}(z) \ = \ G_{i_1 j_1} \dots G_{i_k j_k}.$$

Then, for any $\xi = \xi(N)$ satisfying $\xi \to \infty$ as $N \to \infty$, we have the following pointwise concentration estimate:

(13.2)
$$\mathbb{P}\left[(X_{I_k}(z))_{\mathscr{F}_0} \ = \ O(\xi\Phi) \right] \ \geqslant \ 1 - e^{-[(\xi \log \xi) \wedge \zeta] + O(\log N)}.$$

In order to prove Proposition 13.2, we rely on a more general auxiliary estimate concerned with probability spaces equipped with a family of σ-algebras of the form $\mathscr{F}_\mu$ and $\mathscr{G}_\mu$. In particular, the following result applies to a much wider family of random variables and σ-algebras, and it essentially bounds the fluctuation of a random variable conditioned on $\mathscr{F}_0$ in terms of a priori fluctuation bounds conditioning on the σ-algebras $\mathscr{G}_\mu$, in both a second moment and deterministic sense. In the following statement, we will take $d = d_b$ or $d = d_w$.

Lemma 13.3. *Suppose $\zeta = \zeta(N) > 0$. Suppose X is a $\mathscr{F}_d$-measurable random variable and for each $\mu = 1, \dots, d$ suppose Y_μ is a $\mathscr{G}_\mu$-measurable random variable such that the following probability estimates hold:*

$$\mathbb{P}\left(Y_\mu = O(1) \right) \ \geqslant \ 1 - e^{-\zeta},$$
$$\mathbb{P}\left(Y_\mu \ll N^{O(1)} \right) \ = \ 1.$$

Suppose further that for each μ, we have, almost surely,

(13.3)
$$\left| X_{\mathscr{G}_\mu} \right| \ \leqslant \ Y_\mu,$$

(13.4)
$$\mathbb{E}_{\mathscr{G}_\mu} \left| X_{\mathscr{G}_\mu} \right|^2 \ \leqslant \ d^{-1} Y_\mu.$$

Then, for any $\xi = \xi(N) > 0$, we have the following probability estimate for sufficiently large $N \gg 1$:

(13.5)
$$\mathbb{P}\left(X_{\mathscr{F}_0} \ = \ O(\xi) \right) \ \geqslant \ 1 - e^{-[(\xi \log \xi) \wedge \zeta] + O(\log N)}.$$

A detailed proof of Lemma 13.3 may be found in [4]. We briefly remark on the strategy of the proof; the key insight is to realize the family of random variables $X_\mu = \mathbb{E}_{\mathscr{F}_\mu} X$ defines a martingale with respect to the filtration $\{\mathscr{F}_\mu\}_{\mu \in [[1,d]]}$. The proof of Lemma 13.3 amounts to studying this martingale with martingale concentration estimates and stopping times. We now use Lemma 13.3 to prove Proposition 13.2.

Proof of Proposition 13.2. We appeal to Lemma 13.3, letting $X = \Phi^{-1} G_{ij}$ and $Y_\mu = C_k \Gamma_\mu^{D_k}$, where $C_k, D_k = O_k(1)$ are constants depending only on $k = O(1)$. By the Ward Identity on $\mathbf{D}_{N,\delta,\xi}$ and the assumptions of Proposition 13.2, the assumptions on X, Y_μ in Lemma 13.3 are satisfied.

We first assume $k = 1$ for convenience; from this case, Proposition 13.2 follows for all $k = O(1)$ by a straightforward calculation that can be found in the proof of Proposition 4.1 in [4]. To prove the almost sure estimate (13.3), we appeal to Lemma 12.1. Taking expectation of (12.1) conditioning on $\mathscr{G}_\mu$ and $\widetilde{p}_\mu$, we first see

(13.6)
$$\left| X - \mathbb{E}_{\mathscr{G}_\mu} X \right| \ \leqslant \ \Phi^{-1} \left| G_{ij} - \mathbb{E}_{\mathscr{G}_\mu} G_{ij} \right| \ \leqslant \ O\left(d^{-1/2} \Phi^{-1} \Gamma_\mu \Gamma \right) \ \leqslant \ O\left(\Gamma_\mu^2 \right),$$

where the last bound follows from Corollary 12.2 and the assumption $\Gamma = O(1)$.

To prove the second moment estimate (13.4), we first appeal to the following lemma taken from [4] which helps us compute second moments of $\mathscr{G}_\mu$-fluctuations.

Lemma 13.4. *For any $\mathscr{F}_d$-measurable random variable and any label μ, we have*

$$\mathbb{E}_{\mathscr{G}_\mu} \left| X_{\mathscr{G}_\mu} \right|^2 \;=\; \frac{1}{2} \, \mathbb{E}_{\mathscr{G}_\mu} \left| X - \widetilde{X} \right|^2.$$

Thus, it suffices to estimate the conditional expectation of $|X - \widetilde{X}|^2$. Appealing, again, to Lemma 12.1, we see

$$\left| G_{ij} - \widetilde{G}_{ij} \right|^2 \;\leqslant\; 2d_w^{-1} \left| \sum_{k=1}^{M} \sum_{\ell=M+1}^{M+N} G_{ik}(\widetilde{X} - X)_{k\ell} \widetilde{G}_{\ell j} \right|^2$$

(13.7)
$$+ \; 2d_w^{-1} \left| \sum_{k=M+1}^{M+N} \sum_{\ell=1}^{M} G_{ik}(\widetilde{X} - X)_{k\ell} \widetilde{G}_{\ell j} \right|^2.$$

We now focus on bounding the first term on the RHS of (13.7); bounding the second term will follow from the same argument. By Lemma 11.8, we first note

(13.8)
$$\left| \sum_{k=1}^{M} \sum_{\ell=M+1}^{M+N} G_{ik}(\widetilde{X} - X)_{k\ell} \widetilde{G}_{\ell j} \right|^2 \;\leqslant\; \sum_{x,y,z,w=1}^{O(1)} \left| G_{ix} G_{iy} \widetilde{G}_{zj} \widetilde{G}_{wj} \right|.$$

We first define χ to be the indicator random variable of the event where $\widetilde{H} - H$ is a sum of terms Δ_{xy} such that, conditioned on $\mathscr{G}_\mu$, at least one of x, y is approximately uniform. Conditioning on $\chi = 1$, suppose z, w as in (13.8) are approximately uniform; the other cases for other pairs approximately uniform are analogous. Taking conditional expectation with respect to $\mathscr{G}_\mu$, by (12.3) in Lemma 12.1 and the Schwartz inequality, we finally see

(13.9)
$$\mathbb{E}_{\mathscr{G}_\mu, \chi=1} \left[2d_w^{-1} \left| \sum_{k=1}^{M} \sum_{\ell=M+1}^{M+N} G_{ik}(\widetilde{X} - X)_{k\ell} \widetilde{G}_{\ell j} \right|^2 \right] \;\leqslant\; O(d^{-1}\Phi^2\Gamma^4).$$

Conditioning on $\chi = 0$, which holds with probability at most $O(dM^{-1})$ by Lemma 11.7, we thus immediately see

(13.10)
$$\mathbb{E}_{\mathscr{G}_\mu, \chi=0} \left[2d_w^{-1} \left| \sum_{k=1}^{M} \sum_{\ell=M+1}^{M+N} G_{ik}(\widetilde{X} - X)_{k\ell} \widetilde{G}_{\ell j} \right|^2 \right] \;\leqslant\; O(d^{-1}\Phi^2\Gamma_\mu^4).$$

This yields the desired second moment estimate and thus completes the proof of Proposition 13.2 for the case $k = 1$.

$\square$

14. The Self-Consistent Equation

We now use Lemma 12.6 and Proposition 13.2 to derive a high-probability stability estimate for the diagonal entries of the Green's function G. This stability estimate will follow from studying a *self-consistent equation*, which will allow us to compare the Stieltjes transforms s_b and s_w to their respective limits. Following this stability estimate, we will again appeal to Lemma 7.3 to deduce similar stability estimates for the Green's functions G_* and $G_{*,+}$.

Derivation of the Equation. Before we begin the derivation of the stability estimate, we introduce the following two pieces of notation for a random vector $Z = (Z_i)_{i \in [[1,M]]}$ and $\widetilde{Z} = (\widetilde{Z}_k)_{k \in [[M+1,M+N]]}$:

(14.1)
$$\mathbb{E}_{(i)} \, Z \;=\; \frac{1}{M} \sum_{i=1}^{M} Z_i, \quad \mathbb{E}_{(k)} \, \widetilde{Z} \;=\; \frac{1}{N} \sum_{k=M+1}^{M+N} \widetilde{Z}_k.$$

We begin with the derivation by looking at the matrix equation $HG = zG + \text{Id}$ and computing the diagonal entries of both sides. The (i,i)-entry of the RHS is clearly given by $zG_{ii} + 1$. We now study the LHS, considering the (k,k)-entry for $k \in [[M, M+1]]$. By matrix multiplication we have

$$(14.2) \qquad (HG)_{kk} = \sum_{i=1}^{M} H_{ki}G_{ik} = d_w^{-1/2} \sum_{i=1}^{M} \sum_{\nu=1}^{d_w} \left(\delta_{i,v_{b,\nu}} - \frac{1}{M} \right) G_{ik}$$

$$(14.3) \qquad = d_w^{-1/2} \sum_{\nu=1}^{d_w} \left(G_{v_{b,\nu}k} - \mathbb{E}_{(i)}\, G_{ik} \right),$$

where we used the relation $Md_b = Nd_w$. To compute the averaged difference term in (14.3), we appeal to Lemma 12.6 to deduce the following identity:

$$(14.4) \qquad \mathbb{E}_{\mathscr{F}_0}(HG)_{kk} = -\mathbb{E}_{\mathscr{F}_0} \left(\sum_{\nu=1}^{d_w} \left[\frac{1}{d_w} s_b G_{ii} + O(d_w^{-1}\Phi) \right] \right) = -\mathbb{E}_{\mathscr{F}_0}\, s_b G_{ii} + O(\Phi).$$

Taking an expectation conditioning on $\mathscr{F}_0$ in the matrix equation $HG = zG + \text{Id}$, we see

$$(14.5) \qquad 1 + z\,\mathbb{E}_{\mathscr{F}_0}\, G_{kk} = -\mathbb{E}_{\mathscr{F}_0}\, s_b G_{kk} + O(\Phi).$$

Using Proposition 13.2 to account for the $\mathscr{F}_0$-fluctuation of the Green's function terms, we ultimately deduce a stability equation for the diagonal (k,k)-entries of G, with $k > M$. We may run a similar calculation for indices $i \in [[1, M]]$ and derive the following system of equations:

$$(14.6) \qquad 1 + (z + \gamma s_w)G_{ii} = O\left((1 + |z|)\xi\Phi\right),$$

$$(14.7) \qquad 1 + (z + s_b)G_{kk} = O\left((1 + |z|)\xi\Phi\right).$$

Although this system is a priori coupled, we now appeal to Corollary 7.4 to decouple the equations. More precisely, we deduce the following system of *decoupled* equations:

$$(14.8) \qquad 1 + \left(z + s_b + \frac{1-\gamma}{z} \right) G_{ii} = O\left((1 + |z|)\xi\Phi\right),$$

$$(14.9) \qquad 1 + \left(z + \gamma s_w + \frac{\gamma - 1}{z} \right) G_{kk} = O\left((1 + |z|)\xi\Phi\right).$$

From here, we may proceed in two fashions. First, we may use Lemma 7.3 and Corollary 7.4 to deduce stability equations for the Green's functions G_* and $G_{*,+}$, relating the diagonal entries of these Green's functions to the Stieltjes transforms s_* and $s_{*,+}$. On the other hand, we may also average over the diagonal entries and deduce self-consistent equations for the Stieltjes transforms s_b, s_w and s_*, $s_{*,+}$. We summarize these estimates in the following proposition.

Proposition 14.1. *Suppose $\Gamma = O(1)$ with t-HP, and let $z = E + i\eta \in U_\varepsilon$ satisfy $\eta \gg N^{-1}$. Then for any $i \in [[1, M]]$ and $k \in [[M+1, M+N]]$, we have the following equations uniformly over such z with t-HP:*

$$(14.10) \qquad 1 + \left(z + s_b + \frac{1-\gamma}{z} \right) G_{ii} = O((1 + |z|)\xi\Phi),$$

$$(14.11) \qquad 1 + \left(z + \gamma s_w + \frac{\gamma - 1}{z} \right) G_{kk} = O((1 + |z|)\xi\Phi),$$

$$(14.12) \qquad 1 + (z + 1 - \gamma + zs_{*,+})\,[G_{*,+}]_{ii} = O((1 + |z|^{1/2})\xi\Phi) = O((1 + |z|)\xi\Phi),$$

$$(14.13) \qquad 1 + (z + \gamma - 1 + \gamma zs_*)\,[G_*]_{kk} = O((1 + |z|^{1/2})\xi\Phi) = O((1 + |z|)\xi\Phi).$$

Moreover, we have the following averaged equations uniformly over such z with t-HP:

$$(14.14) \qquad 1 + \left(z + s_b + \frac{1-\gamma}{z}\right) s_b \;=\; O((1+|z|)\xi\Phi),$$

$$(14.15) \qquad 1 + \left(z + \gamma s_w + \frac{\gamma - 1}{z}\right) s_w \;=\; O((1+|z|)\xi\Phi),$$

$$(14.16) \qquad 1 + (z + 1 - \gamma + z s_{*,+})\, s_{*,+} \;=\; O((1+|z|^{1/2}\xi\Phi) \;=\; O((1+|z|)\xi\Phi),$$

$$(14.17) \qquad 1 + (z + \gamma - 1 + \gamma z s_*)\, s_* \;=\; O((1+|z|^{1/2})\xi\Phi) \;=\; O((1+|z|)\xi\Phi).$$

Proof. It remains to upgrade the self-consistent equations (14.10) – (14.17) to hold over all such $z = E + i\eta$ with t-HP. To this end, we appeal to the Lipschitz continuity of the Green's function entries on a sufficiently dense lattice as in the proof of Lemma 8.5. $\qquad\square$

Analysis of the Equation. We recall from a previous discussion in this chapter that the Stieltjes transform m_∞ of the Marchenko-Pastur law with parameter $\gamma \leqslant 1$ is given by the following self-consistent equation:

$$(14.18) \qquad \gamma z m_\infty^2(z) + (\gamma + z - 1)\, m_\infty(z) + 1 \;=\; 0.$$

For the augmented Stieltjes transform $m_{\infty,+}$, we may similarly deduce a self-consistent equation. In our analysis, we will be concerned with providing full details for the Stieltjes transform m_∞ *only*, as the estimate for the augmented transform $m_{\infty,+}$ will follow from the estimate on m_∞. We now note Proposition 14.1 implies the Stieltjes transform s_* solves the same self-consistent equation with an error of $o(1)$ throughout the domain $\mathbf{D}_{N,\delta,\xi} \cap U_\varepsilon$, with t-HP. Our goal will be to use the stability of the self-consistent equation (14.18) under $o(1)$ perturbations to compare s_* and m_∞. This is the content of the following result. To state it, we recall the following control parameter:

$$(14.19) \qquad F_z(r) \;=\; F(r) \;:=\; \left[\left(1 + \frac{1}{\sqrt{(\lambda_+ - z)(z - \lambda_-)}}\right) r\right] \wedge \sqrt{r}.$$

Proposition 14.2. *Let $m : \mathbb{C}_+ \to \mathbb{C}_+$ be the unique solution to the following equation:*

$$(14.20) \qquad \gamma z m^2 + (\gamma + z - 1)m + 1 \;=\; 0.$$

Suppose $s : \mathbb{C}_+ \to \mathbb{C}_+$ is continuous and let

$$(14.21) \qquad R := \gamma z s^2 + (\gamma + z - 1)s + 1.$$

Fix an energy $E \in \mathbb{R} \setminus [-\varepsilon, \varepsilon]$ for $\varepsilon > 0$ small and scales $\eta_0 < C(E)$ and $\eta_\infty \leqslant N$, where $C(E) = O_E(1)$ is a constant to be determined. Suppose we have

$$(14.22) \qquad |R(E + i\eta)| \;\leqslant\; (1 + |z|) r(E + i\eta)$$

for a nonincreasing function $r : [\eta_0, \eta_\infty] \to [0, 1]$. Then for all $z = E + i\eta$ for $\eta \in [\eta_0, \eta_\infty]$, we have the following estimate for sufficiently large N:

$$(14.23) \qquad |m - s| \;=\; O(F(r)).$$

Here, the constant $C(E)$ is determined by

$$(14.24) \qquad \mathrm{Im}\left(\frac{1 - \gamma}{E + i\eta}\right) \;>\; 3\alpha^{1/2}\varepsilon^{-1/2}$$

for all $\eta \leqslant C(E)$.

Before we proceed with the proof of Proposition 14.2, we introduce the following notation.

Notation 14.3. *We denote the solutions to the equation* (14.18) *by* $m_\pm$, *where* m_+ *maps the upper-half plane to itself, and* m_- *maps the upper-half plane to the lower-half plane.*

Moreover, we define the following error functions:

$$v_\pm = |m_\pm - s|. \tag{14.25}$$

Having established this notation, because m_+ takes values in the upper-half plane, we deduce the following upper bound on the values taken by the imaginary part of m_- as follows:

$$\mathrm{Im}(m_-(z)) \leqslant -\mathrm{Im}\left(\frac{1-\gamma}{E+i\eta}\right) < -3\alpha^{1/2}\varepsilon^{-1/2} \tag{14.26}$$

for scales $\eta < C(E)$.

We now proceed to derive an a priori estimate on the error functions $v_\pm$.

Lemma 14.4. *Under the assumptions and setting of Proposition 14.2, we have*

$$|v_+| \wedge |v_-| \leqslant 3\alpha^{1/2}\varepsilon^{-1/2}F(r). \tag{14.27}$$

Proof. We appeal to the following inequality which holds for any branch of the complex square root $\sqrt{\cdot}$ and any complex parameters w, ζ for which the square root is defined:

$$|\sqrt{w+\zeta} - \sqrt{w}| \wedge |\sqrt{w+\zeta} + \sqrt{w}| \leqslant \frac{|\zeta|}{\sqrt{|w|}} \wedge \sqrt{|\zeta|}. \tag{14.28}$$

In particular, this implies the following string of inequalities:

$$|v_+| \wedge |v_-| \leqslant \frac{1}{2\gamma z}\left(\frac{|4\gamma z R|}{\sqrt{|(\gamma+z-1)^2 - 4\gamma z|}} \wedge \sqrt{|4\gamma z R|}\right) \tag{14.29}$$

$$\leqslant \frac{2|R|}{\sqrt{|(\gamma+z-1)^2 - 4\gamma z|}} \wedge \sqrt{\varepsilon\alpha R} \tag{14.30}$$

$$\leqslant \frac{2\varepsilon^{1/2}(1+|z|)r(E+i\eta)}{\sqrt{|(\gamma+z-1)^2 - 4\gamma z|}} \wedge \sqrt{\varepsilon\alpha(1+|z|)r(E+i\eta)} \tag{14.31}$$

where the second inequality follows from the assumption $|z| \geqslant |E| \geqslant \varepsilon$ and the last bound follows if we choose $\varepsilon \leqslant 1$. But this is bounded by $3\alpha^{1/2}\varepsilon^{-1/2}F(r)$ for any $r \in [0,1]$. $\qquad\square$

We now proceed with the proof of Proposition 14.2. Before we present the details of the argument, we provide an outline of the proof. Recall the result of Proposition 14.2 is concerned with an estimate on the quantity v_+. The idea behind the proof is to consider two major regimes for the parameter $z = E + i\eta$. The first regime is the regime where the difference between the solutions m_+ and m_- to the equation (14.18) is dominated by the error term R. In this regime, it suffices to compute an estimate for the minimum $v_+ \wedge v_-$.

In the second regime, we need to run a continuity argument that is structured as follows. The first step is deriving a lower bound for the difference $|v_+ - v_-|$ between the error parameters in the scale window $\eta \in [\eta_0, \eta_\infty]$. Coupled with the estimate on the minimum $v_+ \wedge v_-$, the continuity of the solutions $m_\pm$ and s implies it suffices to estimate v_+ for the boundary value $\eta = \eta_0 = C(E)$ for sufficiently large N. We now make this heuristic precise. For clarity of presentation, we will consider the regimes in the order opposite to the order of presentation in this outline.

Proof. (of Proposition 14.2).

We consider two different regimes. First consider the regime where $|m_+ - m_-| > (1+|z|)r(\eta)$. Precisely, this is the regime defined by $\eta > C(E)$ and the energy-dependent constant $D(E)$ such that

$$(1+|z|)r(\eta) < \frac{|(\gamma+z-1)^2 - 4\gamma z|}{D(E)}; \tag{14.32}$$

the constant $D(E)$ will be determined later. We note by Lemma 14.4, in this regime it suffices to prove the following bound:

$$(14.33) \qquad |v_-| \; > \; |v_-| \wedge |v_+|.$$

We now choose an energy-dependent constant $\kappa(E)$ such that for all $\eta \in [C(E), \eta_\infty]$, we have the bound

$$(14.34) \qquad \alpha^{1/2}(1+|z|)r(\eta) \; \leqslant \; \kappa(E)(1 + |E + iC(E)|).$$

Moreover, note $|(\gamma + z - 1)^2 - 4\gamma z|$ is increasing in η, as seen by translating $z = w + 1 - \gamma$ and computing

$$|(\gamma + z - 1)^2 - 4\gamma z| \;=\; |w^2 - 4\gamma w + X| \;=\; |w(w - 4\gamma) + X|,$$

where $X \in \mathbb{R}$. Because $r(\eta)$ is non-increasing in η, for all $z = E + i\eta$ with $\eta \in [C(E), \eta_\infty]$, we have

$$(14.35) \qquad (1+|z|)r(\eta) \; < \; \frac{\kappa(E)}{D(E)}|(\gamma + z - 1)^2 - 4\gamma z|.$$

We first compute a uniform lower bound on the difference term as follows:

$$
\begin{aligned}
|v_+ - v_-| \;&=\; \frac{|(\gamma + z - 1)^2 - 4\gamma z|}{2\gamma|z|} \\
&\geqslant \; \frac{1}{2\gamma|E + i\eta_\infty|}\left(\frac{D(E)}{\kappa(E)} \frac{(1+|z|)r(\eta)}{\sqrt{|(\gamma+z-1)^2 - 4\gamma z|}} \wedge \sqrt{\frac{D(E)}{\kappa(E)}}\sqrt{\alpha(1+|z|)r(\eta)} \right) \\
&> \; 0.
\end{aligned}
$$

By continuity of s and the estimate Lemma 14.4, choosing $D(E)$ large enough as a function of $\kappa(E), E, \eta_\infty, \varepsilon$, it suffices to prove the estimate (14.23) for some $\eta \in [C(E), \eta_\infty]$. But this follows from Lemma 14.4; in particular, we have at $\eta = C(E)$ and N sufficiently large,

$$|v_-| \;\geqslant\; |\operatorname{Im}(s) - \operatorname{Im}(m_-)| \;\geqslant\; |\operatorname{Im}(m_-)| \;>\; 3\varepsilon^{-1/2} \;\geqslant\; 3\varepsilon^{-1/2}F(r) \;\geqslant\; |v_+| \wedge |v_-|.$$

Thus, we have $|v_+| = |v_+| \wedge |v_-|$ in this first regime, implying the stability estimate (14.23).

Now, we take the regime where the a priori estimate

$$(14.36) \qquad |(\gamma + z - 1)^2 - 4\gamma z| \;=\; O((1+|z|)r(\eta))$$

holds. Thus, we know

$$
\begin{aligned}
|v_-| \;&\leqslant\; |v_+| + \frac{\sqrt{|(\gamma+z-1)^2 - 4\gamma z|}}{2\varepsilon} \;=\; |v_+| + O\left(\frac{(1+|z|)r(\eta)}{\sqrt{|(\gamma+z-1)^2 - 4\gamma z|}} \wedge \sqrt{(1+|z|)r} \right) \\
&=\; |v_+| + O(F(r)),
\end{aligned}
$$

implying the estimate in the second regime as well. $\qquad\square$

We conclude the discussion of the self-consistent equation by noting that Corollary 7.4 allows us to deduce the following local law for the Stieltjes transform $s_{*,+}$:

$$(14.37) \qquad |s_{*,+} - m_{\infty,+}| \;=\; O(F(r)).$$

This estimate holds in the regime $z = E + i\eta$ for $\eta \in [\eta_0, \eta_\infty]$.
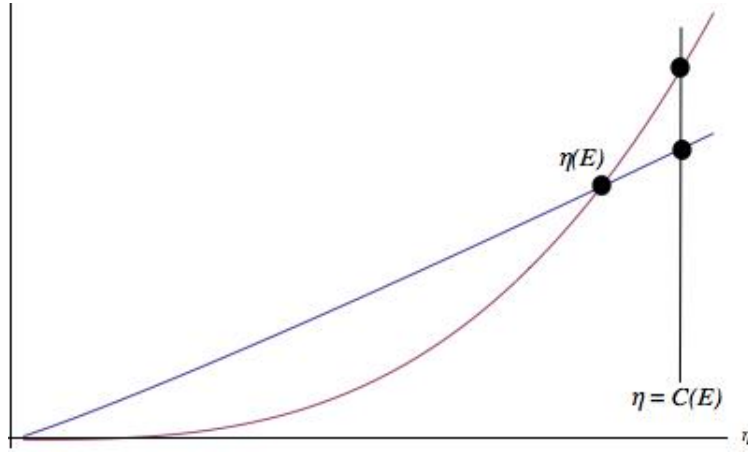
FIGURE 8. This figure shows a contradiction to the inequality $v_- < v_+$ for scales $\eta \in [\eta_0, \eta_\infty]$. Here, the pink curve denotes a model curve for v_- and the purple curve denotes a model curve for v_+. We show $v_- > v_+$ at $\eta = C(E)$. By continuity, if this inequality occurs, then the point $\eta(E)$ where $v_+ = v_-$ must exist. But this contradicts the established lower bound on $|v_+ - v_-|$.

15. THE FINAL GREEN'S FUNCTION ESTIMATES

We conclude this chapter by deducing the estimates for diagonal entries and off-diagonal entries of the Green's functions G_* and $G_{*,+}$ from the local laws for s_* and $s_{*,+}$. In doing so, we will prove Proposition 8.4. We will prove the entry-wise estimates for the Green's function G_*; the details for the entry-wise estimates of $G_{*,+}$ follows from a similar argument.

Fix an index $k \in [[1, N]]$, and for notational convenience, *for this calculation only*, we let G denote the Green's function G_*. Consider the following approximate stability equation for the diagonal entry G_{kk}:

$$(15.1) \qquad 1 + (z + \gamma - 1 + \gamma z s_*)G_{kk} = O\left((1 + |z|)\xi\Phi\right).$$

We now appeal to the following estimate which will allow us to study the stability of this equation upon the replacement $s_* \to m_\infty$ that holds with t-HP:

$$(15.2) \qquad |zG_{kk}| = O(1).$$

Indeed, if $\eta \gg 1$, we have

$$(15.3) \qquad |zG_{kk}| \leqslant \left|\frac{E + i\eta}{\eta}\right| = O(1).$$

If $E \gg 1$, we appeal to the spectral representation of G_{kk} and deduce

$$(15.4) \qquad |zG_{kk}| \leqslant C\left|\frac{1}{N}\sum_\lambda \frac{|\mathbf{u}_\lambda(k)|^2 \times E}{E - \lambda + i\eta}\right|$$

$$(15.5) \qquad \leqslant \frac{C}{N}\sum_\lambda \left|\frac{E}{E - \lambda + i\eta}\right|$$

$$(15.6) \qquad = O(1),$$

where we used the uniform a priori bound $\lambda = O(1)$. If $\eta, E \lesssim 1$, then we appeal to the a priori bound $\Gamma = O(1)$ with t-HP and the trivial bound $z = O(1)$. Thus, by Proposition 14.2, we have

$$(15.7) \qquad 1 + (z + \gamma - 1 + \gamma z m_\infty)G_{kk} = O\left((1 + |z|)\xi\Phi\right) + O(F(\xi\Phi)).$$

On the other hand, the self-consistent equation (14.18) implies the Green's function term on the LHS may be written as $-G_{kk}/m_\infty$. Moreover, because $m_\infty = O(1)$ uniformly on the domain U_ε, we establish the following estimate that holds

over all $z \in \mathbf{D}_{N,\delta,\xi} \cap U_\varepsilon$ with t-HP:

(15.8)
$$m_\infty - G_{kk} = O\left(F(\xi\Phi)\right),$$

where we use the estimate $(1 + |z|)m_\infty = O(1)$. This completes the proof of the local law along the diagonal of $G = G_*$.

To derive the estimate for the off-diagonal entries, we appeal to the Green's function $G(z) = (X - z)^{-1}$ of the linearization X. Note this is *no longer* the Green's function G_* of the covariance matrix X_*. In particular, we appeal to the following entry-wise representation of a matrix equation (for indices $i, j > M$):

(15.9)
$$G_{ij}(HG)_{ii} - G_{ii}(HG)_{ij} = G_{ij}.$$

As in the derivation of the stability equations in Proposition 14.1, by Lemma 12.6 the expectation of the LHS is given by

$$\mathbb{E}_{\mathscr{F}_0}\left[G_{ij}d_w^{-1/2}\underline{s}G_{ii}\right] - \mathbb{E}_{\mathscr{F}_0}\left[G_{ii}d_w^{-1/2}\underline{s}G_{ij}\right] + O(\Phi) = O(\Phi).$$

Thus, at the cost of a concentration estimate in Proposition 13.2, we deduce

(15.10)
$$|G_{ij}| = O(\xi\Phi),$$

which yields the estimate for the off-diagonal entries. By Lemma 7.3, this gives the desired estimate for the off-diagonal entries of the Green's function G_* with t-HP. An analogous calculation proves the desired entry-wise estimates for the Green's function $G_{*,+}$, which completes the proof of Theorem 4.1.

38

Chapter II: Dyson's Brownian Motion for Biregular Bipartite Graphs

I. Dyson Brownian Motion for Adjacency Matrices

1. THE GAUSSIAN MEASURE ON HILBERT SPACES

To define Brownian dynamics for matrices, we begin with the following result constructing Gaussian measures on Hilbert spaces invariant under isometries.

Proposition 1.1. *Suppose $\mathscr{H}$ is a finite-dimensional Hilbert space with basis $\{\mathbf{f}_\alpha\}_\alpha$. Let $\{z_\alpha\}_\alpha$ denote a (finite) collection of scalar-valued Gaussian random variables. Then there exists a Gaussian measure on $\mathscr{H}$ given by the following random vector:*

$$\omega = \sum_\alpha z_\alpha \mathbf{f}_\alpha \tag{1.1}$$

such that the measure induced by ω is invariant under isometries of $\mathscr{H}$. In particular, the Gaussian measure is independent of the choice of basis $\{\mathbf{f}_\alpha\}_\alpha$.

One consequence of the Gaussian measure is the existence of Brownian motions; for any basis $\{\mathbf{f}_\alpha\}_\alpha$, we may define the Brownian motion on $\mathscr{H}$ as

$$B(t) = \sum_\alpha B_\alpha(t)\mathbf{f}_\alpha, \tag{1.2}$$

where the $\{B_\alpha(t)\}_\alpha$ are independent standard one-dimensional Brownian motions. Because the Gaussian measure is invariant under isometries of $\mathscr{H}$ by Proposition 1.1, the Brownian motion $B(t)$ is also invariant under isometries of $\mathscr{H}$.

2. THE GAUSSIAN MEASURE ON ADJACENCY MATRICES

We now consider the Hilbert space $\mathbf{M}_{M \times N}(\mathbb{R})$ of real $M \times N$ matrices equipped with the following trace form:

$$(A, B) = \mathrm{Tr}(A^* B), \tag{2.1}$$

where the star denotes the adjoint of a matrix. Thus, $\mathbf{M}_{M \times N}(\mathbb{R})$ comes equipped with a Gaussian measure. Similarly, we next define the following space of matrices:

$$\mathscr{M} := \left\{ X = \begin{pmatrix} 0 & H \\ H^* & 0 \end{pmatrix} : H \in \mathbf{M}_{M \times N}(\mathbb{R}) \right\}. \tag{2.2}$$

Notation 2.1. *For convenience, if the blocks of $X \in \mathscr{M}$ are given by $H \in \mathbf{M}_{M \times N}(\mathbb{R})$, we will write $X = (H, H^*)$.*

We first note that $\mathscr{M}$ consists only of real symmetric matrices. Moreover, the space $\mathscr{M}$ comes equipped with a Hilbert space structure induced by that of $\mathbf{M}_{M \times N}(\mathbb{R})$, and thus a Gaussian measure as well.

From the perspective of adjacency matrices of bipartite graphs, the space $\mathscr{M}$ is not the space of primary interest, as this space does not reflect their algebraic structure. In particular, such adjacency matrices share the common eigenvector:

$$\mathbf{e}(i) = \begin{cases} \frac{1}{\sqrt{M}} & 1 \leqslant i \leqslant M \\ \frac{1}{\sqrt{N}} & M+1 \leqslant i \leqslant M+N \end{cases}. \tag{2.3}$$

For any (d_b, d_w)-regular bipartite graph with adjacency matrix X, we have

$$X\mathbf{e} = \sqrt{d_b d_w}\,\mathbf{e}. \tag{2.4}$$

Upon the following deterministic shift and scaling in the matrix entries of X, we may assume the vector $\mathbf{e}$ instead corresponds to the eigenvalue $\lambda = 0$:

$$X(0) = d_w^{-1/2} \begin{pmatrix} 0 & A - \frac{d_b}{N}(\mathbf{1}) \\ A^* - \frac{d_w}{M}(\mathbf{1}) & 0 \end{pmatrix}, \tag{2.5}$$

where A denotes the upper-right block of an adjacency matrix and the matrix $\mathbf{1}$ denotes a matrix of the appropriate dimension whose entries are all equal to 1. The matrix $X(0)$ with the above shift and normalization will be referred to as a *normalized adjacency matrix*, or the *normalization* of the adjacency matrix. The details of this transformation in matrix entries are discussed in Chapter I of this thesis. To account for this algebraic invariant of biregular bipartite graphs, we define the following space of matrices:

$$(2.6) \qquad \mathscr{M}_{\mathbf{e}} \;=\; \{X \in \mathscr{M} :\; X\mathbf{e} = 0\}.$$

Clearly, $\mathscr{M}_{\mathbf{e}}$ is a subspace of $\mathscr{M}$ and thus inherits a Gaussian measure that coincides with the Gaussian measure on $\mathscr{M}$ upon integrating out $\mathscr{M}_{\mathbf{e}}^{\perp}$. We now proceed to introduce the following stochastic differential equation (SDE) as motivated by earlier works in universality:

$$(2.7) \qquad \mathrm{d}X(t) \;=\; \frac{1}{\sqrt{N}}\mathrm{d}\mathbf{B}(t) \;-\; \frac{1}{2}X(t)\,\mathrm{d}t,$$

where $\mathbf{B}(t)$ denotes the standard Brownian motion on $\mathscr{M}_{\mathbf{e}}$. The SDE (2.7) is a matrix-valued Ornstein-Uhlenbeck process. We note that on the entire space $\mathscr{M}$, the SDE can be written in coordinates by running each independent matrix entry through a one-dimensional Ornstein-Uhlenbeck process.

Remark 2.2. Although $X(t)$ contains only $N-1$ nontrivial eigenvalues, we retain the normalization $N^{-1/2}$ in the SDE (2.7).

To better study the above matrix-valued SDE, however, we still need to choose coordinates. The advantage of the abstract construction of the Gaussian measure is the confidence that we may freely choose coordinates without changing the measure. The suitable choice of coordinates is given by the singular value decomposition of matrices in $\mathbf{M}_{M \times N}(\mathbb{R})$. We summarize this in the following parameterization of $\mathscr{M}_{\mathbf{e}}$.

Lemma 2.3. *Suppose $X \in \mathscr{M}_{\mathbf{e}}$ has the block representation $X = (H, H^*)$. Then, for some matrix $\widetilde{X} \in \mathbf{M}_{(M-1)\times(N-1)}(\mathbb{R})$,*

$$(2.8) \qquad H \;=\; O(M) \times (\widetilde{X} \oplus 0) \times O(N)^*,$$

where the matrices $O(M), O(N)$ are orthogonal of dimension M and N, respectively. Here, multiplication on the RHS is multiplication as matrices. Moreover, under the induced map

$$(2.9) \qquad \mathscr{M}_{\mathbf{e}} \;\to\; \mathbf{M}_{(M-1)\times(N-1)}(\mathbb{R}),$$

the Gaussian measure is invariant, where the latter space is equipped with the same inner product given by (2.1).

Proof. The parameterization (2.8) comes from the singular value decomposition of H. The invariance of the Gaussian measure is a consequence of the following straightforward calculation:

$$(2.10) \qquad \mathrm{Tr}(A^*B) \;=\; \mathrm{Tr}\left(O(N)\left(\widetilde{A}^* \oplus 0\right)O(M)^*O(M)\left(\widetilde{B} \oplus 0\right)O(N)^*\right)$$

$$(2.11) \qquad\qquad\quad =\; \mathrm{Tr}\left(O(N)\left(\widetilde{A}^*\widetilde{B} \oplus 0\right)O(N)^*\right)$$

$$(2.12) \qquad\qquad\quad =\; \mathrm{Tr}\left(\widetilde{A}^*\widetilde{B}\right).$$

$\square$

3. Analysis of Switchings on the Generator

We now briefly review the background on switchings on graphs discussed in detail in Chapter I. Suppose E is a biregular bipartite graph, and consider a pair of edges $e_1 = ij, e_2 = mn \in E$ with four distinct vertices. We let E_{e_1,e_2} denote the subgraph of E with edges e_1, e_2.

Definition 3.1. We say a *simple switching* of E at E_{e_1,e_2} is the following perturbed graph:

$$(3.1) \qquad E_s \;=\; E - E_{e_1,e_2} + E_{in,mj},$$

where the operations on the RHS are understood in the sense of adjacency matrices.

Remark 3.2. We note that the switched graph E_s suppresses from its notation the dependence of e_1, e_2. This will not be important, but we emphasize it now for clarity's sake.
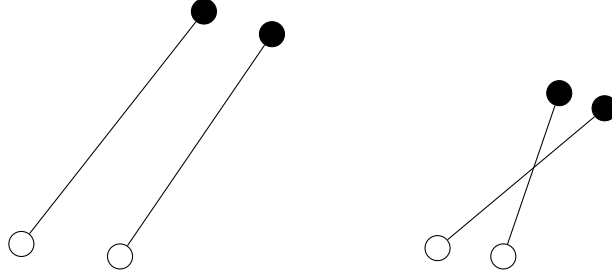
FIGURE 9. We illustrate a simple switching on E_{e_1,e_2} on the LHS and E_s on the RHS.

As in Chapter I of this thesis, we now interpret the combinatorics of switchings in terms of the corresponding adjacency matrices. To do so, we first define the following matrix:

$$(3.2) \qquad \xi_{ij}^{mn} = \Delta_{ij} + \Delta_{mn} - \Delta_{in} - \Delta_{mj},$$

where Δ_{xy} denotes the adjacency matrix of the graph whose only edge is xy. With this notation, we may easily deduce

$$(3.3) \qquad A(E_s) = A(E) + \xi_{ij}^{mn},$$

where $A(E_s)$ (resp. $A(E)$) denotes the adjacency matrix of the graph E_s (resp. E). We now define the following:

$$(3.4) \qquad X_{ij}^{mn} = \mathrm{Tr}\left(\xi_{ij}^{mn} X\right) = 2\left(X_{ij} + X_{mn} - X_{in} - X_{mj}\right),$$

$$(3.5) \qquad \partial_{ij}^{mn} = \mathrm{Tr}\left(\xi_{ij}^{mn} \partial\right) = 2\left(\partial_{ij} + \partial_{mn} - \partial_{in} - \partial_{mj}\right).$$

To motivate the above definitions (3.4) and (3.5), we compute the following Taylor approximation: for any $F \in C^2(\mathscr{M}_\mathbf{e})$,

$$(3.6) \qquad F(X + \xi_{ij}^{mn}) = F(X) + \left[\partial_{ij}^{mn} F\right](X) \times X_{ij}^{mn} + \frac{1}{2}\left[\partial_{ij}^{mn} F\right](X) \times (X_{ij}^{mn})^2 + \mathrm{Error}.$$

In particular, (3.4) and (3.5) come from Taylor expanding. This will be crucial in analyzing the generator of (2.7) as we will need to compute Taylor estimates along matrices corresponding to switchings ξ_{ij}^{mn}. To formalize the directions along which we will Taylor expand, we define the following set of all switchings:

$$(3.7) \qquad \mathscr{X} := \bigcup_{(i,j)\in V_b} \bigcup_{(m,n)\in V_b} \{\xi_{ij}^{mn}\}.$$

The terms (3.4) and (3.5) also determine the structure of the generator of (2.7), which we state in the following result.

Proposition 3.3. *The generator of the OU process on $\mathscr{M}_\mathbf{e}$ is given by*

$$(3.8) \qquad \mathscr{L} = \frac{1}{8MN^2} \sum_{i,j,k,\ell} \left(\partial_{ij}^{k\ell}\right)^2 - \frac{1}{16MN} \sum_{i,j,k,\ell} X_{ij}^{k\ell} \partial_{ij}^{k\ell}.$$

I.e., for any $F \in C^2(\mathscr{M}_\mathbf{e})$, we have

$$(3.9) \qquad \partial_t \,\mathbb{E}\, F(X(t)) = \mathbb{E}(\mathscr{L}F)(X(t)).$$

We now devote the remainder of this discussion towards proving Proposition 3.3. We begin by recalling the following notions of black and white vertices from Chapter I.

Definition 3.4. Define V_b to be the set of indices $\{(i,j)\}$ such that $i, j - M \in [1, M]$. Similarly, define V_w to be those indices (k, ℓ) such that $\ell, k - M \in [1, M]$.

The proof of Proposition 3.3 will amount to an application of the Ito formula. To apply this formula, however, we need to compute the quadratic covariations of the matrix entries $\mathrm{d}X_{ij}$, where we choose coordinates as in Lemma 2.3. We record these covariation processes in the following lemma.

Lemma 3.5. *Fix any two indices (i, j) and (k, ℓ). Then, we have*

(3.10)
$$\mathrm{d}\langle X_{ij}, X_{k\ell}\rangle \;=\; \begin{cases} \frac{1}{N}\left(\delta_{ik} - \frac{1}{M}\right)\left(\delta_{j\ell} - \frac{1}{N}\right): & (i,j), (k,\ell) \in V_b; \\ \frac{1}{N}\left(\delta_{ik} - \frac{1}{N}\right)\left(\delta_{j\ell} - \frac{1}{M}\right): & (i,j), (k,\ell) \in V_w; \\ \frac{1}{N}\left(\delta_{i\ell} - \frac{1}{M}\right)\left(\delta_{jk} - \frac{1}{N}\right): & (i,j) \in V_b, (k,\ell) \in V_w; \\ \frac{1}{N}\left(\delta_{i\ell} - \frac{1}{N}\right)\left(\delta_{jk} - \frac{1}{M}\right): & (i,j) \in V_w, (k,\ell) \in V_b. \end{cases}$$

Proof. We consider the case $(i, j), (k, \ell) \in V_b$; the other cases follow similarly. As in Lemma 2.3, we may assume the normalized adjacency matrix X is of the form

(3.11)
$$X = O(M)\left(\widetilde{X} \oplus 0\right)O(N)^*$$

where $\widetilde{X} \in \mathbf{M}_{(M-1)\times(N-1)}(\mathbb{R})$ and the orthogonal component corresponds to the span of the eigenvector $\mathbf{e}$ and is thus constant in time. Moreover, we may assume the Gaussian measure on $\mathbf{M}_{(M-1)\times(N-1)}(\mathbb{R})$ is given by drawing each entry from independent standard one-dimensional Gaussian distributions. This also implies that the Ornstein-Uhlenbeck process on $X \in \mathscr{M}_{\mathbf{e}}$ is equal, in law, to letting each matrix entry in $\widetilde{X}$ run independent one-dimensional Ornstein-Uhlenbeck processes up to the symmetry constraint. This implies the quadratic covariation process for $\widetilde{X}(t)$ is given by

(3.12)
$$\mathrm{d}\langle \widetilde{X}_{ij}(t), \widetilde{X}_{k\ell}(t)\rangle \;=\; \frac{1}{N}\delta_{ik}\delta_{j\ell}$$

where we use the assumption $(i, j), (k, \ell) \in V_b$. Thus, under this same assumption, we compute

(3.13)
$$\mathrm{d}\langle H_{ij}, H_{k\ell}\rangle \;=\; \sum_{m,n}\sum_{x,y} [O(M)_{im}O(N)_{jn}O(M)_{kx}O(N)_{\ell y}]\,\mathrm{d}\langle \widetilde{X}_{mn}, \widetilde{X}_{xy}\rangle$$

(3.14)
$$= \sum_{m,n}\sum_{x,y} [O(M)_{im}O(N)_{jn}O(M)_{kx}O(N)_{\ell y}] \times \frac{1}{N}\delta_{mx}\delta_{ny}$$

(3.15)
$$= \frac{1}{N}\sum_{m,n} O(M)_{im}O(N)_{jn}O(M)_{km}O(N)_{\ell n}$$

(3.16)
$$= \frac{1}{N}\left(\sum_{m=1}^{M-1} O(M)_{im}O(M)_{km}\right)\left(\sum_{n=1}^{N-1} O(N)_{jn}O(N)_{\ell n}\right).$$

Because the matrices $O(M), O(N)$ give the SVD of the matrix $H(t)$, we know $O(M)_{iM} = M^{-1/2}$ and $O(N)_{kN} = N^{-1/2}$ for any suitable indices i, k. With this and the assumption that $O(M), O(N)$ are orthogonal matrices, we see

(3.17)
$$\sum_{m=1}^{M-1} O(M)_{im}O(M)_{km} \;=\; \delta_{ik} - \frac{1}{M},$$

(3.18)
$$\sum_{n=1}^{N-1} O(N)_{jn}O(N)_{\ell n} \;=\; \delta_{j\ell} - \frac{1}{N},$$

which completes the derivation of the covariation processes in the case $(i, j), (k, \ell) \in V_b$. □

We are now in a position to deduce Proposition 3.3 from the Ito formula. First, we establish the following shorthand for differentiation of any sufficiently smooth function F on a matrix space and indices (i, j):

(3.19)
$$\partial_{ij}F(H) \;:=\; \left[\partial_{H_{ij}}F\right](H).$$

We now proceed with the proof of Proposition 3.3. By the Ito formula, for any $F \in C^2(\mathscr{M}_{\mathbf{e}})$, we have

(3.20)
$$\mathrm{d}F(H) \;=\; \sum_{i,j=1}^{M+N} [\partial_{ij}F](H)\,\mathrm{d}H_{ij} + \frac{1}{2}\sum_{i,j,k,\ell=1}^{M+N} [\partial_{ij}\partial_{k\ell}F](H)\,\mathrm{d}\langle H_{ij}, H_{k\ell}\rangle.$$

Taking expectation, the martingale term in $\mathrm{d}H_{ij}$ vanishes; by construction, we thus have

$$(3.21) \qquad \mathrm{d}\,\mathbb{E}\,F(H) \;=\; -\frac{1}{2}\sum_{i,j=1}^{M+N}\mathbb{E}\left[H_{ij}\left[\partial_{ij}F\right](H)\right]\mathrm{d}t \;+\; \frac{1}{2}\sum_{i,j,k,\ell=1}^{M+N}\mathbb{E}\left[\left[\partial_{ij}\partial_{k\ell}F\right](H)\right]\mathrm{d}\langle H_{ij},H_{k\ell}\rangle.$$

It remains to compute the covariation $\mathrm{d}\langle H_{ij},H_{k\ell}\rangle$; this is computed in Lemma 3.5. Thus, we have

$$(3.22) \qquad MN^2\sum_{i,j,k,\ell=1}^{M+N}\mathbb{E}\left[\left[\partial_{ij}\partial_{k\ell}F\right](H)\right]\mathrm{d}\langle H_{ij},H_{k\ell}\rangle \;=\; \sum_{(i,j),(k,\ell)\in V_b}\mathbb{E}\left[\left[\partial_{ij}(\partial_{ij}+\partial_{k\ell}-\partial_{i\ell}-\partial_{jk})F\right](H)\right]\mathrm{d}t$$

$$(3.23) \qquad\qquad\qquad\qquad + \sum_{(i,j),(k,\ell)\in V_w}\mathbb{E}\left[\left[\partial_{ij}(\partial_{ij}+\partial_{k\ell}-\partial_{i\ell}-\partial_{jk})F\right](H)\right]\mathrm{d}t$$

$$(3.24) \qquad\qquad\qquad\qquad + \sum_{(i,j)\in V_b,(k,\ell)\in V_w}\mathbb{E}\left[\left[\partial_{ij}(\partial_{ij}+\partial_{k\ell}-\partial_{ik}-\partial_{j\ell})F\right](H)\right]\mathrm{d}t$$

$$(3.25) \qquad\qquad\qquad\qquad + \sum_{(i,j)\in V_w,(k,\ell)\in V_b}\mathbb{E}\left[\left[\partial_{ij}(\partial_{ij}+\partial_{k\ell}-\partial_{ik}-\partial_{j\ell})F\right](H)\right]\mathrm{d}t.$$

We note however, upon the bijection $(i,j)\mapsto(k,\ell)$, that the summations given by the RHS of (3.22) and (3.23) are equal. Similarly, we see the summations given by (3.24) and (3.25) are also equal. Lastly, we see (3.22) and (3.24) are equal upon switching the indices k,ℓ. Thus, because the process $\mathbb{E}\,F(H)$ contains no diffusion term,

$$(3.26) \qquad \mathscr{L} \;=\; \frac{2}{MN^2}\sum_{(i,j),(k,\ell)\in V_b}\partial_{ij}\left(\partial_{ij}+\partial_{k\ell}-\partial_{i\ell}-\partial_{jk}\right) \;-\; \frac{1}{2}\sum_{i,j=1}^{M+N}H_{ij}\partial_{ij}$$

$$(3.27) \qquad = \frac{2}{MN^2}\sum_{(i,j),(k,\ell)\in V_b}\partial_{ij}\left(\partial_{ij}+\partial_{k\ell}-\partial_{i\ell}-\partial_{jk}\right) \;-\; \sum_{(i,j)\in V_b}H_{ij}\partial_{ij},$$

where the second equality (3.27) holds since H is symmetric. To understand the second-order terms in (3.27), we claim

$$(3.28) \qquad \sum_{(i,j),(k,\ell)\in V_b}\partial_{ij}\left(\partial_{ij}+\partial_{k\ell}-\partial_{i\ell}-\partial_{jk}\right) \;=\; \frac{1}{4}\sum_{(i,j),(k,\ell)\in V_b}\left(\partial_{ij}+\partial_{k\ell}-\partial_{i\ell}-\partial_{jk}\right)^2.$$

Indeed, (3.28) follows from the fact that we are summing over all indices $(i,j),(k,\ell)\in V_b$. For the same reason, as well as the assumed relations $\sum_j H_{ij}=\sum_i H_{ij}=0$, we also have

$$(3.29) \qquad \sum_{(i,j)\in V_b}H_{ij}\partial_{ij} \;=\; \frac{1}{4MN}\sum_{(i,j),(k,\ell)\in V_b}\left(H_{ij}+H_{k\ell}-H_{i\ell}-H_{jk}\right)\left(\partial_{ij}+\partial_{k\ell}-\partial_{i\ell}-\partial_{jk}\right).$$

This completes the proof of Proposition 3.3. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We end this discussion by remarking on the nature of the generator $\mathscr{L}$, in particular its representation via differential operators with coefficients all coming from simple switchings of graphs in Ω. This discrete-probabilistic interpretation of the generator $\mathscr{L}$ will be important in studying the short-time stability of certain matrix statistics.

4. The Main Result: Eigenvalue Statistics

We now state the main result of this chapter, giving short-time stability of eigenvalue statistics. To make this precise, we introduce the following eigenvalue statistic.

Definition 4.1. For a random matrix ensemble H of dimension N, we let $\varrho_{H,N}(\lambda_1,\ldots,\lambda_N)$ denote the density function for the law of the eigenvalues of H on the simplex $\lambda_1<\lambda_2<\ldots<\lambda_N$. For any $1\leqslant n\leqslant N$, we define the *n-point correlation function*, denoted by $\varrho_{H,N}^{(n)}$, by the following marginal formula for any $\lambda_1<\ldots<\lambda_n$:

$$(4.1) \qquad \varrho_{H,N}^{(n)}(\lambda_1,\ldots,\lambda_n) \;=\; \int_{\lambda_n<x_{n+1}<\ldots<x_N}\varrho_{H,N}(\lambda_1,\ldots,\lambda_n,x_{n+1},\ldots,x_N)\,\mathrm{d}x_{n+1}\ldots\mathrm{d}x_N.$$

We now make precise a notion of coincidence of eigenvalue correlation statistics, in an averaged sense.

Definition 4.2. Suppose H_1 and H_2 are two random matrix ensembles of equal dimension N in a matrix space $\mathbf{M}$, e.g. Wigner matrices, Wishart matrices, and the space $\mathscr{M}_\mathbf{e}$. We say the *averaged bulk eigenvalue correlation statistics* of H_1 and H_2 coincide at the energy E_0 if the following holds.

For any $n \in \mathbb{Z}_{\geqslant 0}$, any test function $\varphi \in C_c^\infty(\mathbb{R})$, and a constant $c > 0$ sufficiently small, we have for $b = N^{-1+c}$

$$(4.2) \qquad \frac{1}{2b} \int_{E_0 - b}^{E_0 + b} \mathrm{d}E' \int_{\mathbb{R}^n} \varphi(x_1, \ldots, x_n) N^n \left(\varrho_{H_1}^{(n)} - \varrho_{H_2}^{(n)} \right) \left(E' + \frac{dx_1}{N\varrho_\infty(E_0)}, \ldots, E' + \frac{dx_n}{N\varrho_\infty(E_0)} \right) = o_{N\to\infty}(1)$$

where $\varrho_{H_i}^{(n)}$ denotes the n-point correlation function of the matrix ensemble H_i (for $i = 1, 2$). Here, we also use ϱ_∞ to denote the density function of either the Marchenko-Pastur law or the semicircle law depending on if the random matrix ensembles H_1, H_2 are covariance matrix ensembles or Wigner matrix ensembles, respectively.

In particular, Definition 4.2 requires a small average around the energy E_0. We note that there are results, e.g. in [16] in the ensemble of Wigner matrices, that provide similar results without an average of the energy; universality results along this line are known as *fixed energy* universality results. Although it is believed that the arguments in [16] extend to linearized covariance matrices, we do not pursue that in this thesis.

To state the main theorem of this chapter, we will introduce the following notation for the covariance matrix ensembles (and their linearization ensembles) at a given time.

Notation 4.3. *For a given time $t \geqslant 0$, we let $\mathscr{X}_*(t)$ denote the random matrix ensemble of matrices $X_*(t) = H(t)^* H(t)$, where the matrix $H(t)$ solves the matrix-valued Ornstein-Uhlenbeck equation*

$$(4.3) \qquad \mathrm{d}H(t) = \frac{1}{\sqrt{N}} \mathrm{d}\mathbf{B}(t) - \frac{1}{2} H(t)\, \mathrm{d}t$$

with initial data $H(0)$ the upper-right block of the normalized adjacency matrix $X(0) = (H(0), H^*(0))$ of a graph in Ω.

Similarly, we let $\mathscr{X}(t)$ denote the random matrix ensembles of linearizations $X(t) = (H(t), H(t)^*)$.

We now state the main theorem, which states that for any energy E in the bulk of ϱ_{MP}, the averaged bulk eigenvalue correlation statistics of the random matrix ensembles $\mathscr{X}_*(0)$ and $\mathscr{X}_*(N^{-1-\zeta}D^{1/2})$ agree for a small $\zeta > 0$. Here, we define the bulk of ϱ_{MP} to be those energies E a fixed distance from the edges. More precisely, the bulk is defined as

$$(4.4) \qquad \mathscr{I}_{\mathrm{MP}, \varepsilon} = \left[\varepsilon, \, (1 - \varepsilon)(1 + \sqrt{\gamma})^2 \right].$$

Here, $\varepsilon > 0$ is a fixed (small) constant as in the definition of the domains U_ε and $U_{\varepsilon, \delta}$. We also recall the definition $\gamma := 1/\alpha$. Similarly, we may define the bulk of the linearization to be

$$(4.5) \qquad \mathscr{I}_{\mathrm{linear}, \varepsilon} = \pm\sqrt{\mathscr{I}_{\mathrm{MP}, \varepsilon}} = \pm\left[\sqrt{\varepsilon}, \sqrt{1 - \varepsilon}(1 + \sqrt{\gamma}) \right]$$

We now state the main theorem, which serves as the second step in the three-step strategy discussed in the introduction of this thesis.

Theorem 4.4. *Suppose $\varepsilon > 0$ and $\zeta > 0$ are fixed constants. Then, for any $t \in [0, N^{-1-\zeta}D^{1/2}]$ and any energy $E \in \mathscr{I}_{\mathrm{linear}, \varepsilon}$, the averaged bulk eigenvalue correlation statistics of $\mathscr{X}(0)$ and $\mathscr{X}(t)$ coincide.*

Thus, for any $t \in [0, N^{-1-\zeta}D^{1/2}]$ and any energy $E \in \mathscr{I}_{\mathrm{MP}, \varepsilon}$, the averaged bulk correlation statistics of $\mathscr{X}_*(0)$ and $\mathscr{X}_*(t)$ coincide.

II. Short Time Stability of Dyson's Brownian Motion for Graphs

The remainder of this chapter will now be devoted towards studying the stability of eigenvalue statistics along the following matrix-valued Ornstein-Uhlenbeck dynamics:

$$(5.1) \qquad \mathrm{d}X(t) \;=\; \frac{1}{\sqrt{N}}\,\mathrm{d}\mathbf{B}(t) \;-\; \frac{1}{2}X(t)\,\mathrm{d}t,$$

where we recall $X(0) = (H(0), H(0)^*)$ is a normalized adjacency matrix. We recall the generator of the SDE is given by

$$(5.2) \qquad \mathscr{L} \;=\; \frac{1}{8MN^2} \sum_{i,j,k,\ell} \left(\partial_{ij}^{k\ell}\right)^2 \;-\; \frac{1}{16MN} \sum_{i,j,k,\ell} X_{ij}^{k\ell}\partial_{ij}^{k\ell}.$$

Our goal now will be to estimate the stability of eigenvalue statistics driven by this generator. To state the general result, we first define the following deterministic norms beginning with a L^r-seminorm on $C^0(\mathscr{M}_{\mathbf{e}})$:

$$(5.3) \qquad \|F\|_{r,t} \;=\; \left(\mathbb{E}\,|F(X(t))|^r\right)^{1/r},$$

where the expectation is taken over the randomness of $X(t)$. We extend this seminorm for derivatives: for any $F \in C^k(\mathscr{M}_{\mathbf{e}})$:

$$(5.4) \qquad \|\partial^k F\|_{r,t} \;:=\; \left\| \sup_{\theta_i \in [0,1]} \sup_{X_i \in \mathscr{X}} \left| \partial_{X_1} \dots \partial_{X_k} F\left(\cdot + d_b^{-1/2} \sum_{i=1}^k \theta_i X_i \right) \right| \right\|_{r,t}.$$

In particular for $k = 0$ the seminorms (5.3) and (5.4) coincide. We now recall the sparsity parameter from Chapter I:

$$(5.5) \qquad D \;=\; d_b \wedge \frac{N^2}{d_b^3}.$$

Remark 5.1. In this thesis we will take $d \in [N^\gamma, N^{2/3-\gamma}]$, which gives us the lower bound $D \geqslant N^\gamma$. This lower bound will be important in the arguments to follow.

We may now state the stability estimate. All adjacency matrices are of bipartite graphs discussed in Chapter I.

Theorem 5.2. *Suppose $X(t)$ solves the SDE* (5.1) *with initial condition $X(0)$ a normalized adjacency matrix. Moreover, suppose $r(\varepsilon)$ is sufficiently large as a function of a fixed $\varepsilon > 0$. Then, for any $F \in C^4(\mathscr{M}_{\mathbf{e}})$, we have*

$$(5.6) \qquad \mathbb{E}\,F(X(t)) - \mathbb{E}\,F(X(0)) \;=\; O\left(D^{-1/2} N^{1+\varepsilon} \max_{1 \leqslant i \leqslant 4} \int_0^t \|\partial^i F\|_{r,s}\,\mathrm{d}s \right).$$

Remark 5.3. For $F \in C^4(\mathscr{M}_{\mathbf{e}})$, if the norms $\|\partial^i F\|_{r,s}$ are uniformly bounded, then for short times the RHS is $o(1)$ for large N. If the function F encodes local eigenvalue statistics, then we have at least one notion of comparing eigenvalue statistics of the initial data $X(0)$ and $X(t)$ for short times.

However, if we allow longer times, then the error term on the RHS diverges. *This is characteristic, as of now, to the sparsity of the matrix.* More precisely, to upgrade the stability to longer times, we need a stronger bound on $D^{-1/2}$ which is not achievable for any allowed value of d, i.e. any sparse matrix.

We now outline the proof of Theorem 5.2. The idea of the proof relies on exploiting the graph structure appearing in the first- and second-order coefficients in the generator $\mathscr{L}$ given in (5.2). As noted in an earlier discussion, the differential operators, and the coefficient for the first-order term, in $\mathscr{L}$ may be viewed as coming from differentiating along directions

ξ_{ij}^{mn} induced by simple switchings of graphs. In fact, in proving Theorem 5.2, we will use the switching dynamics to construct a Markovian jump process with generator $\mathcal{Q}$ that approximates (5.1) as follows:

$$(5.7) \qquad \mathbb{E}\,\mathcal{Q}F(X(t)) \,=\, \mathbb{E}\,\mathscr{L}F(X(t)) \,+\, O\left(D^{-1/2}N^{1+\varepsilon}\right) \max_{1\leqslant i\leqslant 4} \|\partial^i F\|_{r(\varepsilon),0}.$$

In other words, we will use the underlying graph structure of $\mathscr{L}$ to compare it to the generator of a continuous-time jump process that will be accessible to study via tractable combinatorial methods. We give a detailed construction of the process determined by $\mathcal{Q}$ in a later discussion; most important is the invariance of the uniform measure on Ω with respect to the dynamics generated by $\mathcal{Q}$. This resembles the invariance of the uniform probability measure under switching dynamics in Chapter I. As a consequence, the LHS vanishes. By the Dynkin formula, we will then deduce the desired short-time stability.

6. The Jump Process and Short-Time Stability

In order to define the jump process, we discuss more in detail the dynamics on Ω generated by simple switchings. For the remainder of this discussion, we will often use the notation A to denote both a biregular bipartite graph and its unnormalized adjacency matrix. In this spirit, we will retain the notation from Chapter I and let $[A]$ denote the set of vertices in A. Moreover, the following discussion will resemble much of the discussion on switching dynamics in Chapter I.

We begin by defining the following (random) indicator function for vertices $i, j, m, n \in [A]$:

$$(6.1) \qquad I_{ij}^{mn}(A) \,=\, A_{ij}A_{mn}\left(1 - A_{in}\right)\left(1 - A_{mj}\right).$$

In words, the indicator function I_{ij}^{mn} detects whether or not the edges ij and mn exist in A. Moreover, if these edges exist in A, then the function also detects whether or not the subgraph $A_{ij,mn}$ is 1-regular. We now use the function I_{ij}^{mn} to define the following Markovian jump process via its generator:

$$(6.2) \qquad \mathcal{Q}f(A) \,=\, \frac{1}{4Nd_w} \sum_{(i,j)\in V_b} \sum_{(m,n)\in V_b} I_{ij}^{mn}(A)\left[f(A - \xi_{ij}^{mn}) - f(A)\right].$$

Remark 6.1. By the basic theory of point processes, we may give the following description for the process generated by $\mathcal{Q}$. Fix a graph $A \in \Omega$. Events defining this process are dictated by a Poisson clock depending on N and d_b. At each event, two edges $ij, mn \in A$ are selected uniformly at random. If $I_{ij}^{mn}(A) = 1$, then perform a simple switching. The normalization factor serves to match with the generator $\mathscr{L}$ as in Theorem 5.2.

Using this probabilistic description of the jump process defined by $\mathcal{Q}$, we may deduce the following result describing the generator $\mathcal{Q}$ in relation to the uniform probability measure on Ω.

Proposition 6.2. *Let μ_{unif} denote the uniform probability measure on Ω. Then, μ_{unif} is invariant under the generator $\mathcal{Q}$.*

The idea of the proof of Proposition 6.2 is as follows. Because simple switchings of graphs are an invertible operation, the uniform probability measure on the discrete space Ω should be invariant upon applying simple switchings to graphs. To give a rigorous proof of Proposition 6.2, however, we need to rigorously define dynamics on Ω induced by simple switchings. For now, we will take Proposition 6.2 for granted and use it to prove Theorem 5.2. We will address the details of the simple switching dynamics in a later discussion; this discussion has been carried out rigorously in Chapter I in the case of double switchings, but we will still provide a discussion for simple switchings in detail.

In order to compare the generator $\mathcal{Q}$ and the generator $\mathscr{L}$, we first note these two operators act on different function spaces. More precisely, $\mathcal{Q}$ acts on functions on Ω, i.e. on unnormalized adjacency matrices. On the other hand, $\mathscr{L}$ acts on functions on $\mathcal{M}_{\mathbf{e}}$, within which Ω embeds by taking the normalization of an adjacency matrix. Thus, for these operators to agree, we need to establish the following convention. First, we recall the normalization of an adjacency matrix with upper-right off-diagonal block A as follows:

$$(6.3) \qquad X_A \,=\, d_w^{-1/2}\begin{pmatrix} 0 & A - \frac{d_b}{N}(\mathbf{1}) \\ A^* - \frac{d_b}{N}(\mathbf{1}) & 0 \end{pmatrix}.$$

Here, we are abusing notation by now letting A denote the off-diagonal block of an adjacency matrix. This will not be important to the subsequent discussion, but we emphasize this abuse of notation for sake of clarity. Moreover, we recall the notation $\mathbf{1}$ denotes a matrix of the appropriate dimension whose entries are all equal to 1.

We now introduce the convention: for any function $F \in C^4(\mathscr{M}_{\mathbf{e}})$, we define the function $f_F : \Omega \to \mathbb{R}$ by the following equation, letting X_A denote the normalization of A as in (6.3):

$$(6.4) \qquad\qquad f(A) \;=\; f_F(A) \;=\; F(X_A).$$

We now extend this convention to all matrices $A \in \mathscr{M}$. In particular, for any fixed adjacency matrix A (*not* block), suppose $A(t)$ solves the matrix-valued SDE (5.1) with initial data $A(0) = A$. Then, for all times $t \geqslant 0$ we may similarly define $f(A(t)) = F(X_{A(t)})$, where we replace A with $A(t)$ in the normalization given by (6.3).

Remark 6.3. We briefly remark here that the a priori C^4-regularity on the function F is an unnecessary assumption in establishing the above convention. However, our short-time stability results depend on this assumed regularity, so we will assume F is C^4 throughout for sake of simplicity.

Remark 6.4. We now remark that, although $f(A(t)) = F(X_{A(t)})$ holds for all times $t \geqslant 0$, the probabilistic result in Proposition 6.2 holds *only* for the initial data $t = 0$. This is because the proof of Proposition 6.2 relies on the graph structure underlying the initial data matrices and the switchings on said graphs.

In addition to the probabilistic result concerning the discrete generator $\mathcal{Q}$ in Proposition 6.2, we will use the following stronger comparison result to deduce the short-time stability in Theorem 5.2.

Proposition 6.5. *Fix $\varepsilon > 0$ and $r(\varepsilon)$ sufficiently large depending on ε. For any $F \in \mathscr{C}^4(\mathscr{M}_{\mathbf{e}})$, we have*

$$(6.5) \qquad\qquad \mathcal{Q} f_F(A) \;=\; \mathscr{L} F(X) + R,$$

where the error term R satisfies the following bound in expectation:

$$(6.6) \qquad\qquad \mathbb{E}_{\mu_{\mathrm{unif}}} R \;=\; O\left(D^{-1/2} N^{1+\varepsilon}\right) \max_{1 \leqslant i \leqslant 4} \|\partial^i F\|_{r(\varepsilon),0}.$$

I.e., the expectation is taken over the randomness of the uniform probability measure on Ω.

Remark 6.6. We remark here that Proposition 6.5 is an estimate only for time $t = 0$, as otherwise the action of the generator $\mathcal{Q}$ on the function f_F may not be well-defined. Thus, in some sense Proposition 6.5 gives is much weaker result than that of Theorem 5.2. This issue will be remedied in directly studying a weak solution to the SDE (5.1) by separating the solution into two terms: the initial data $X(0)$ and a Gaussian perturbation term.

As with Proposition 6.2, we will now take Proposition 6.5 for granted, and we will use both results together to give a proof of Theorem 5.2. Before we give the details of the proof, we now provide an outline building on Remark 6.6 to clarify the forthcoming decomposition of the generator $\mathscr{L}$ and other technical details in the argument. Recall the solution to a one-dimensional Ornstein-Uhlenbeck SDE with initial data $x(0)$ is given, in law, by

$$(6.7) \qquad\qquad x(t) \;\overset{d}{=}\; e^{-t/2} x(0) + (1 - e^{-t})^{1/2} \omega,$$

where ω is a standard Gaussian random variable independent of $x(0)$. An analogous identity holds for the matrix-analog of the Ornstein-Uhlenbeck SDE, which we discuss in detail during the course of the proof of Theorem 5.2; the scalar random variables in the above solution become random matrices instead. We now decompose the action of the generator $\mathscr{L}$ into the action along each component $X(0)$ and ω, i.e. we rewrite

$$(6.8) \qquad\qquad \mathscr{L} \;=\; \mathscr{L}_X + \mathscr{L}_w$$

where the operator $\mathscr{L}_X$ conditions on the Gaussian term ω and the operator $\mathscr{L}_W$ conditions on the initial data $X(0)$. Here, we use the independence of $x(0)$ and ω.

First, we address the Gaussian term $\mathscr{L}_X$. Because the Gaussian distribution is an invariant measure under the one-dimensional Ornstein-Uhlenbeck SDE, similar to the underlying mechanism of Proposition 6.2, the action of the Gaussian term of $\mathscr{L}_x$ will vanish. Thus, we may consider only the randomness in the initial data $X(0)$, given by the normalized adjacency matrices, while conditioning on the Gaussian perturbation ω. This reduces the stability result into analyzing the randomness of bipartite graphs. Combining now Proposition 6.2 and Proposition 6.5 and integrating over time, we will be able to deduce the desired short-time stability.

Before we provide the details of this heuristic outline, we will introduce the following definitions which will be important for the proof of Theorem 5.2 in conditioning on initial data versus Gaussian perturbations. For any function $F \in C^4(\mathscr{M}_{\mathbf{e}})$, we define

(6.9)
$$F_X(W) \; := \; F_W(X) \; := \; F\left(e^{-t/2}X(0) + (1 - e^{-t})^{1/2}W\right),$$

where W is a random matrix whose distribution follows the standard Gaussian measure, normalized by multiplying the entires of W by $N^{-1/2}$. In words, $F_X(W)$ allows us to condition on the initial data $X(0)$, and $F_W(X)$ allows us to condition on the Gaussian term W. The matrix on the far RHS interpolating between $X(0)$ and W through the time variable t is the unique weak solution to the matrix-valued SDE (5.1) of primary interest. Lastly, we note that here we use $X = X(0)$ as shorthand notation.

6.1. **Proof of Theorem 5.2.** As noted in the preceding outline, the matrix-valued Ornstein-Uhlenbeck SDE (5.1) has the following unique weak solution:

(6.10)
$$X(t) \; \overset{d}{=} \; e^{-t/2}X(0) + (1 - e^{-t})^{1/2}W,$$

where W is a random matrix whose probability distribution is the Gaussian measure on $\mathscr{M}_{\mathbf{e}}$. The derivation of this weak solution, and its uniqueness, may be seen by changing coordinates via SVD as in Lemma 2.3 and solving the matrix-valued equation in standard coordinates. We will take this for granted as it is standard in Ito calculus.

Next, we will decompose the action of $\mathscr{L}$ on a function $F \in C^4(\mathscr{M}_{\mathbf{e}})$ into its action on the two functions F_W and F_X. This is summarized as the following result, whose proof is a straightforward calculation; the reader is welcome to skip the proof of this auxiliary lemma and continue on to the derivation of the short-time stability and the rest of Theorem 5.2.

Lemma 6.7. *For any $F \in C^2(\mathscr{M}_{\mathbf{e}})$, we have the following decomposition of the action of the generator $\mathscr{L}$:*

(6.11)
$$\mathscr{L}F(X(t)) \; = \; \mathscr{L}F_X(W) \; + \; \mathscr{L}F_W(X),$$

where $\mathscr{L}$ is the generator associated to the matrix-valued Ornstein-Uhlenbeck process with initial condition $X = X(0)$.

Proof. Recall the generator $\mathscr{L}$ is given by the following second-order differential operator:

(6.12)
$$\mathscr{L} \; = \; \frac{1}{8MN^2} \sum_{(i,j),(k,\ell)\in V_b} \left(\partial_{ij}^{k\ell}\right)^2 \; - \; \frac{1}{16MN} \sum_{(i,j),(k,\ell)\in V_b} X_{ij}^{k\ell}\partial_{ij}^{k\ell}.$$

We first address the second-order terms; in particular, for any fixed pair (i, j), we have

(6.13)
$$\partial_{ij}\partial_{k\ell}F\left(e^{-t/2}X + (1 - e^{-t})^{1/2}W\right) \; = \; \partial_{ij}\left[e^{-t/2}\partial_{X_{k\ell}}F + (1 - e^{-t})^{1/2}\partial_{W_{k\ell}}F\right]$$

(6.14)
$$= \; e^{-t}\partial_{X_{ij}}\partial_{X_{k\ell}}F \; + \; (1 - e^{-t})\partial_{W_{ij}}\partial_{W_{k\ell}}F,$$

where (6.14) follows from the independence of $X(0)$ and W. However, (6.14) is exactly equal to

(6.15)
$$\partial_{ij}\partial_{k\ell}F_X(W) \; + \; \partial_{ij}\partial_{k\ell}F_W(X).$$

Thus, the second-order terms on the LHS and RHS, respectively, of (6.11) agree. To show the first-order terms agree, this amounts to similar identities given as follows:

(6.16)
$$\partial_{ij}F(X(t)) \; = \; e^{-t/2}\partial_{ij}F_W(X) \; + \; (1 - e^{-t})^{1/2}F_X(W).$$

In particular, averaging over indices $(i, j), (k, \ell) \in V_b$ leads to the vanishing of terms of the form $W_{ij}\partial_{X_{ij}}$ and $X_{ij}\partial_{W_{ij}}$ coming from the first-order terms in $\mathscr{L}$. $\qquad\square$

We now continue with the proof of Theorem 5.2. Using Lemma 6.7 and taking an expectation over the randomness in the process $X(t)$, we deduce

$$\mathbb{E}\,\mathscr{L}F(X(t)) \;=\; \mathbb{E}\,\mathscr{L}F_X(W) \;+\; \mathbb{E}\,\mathscr{L}F_W(X). \tag{6.17}$$

We reemphasize here that the expectation is taken both over the randomness of the initial data $X(0)$ and the Gaussian term W, where we condition on the appropriate component in each expectation term on the RHS. We note this holds, again, due to the independence of $X(0)$ and W. Lastly, we note that the first expectation term on the RHS vanishes. This is because the underlying distribution of the matrix W is invariant under the matrix-valued Ornstein-Uhlenbeck SDE. Thus, we find

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}\,F(X(t)) \;=\; \mathbb{E}\,\mathscr{L}F(X(t)) \;=\; \mathbb{E}\,\mathscr{L}F_W(X) \tag{6.18}$$

$$= \; \mathbb{E}\,\mathcal{Q}f_F(A) \;+\; O\left(D^{-1/2}N^{1+\varepsilon}\right) \max_{1\leqslant i\leqslant 4}\|\partial^i F\|_{r(\varepsilon),0} \tag{6.19}$$

for some positive exponent $r(\varepsilon) > 0$ depending on a fixed, sufficiently small $\varepsilon > 0$. The first equation follows from the Dynkin formula, and the last equation follows from Proposition 6.5. By Proposition 6.2, we have

$$\frac{d}{dt}\,\mathbb{E}\,F(X(t)) \;=\; O\left(D^{-1/2}N^{1+\varepsilon}\right) \max_{1\leqslant i\leqslant 4}\|\partial^i F\|_{r(\varepsilon),0}. \tag{6.20}$$

Integrating over time $s \in [0,t]$, we deduce Theorem 5.2. □

It now remains to prove Propositions 6.2 and 6.5. The remaining discussion on the short-time stability will now be dedicated to proving these two results, which are both discussed in detail for the case of d-regular graphs in [3]. The first result, Proposition 6.2 is also discussed in more detail in Chapter I of this thesis, as well as [4] by which Chapter I is motivated. Moreover, for the sake of local eigenvalue statistics, the reader is invited to skip the details of the proofs of these two results and proceed to the subsequent discussion on using Theorem 5.2 to prove stability of the Stieltjes transform of $X(t)$.

7. Proof of Proposition 6.2

We now rewrite the result in Proposition 6.2 as follows: for any function $f : \Omega \to \mathbb{R}$, we have

$$\mathbb{E}_{\mu_{\mathrm{unif}}}\,\mathcal{Q}f(A) \;=\; 0, \tag{7.1}$$

where μ_{unif} denotes the uniform measure on Ω. Indeed, this is equivalent to the invariance of μ_{unif} by the Dynkin formula.

We now exploit this probabilistic description of the generator $\mathcal{Q}$ discussed in Remark 6.1. The following discussion is carried out in detail in Chapter I of this thesis with regards to double switchings instead of simple switchings, but the details are analogous. For that reason, we will not proceed with the same level of rigor and formalism as in Chapter I.

Our goal now is to build dynamics on the probability space Ω. Suppose $S \subseteq E$ is the subgraph of E with vertices $[S] = \{i,j,m,n\}$. We will call such a subgraph E a *neighborhood*. We define the following indicator function:

$$I(S) \;=\; I(S,E) \;=\; \mathbf{1}\left\{|[S]| = 4,\ S \text{ is } 1\text{-regular}\right\} \tag{7.2}$$

$$= \; \sum_{k_1,\ldots,k_4 \in [S]} I^{k_3 k_4}_{k_1 k_2}(S). \tag{7.3}$$

Remark 7.1. We note the notation including the larger graph E to emphasize that we are fixing a graph $E \in \Omega$ when defining the function $I(S)$. We will suppress this notation for convenience, however.

In words, the indicator function $I(S)$ determines whether or not S, as a graph, is composed of two distinct, non-incidental edges. The representation in the last equation follows from a simple unfolding of the functions I^{mn}_{ij} and I. We also note the function $I(S)$ is random only when S is drawn randomly.

To build the desired dynamics on Ω, we will perform local switchings on graphs. To globalize these dynamics and allow the switchings to resample the entire graph E, we will introduce another indicator function as in Chapter I.

Suppose $S_1, S_2 \subseteq E$ are two subgraphs with vertex sets $[S_\ell] = \{i_\ell, j_\ell, m_\ell, n_\ell\}$ for $\ell = 1, 2$. We define the following indicator function J via

(7.4)
$$J(S_1, S_2) = J(S_1, S_2; E) = \mathbf{1}\{[S_1] \cap [S_2] = \emptyset\}.$$

In words, the function $J(S_1, S_2)$ detects whether or not two neighborhoods S_1, S_2 overlap at the level of vertices. Similar to the function $I(S)$, the function $J(S_1, S_2)$ is random only when S_1, S_2 are drawn randomly. Unlike the function $I(S)$, however, the function $J(S_1, S_2)$ is *not* local and detects information on the intersection of two neighborhoods S_1 and S_2. This will help us glue together simple switchings of randomly drawn neighborhoods S_ℓ.

We now give the details of the switching dynamics. Fix any graph $E \in \Omega$ and its associated adjacency matrix A, and fix a subgraph S with $[S] = \{i, j, m, n\}$. We define the local dynamics

(7.5)
$$T_S(A) = \begin{cases} A - \xi_{ij}^{mn} & I(S) = 1, A_{ij} = 1, A_{mn} = 1, \\ A + \xi_{ij}^{mn} & I(S) = 1, A_{in} = 1, A_{mj} = 1, \\ A & I(S) = 0. \end{cases}$$

We now let $T(S)$ denote the subgraph/neighborhood of $T_S(A)$ with vertex set $[S]$. Moreover, for any $S_1, S_2 \subseteq E$, we define

(7.6)
$$T_{S_1, S_2}(A) = \begin{cases} (T_{S_1} \circ T_{S_2})(A) & J(S_1, S_2) = 1, \\ A & J(S_1, S_2) = 0. \end{cases}$$

We note, conditioning on $J(S_1, S_2) = 1$, we have $T_{S_1}(T_{S_2}(A)) = T_{S_2}(T_{S_1}(A))$, thus we may inductively and unambiguously define the global dynamics $T_{S_1, \ldots, S_k}$ for any $k \geqslant 1$.

In order to exploit the probabilistic description of the generator $\mathcal{Q}$, we need to define the following auxiliary probability space on which the dynamics will actually occur. We define the space $\widetilde{\Omega}$ as follows:

(7.7)
$$\widetilde{\Omega} := \{(E, S) : E \in \Omega, S \subseteq E, |[S]| = 4\}.$$

In words, the probability space $\widetilde{\Omega}$, as a set, is the set of all pairs (E, S) where $E \in \Omega$ and $S \subseteq E$ is a neighborhood. We now impose the following probability measure on $\widetilde{\Omega}$: we will draw the graph $E \subset \Omega$ with respect to the uniform probability measure on Ω, and then choose S from the set of neighborhoods in E uniformly at random. In other words, we draw $E \in \Omega$ uniformly at random and conditioning on $E \in \Omega$, the neighborhood S is drawn uniformly at random. We denote this probability measure by $\widetilde{\mu}_{\text{unif}}$.

We now define the switching dynamics. We define the mapping $T : \widetilde{\Omega} \to \widetilde{\Omega}$ as follows: for any $(E, S) \in \widetilde{\Omega}$, we define

(7.8)
$$T((E, S)) = (T_S(A), T(S))$$

where the RHS is well-defined by use of the functions I and J. Here, we identify a graph E with its adjacency matrix A. Thus, to prove Proposition 6.2, it suffices to prove the following result concerning the Markov dynamics T on $\widetilde{\Omega}$.

Lemma 7.2. *The uniform probability measure on $\widetilde{\Omega}$ is invariant under T. Moreover, the dynamics T are reversible with respect to the uniform measure on $\widetilde{\Omega}$.*

Indeed, Lemma 7.2 follows from the construction of T as an involution on $\widetilde{\Omega}$. In particular, given any two edges $\{ij\}, \{mn\}$ with i, j, m, n mutually distinct vertices, there are only two allowed switchings of the neighborhood $\{ij, mn\}$. This implies

(7.9)
$$\mathbb{E}_{\widetilde{\mu}_{\text{unif}}} \mathcal{Q}f(A, S) = 0.$$

For any function $f : \Omega \to \mathbb{R}$, we now lift to a function

(7.10)
$$\widetilde{f} : \widetilde{\Omega} \longrightarrow \mathbb{R}, \quad \widetilde{f}(A, S) = f(A), A \in \Omega.$$

By taking f to be the indicator function for any subset of graphs, we deduce that the canonical projection $\widetilde{\Omega} \to \Omega$ pushforwards the measure $\widetilde{\mu}_{\text{unif}}$ to the measure μ_{unif}. This completes the proof of Proposition 6.2.

8. Proof of Proposition 6.5

We conclude this discussion by deriving the estimate comparing $\mathcal{Q}$ and $\mathscr{L}$. We recall the desired estimate as follows. For any $F \in C^4(\mathscr{M}_{\mathbf{e}})$ and the corresponding function $f_F : \Omega \to \mathbb{R}$, we have

$$\tag{8.1} \mathcal{Q}f_F(A) \; = \; \mathscr{L}F(X) + R,$$

where R satisfies the following expectation bound:

$$\tag{8.2} \mathbb{E}_{\mu_{\mathrm{unif}}} R \; = \; O\left(D^{-1/2}N^{1+\varepsilon}\right) \max_{1 \leqslant i \leqslant 4} \|\partial^i F\|_{r(\varepsilon),0},$$

where $r(\varepsilon)$ depends on $\varepsilon > 0$ chosen to be small. Before we proceed with the details of the proof, we provide a brief outline. Recall the generator $\mathcal{Q}$ is given by

$$\tag{8.3} \mathcal{Q}f(A) \; = \; \frac{1}{4Nd_w} \sum_{(i,j)\in V_b} \sum_{(m,n)\in V_b} I_{ij}^{mn}(A) \left[f(A - \xi_{ij}^{mn}) - f(A) \right].$$

We now Taylor expand the function $f : \Omega \to \mathbb{R}$ in $A \in \Omega$ along the directions in $\mathscr{X}$, which we recall is defined as

$$\tag{8.4} \mathscr{X} \; := \; \bigcup_{(i,j)\in V_b} \bigcup_{(m,n)\in V_b} \{\xi_{ij}^{mn}\}.$$

In particular, we will Taylor expand to fourth-order, which is where the a priori regularity $F \in C^4(\mathscr{M}_{\mathbf{e}})$ is assumed. Approximately upon Taylor expanding, we expect

$$\tag{8.5} \mathcal{Q}f(A) \; \approx \; \frac{1}{4Nd_w} \sum_{(i,j)\in V_b} \sum_{(m,n)\in V_b} A_{ij} A_{mn} \left[-\partial_{ij}^{mn} f(A) + \frac{1}{2}(\partial_{ij}^{mn})^2 f(A) \right].$$

We will rewrite the entries of the adjacency matrix as follows:

$$\tag{8.6} A_{ij} \; = \; \frac{d_b}{N} + \left(A_{ij} - \frac{d_b}{N} \right),$$

$$\tag{8.7} A_{mn} \; = \; \frac{d_b}{N} + \left(A_{mn} - \frac{d_b}{N} \right).$$

Plugging these expansions into the heuristic approximation (8.5), we obtain the generator $\mathscr{L}$ on the RHS upon the convention $f_F(A) = F(H)$ for any suitably regular function $F \in C^4(\mathscr{M}_{\mathbf{e}})$, in addition to other terms we will show are, in fact, error terms. Controlling these error terms will depend on the nonzero correlations between the matrix entries and a combinatorial analysis of biregular bipartite graphs $E \in \Omega$.

We now provide the details of the above proof outline, which follow closely the details given in the proof of Proposition 3.3 in [3]. To make the Taylor expansion (8.5) more precise, we write

$$\tag{8.8} \mathcal{Q}f(A) \; = \; \frac{1}{4Nd_w} \sum_{(i,j)\in V_b} \sum_{(m,n)\in V_b} A_{ij} A_{mn} \left[-\partial_{ij}^{mn} f(A) + \frac{1}{2}(\partial_{ij}^{mn})^2 f(A) \right] \; + \; N^2(R_1 + R_2),$$

where we make the approximation (8.5) by introducing the following error terms:

$$\tag{8.9} R_1 \; = \; O\left(\frac{N}{d_w}\right) \frac{1}{N^4} \sum_{(i,j)\in V_b} \sum_{(m,n)\in V_b} A_{ij} A_{mn}(1 - I(ij,mn)) \sup_{\theta\in[0,1]} \sup_{X\in\mathscr{X}} |\partial_X f(A + \theta X)|,$$

$$\tag{8.10} R_2 \; = \; O\left(\frac{N}{d_w}\right) \frac{1}{N^4} \sum_{(i,j)\in V_b} \sum_{(m,n)\in V_b} A_{ij} A_{mn} \sup_{\theta_i\in[0,1]} \sup_{X_i\in\mathscr{X}} \left| \partial_{X_1}\partial_{X_2}\partial_{X_3} f\left(A + \sum_{i=1}^{3} \theta_i X_i\right)\right|.$$

We proceed with using the representations of the adjacency matrix entries given in (8.6) and (8.7). In particular, given we expect the fluctuation $A_{ij} - d_b/N$ to be small, we use these representations and study the first-order term in (8.8) by grouping

terms as follows:

$$(8.11) \qquad \sum_{(i,j),(m,n)\in V_b} A_{ij}A_{mn}\partial_{ij}^{mn} f(A) = \sum_{(i,j)\in V_b}\sum_{(m,n)\in V_b} \frac{d_b^2}{N^2}\partial_{ij}^{mn} f(A)$$

$$(8.12) \qquad + \sum_{(i,j)\in V_b}\sum_{(m,n)\in V_b} \partial_{ij}^{mn} f(A)\frac{d_b}{N}\left[\left(A_{ij}-\frac{d_b}{N}\right)+\left(A_{mn}-\frac{d_b}{N}\right)\right]$$

$$(8.13) \qquad + N^2 R_3,$$

where the error term R_3 is given by the following equation:

$$(8.14) \qquad \frac{1}{4Nd_w}R_3 = O\left(\frac{N}{d_w}\right)\frac{1}{N^4}\sum_{(i,j),(m,n)\in V_b} \partial_{ij}^{mn} f(A)\left(A_{ij}-\frac{d_b}{N}\right)\left(A_{mn}-\frac{d_b}{N}\right).$$

We first study the term on the RHS of (8.11). We note that, upon averaging over all $(i,j)\in V_b$ and $(m,n)\in V_b$, we have

$$(8.15) \qquad \sum_{(i,j)\in V_b}\sum_{(m,n)\in V_b} \partial_{ij}^{mn} f(A) = 0,$$

upon unfolding the definition of the operator ∂_{ij}^{mn} in terms of partial derivatives along directions in $\Delta_{ij}\in\mathscr{X}$. Thus, the term on the RHS of (8.11), i.e. the supposed leading-order term of the expansion of the first-order differential term, in fact vanishes. This explains the need to appeal to the lower-order term in (8.12) in the above expansion of the first-order differential term. We now study this term (8.12); upon switching indices $(i,j)\leftrightarrow(m,n)\in V_b$, because we average over indices $(i,j)\in V_b$ and $(m,n)\in V_b$ we see this term is given by

$$(8.16) \qquad 2\sum_{(i,j)\in V_b}\sum_{(m,n)\in V_b} \partial_{ij}^{mn} f(A)\frac{d_b}{N}\left(A_{ij}-\frac{d_b}{N}\right) = 2d_b^{1/2}\sum_{(i,j)\in V_b}\sum_{(m,n)\in V_b} \frac{d_b}{N}X_{ij}\partial_{ij}^{mn} f(A),$$

where the equation follows from the definition of the normalization X of A. This, in turn, is equal to

$$(8.17) \qquad \frac{d_b}{2N}\sum_{(i,j),(m,n)\in V_b} X_{ij}^{mn}\partial_{ij}^{mn} F(H)$$

where the factor of $d_b^{-1/2}$ comes from the convention $f(A) = F(X)$, and the replacement of X_{ij} by X_{ij}^{mn} comes with the normalization factor of $1/4$. Thus, we ultimately see, recalling $M=\alpha N$ and $d_w=\alpha d_b$,

$$(8.18) \qquad \frac{1}{4Nd_w}\sum_{(i,j),(m,n)\in V_b} A_{ij}A_{mn}\partial_{ij}^{mn} f(A) = \frac{1}{16MN}\sum_{(i,j),(m,n)\in V_b} X_{ij}^{mn}\partial_{ij}^{mn} F(H) + N^2 R_3.$$

We now study the second-order terms in (8.8). Again using the representations (8.6) and (8.7), this term is given by

$$(8.19) \qquad \frac{1}{4Nd_w}\sum_{(i,j),(m,n)\in V_b} A_{ij}A_{mn}\times\frac{1}{2}\left(\partial_{ij}^{mn}\right)^2 f(A) = \frac{d_b}{8d_w N^3}\sum_{(i,j),(m,n)\in V_b} \left(\partial_{ij}^{mn}\right)^2 f(A)$$

$$(8.20) \qquad + N^2(R_4 + R_5),$$

where the error terms are given by

$$(8.21) \qquad R_4 = O\left(\frac{N}{d_w}\right)\frac{1}{N^4}\sum_{(i,j),(m,n)\in V_b} \left(\left(\partial_{ij}^{mn}\right)^2 f(A)\frac{d_b}{N}\left(A_{ij}-\frac{d_b}{N}\right)\right),$$

$$(8.22) \qquad R_5 = O\left(\frac{N}{d_w}\right)\frac{1}{N^4}\sum_{(i,j),(m,n)\in V_b} \left(\left(\partial_{ij}^{mn}\right)^2 f(A)\left(A_{ij}-\frac{d_b}{N}\right)\left(A_{mn}-\frac{d_b}{N}\right)\right).$$

Compiling these expansions for the first- and second-order differential terms in (8.8), we finally deduce

$$(8.23) \qquad \mathcal{Q}f(A) = \frac{1}{8MN^2}\sum_{(i,j),(m,n)\in V_b} \left(\partial_{ij}^{mn}\right)^2 F(H) - \frac{1}{16MN}\sum_{(i,j),(m,n)\in V_b} H_{ij}\partial_{ij}^{mn} F(H)$$

$$(8.24) \qquad + N^2(R_1 + R_2 + R_3 + R_4 + R_5).$$

Thus, to finish the proof of Proposition 6.5, it suffices to prove the following estimate on the expectation of the error terms $R_1, \ldots, R_5$.

Proposition 8.1. *In the setting of Proposition 6.5, we have*

$$(8.25) \qquad N^2 \, \mathbb{E}\left[\sum_{i=1}^{5} R_i \right] \; = \; O\left(D^{-1/2} N^{1+\varepsilon} \right) \max_{1 \leqslant i \leqslant 4} \| \partial^i F \|_{r(\varepsilon), 0}.$$

The remainder of this discussion will now be devoted to the proof of Proposition 8.1. The methods and estimates used in this proof are exactly those in the d-regular graph ensemble studied in [3]. However, because the graph structure is important in deriving the estimate in Proposition 8.1, we will include the details for sake of completeness. The proofs are almost identical as those in [3], however, and the reader is invited to skip the details and go to the discussion of using Theorem 5.2 to derive stability of local eigenvalue statistics.

8.1. **Proof of Proposition 8.1.** The necessary ingredients for the proof of Proposition 8.1 are moment bounds on the adjacency matrix entries. The first of these bounds is the following estimate, from which we will derive further moment bounds.

Lemma 8.2. *Fix any $p = O(1) > 0$ and consider any p vertices $(i_1, j_1), \ldots, (i_p, j_p) \in V_b$. Then, for any $x \in [1, M + N]$ and $y \in [1, M + N] \setminus \{(i_1, j_1), \ldots, (i_p, j_p)\}$, we have*

$$(8.26) \qquad \mathbb{E}\left[A_{i_1 j_1} \ldots A_{i_p j_p} A_{xy} \right] \; = \; O\left(\frac{d_b}{N} \right) \mathbb{E}\left[A_{i_1 j_1} \ldots A_{i_p j_p} \right].$$

Proof. First, we may assume $(x, y) \in V_b$ or $(x, y) \in V_w$; otherwise, the estimate follows trivially. We derive the estimate for $(x, y) \in V_b$; the case $(x, y) \in V_w$ follows similarly.

We now define the following for notational convenience:

$$(8.27) \qquad I(p) \; := \; \{(i_1 j_1), \ldots, (i_p, j_p)\}, \quad A(p) \; := \; A_{i_1 j_1} \ldots A_{i_p j_p}.$$

Thus, we have

$$(8.28) \qquad \mathbb{E}\left[A(p) \right] \; = \; \frac{1}{d_b} \mathbb{E}\left[A(p) \sum_y A_{xy} \right] \; = \; \frac{1}{d_b} \mathbb{E}\left[A(p) \sum_{y \in I} A_{xy} \right] \; + \; \frac{1}{d_b} \mathbb{E}\left[A(p) \sum_{y \notin I} A_{xy} \right].$$

We note $|I| = O(1)$ and thus $|\{y \notin I\}| = O(N)$. Moreover, noting $0 < A_{ij} = O(1)$ for all i, j, and also noting the law of A_{ij} under the uniform measure on graphs is invariant under relabeling vertices, we deduce

$$(8.29) \qquad \mathbb{E}\left[A(p) \right] \; = \; O\left(\frac{1}{d_b} \right) \mathbb{E}\left[A(p) A_{xy} \right] \; + \; O\left(\frac{N}{d_b} \right) \mathbb{E}\left[A(p) \right],$$

from which the desired estimate follows clearly. □

Remark 8.3. The estimate given above is somewhat of an independence statement under mild conditions; this will help us compute estimates for moments of adjacency matrix entries. In particular, we deduce the following moment estimates.

Lemma 8.4. *Let a, b be integers defined by*

$$(8.30) \qquad \#\{i, j, m, n\} \; = \; 4 - a, \quad \#\{i, j, m, n, k, \ell, p, q\} \; = \; 8 - b.$$

Then, we have

$$(8.31) \qquad \mathbb{E}\left[A_{ij} A_{mn} \right] \; = \; O\left(\frac{d_b}{N} \right)^{2 - \lfloor a/2 \rfloor},$$

$$(8.32) \qquad \mathbb{E}\left[A_{ij} A_{mn} A_{k\ell} A_{pq} \right] \; = \; O\left(\frac{d_b}{N} \right)^{4 - \lfloor b/2 \rfloor}.$$

54

Proof. We first note $A_{ii} = 0$ for all indices i, so we may assume the bounds $a \leqslant 2, b \leqslant 4$. Moreover we also note all moments of A_{ij} are equal. Then, the desired estimates (8.31) and (8.32) follow from Lemma 8.2. We illustrate for the case $a = 1$; suppose, without loss of generality, that $i = m$. Thus, by Lemma 8.2, we have

$$(8.33) \qquad \mathbb{E}\left[A_{ij} A_{in}\right] = O\left(\frac{d_b}{N}\right) \mathbb{E} A_{ij} = O\left(\frac{d_b}{N}\right)^2 \leqslant O\left(\frac{d_b}{N}\right)^{3/2}.$$

$\square$

We now introduce the following notation for convenience:

$$(8.34) \qquad I_1 := I(ij, mn; A), \quad I_2 := I(k\ell, pq; A),$$

$$(8.35) \qquad J_{12} := J(\{ij, mn\}, \{k\ell, pq\}; A), \quad I_{12} := I_1 I_2 J_{12}.$$

Remark 8.5. We recall that $I_1 = 0$ and $J_{12} = 0$ with low probability. We now derive the following estimates conditioning on exceptional events with regards to the simple switchings dynamics on $\widetilde{\Omega}$.

Lemma 8.6. *Let a, b be defined as in Lemma 8.4. Then, we have*

$$(8.36) \qquad \mathbb{E}\left[(A_{ij} A_{mn} + A_{in} A_{mj})(1 - I_1)\right] = O\left(\frac{d_b}{N}\right)^{3-a},$$

$$(8.37) \qquad \mathbb{E}\left[(A_{ij} A_{mn} + A_{in} A_{mj})(A_{k\ell} A_{pq} + A_{kq} A_{p\ell}(1 - I_{12})\right] = O\left(\frac{d_b}{N}\right)^{5-b}.$$

Proof. We first assume $a, b = 0$; for $a, b \neq 0$, the estimates follow directly from Lemma 8.4 and the following inequalities that hold for $a, b \neq 0$:

$$(8.38) \qquad 3 - a \leqslant 2 - \left\lfloor \frac{a}{2} \right\rfloor, \quad 5 - b \leqslant 4 - \left\lfloor \frac{b}{2} \right\rfloor.$$

If $a = 0$, then the event $A_{ij} A_{mn} I_1 = 0$ corresponds to the event that the subgraph restricted to the vertices $(i, j), (m, n)$ is not 1-regular. In particular, we have the following bound:

$$(8.39) \qquad \mathbb{E}\left[A_{ij} A_{mn}(1 - I_1)\right] \leqslant \mathbb{E}\left[A_{ij} A_{mn}(A_{in} + A_{mj})\right] = O\left(\frac{d_b}{N}\right)^3,$$

where the last bound follows by the assumption $a = 0$ and Lemma 8.2. Because the law of A_{ij} is invariant under relabeling vertices, we deduce (8.36).

Similarly, if $b = 0$, then the event $A_{ij} A_{mn} A_{k\ell} A_{pq}(1 - I_{12}) = 0$ corresponds to the event that the subgraph restricted to the vertices $(i, j), (m, n), (k, \ell), (p, q)$ is not bipartite, or if it is bipartite, the subgraphs are not 1-regular. Formally, we have the estimate (conditioning on $b = 0$)

$$(8.40) \qquad \mathbb{E}\left[A_{ij} A_{mn} A_{k\ell} A_{pq}(1 - I_{12})\right] \leqslant \mathbb{E}\left[A_{ij} A_{mn} A_{k\ell} A_{pq} A_\Sigma\right],$$

where we define

$$(8.41) \qquad A_\Sigma = A_{in} + A_{mj} + A_{kq} + A_{p\ell} + A_{iq} + A_{i\ell} + A_{m\ell} + A_{mq} + A_{kj} + A_{kn} + A_{pj} + A_{pn}.$$

Similarly, we deduce the following bound via Lemma 8.2:

$$(8.42) \qquad \mathbb{E}\left[A_{ij} A_{mn} A_{k\ell} A_{pq}(1 - I_{12})\right] \leqslant O\left(\frac{d_b}{N}\right)^5.$$

Relabeling vertices and by the assumption $b = 0$, as in the proof of (8.36) we deduce (8.37). $\square$

As a direct consequence of the estimates in Lemma 8.6, we deduce the following averaged estimates.

Lemma 8.7. *Suppose α, β are defined by the equations*

$$(8.43) \qquad \#\{i, j\} = 2 - \alpha, \quad \#\{i, j, k, \ell\} = 4 - \beta.$$

Then, we have the following estimates:

$$(8.44) \qquad \frac{1}{N^2} \sum_{(m,n) \in V_b} \mathbb{E}\left[(A_{ij}A_{mn} + A_{in}A_{mj})(1 - I_1)\right] = O\left(\frac{d_b}{N}\right)^{3-\alpha},$$

$$(8.45) \qquad \frac{1}{N^4} \sum_{(m,n) \in V_b} \sum_{(p,q) \in V_b} \mathbb{E}\left[(A_{ij}A_{mn} + A_{in}A_{mj})(A_{k\ell}A_{pq} + A_{kq}A_{p\ell})(1 - I_{12})\right] = O\left(\frac{d_b}{N}\right)^{5-\beta}.$$

Moreover, we have

$$(8.46) \qquad \frac{1}{N^4} \sum_{(i,j),(m,n) \in V_b} \mathbb{E}\left[(A_{ij}A_{mn} + A_{in}A_{mj})(1 - I_1)\right] = O\left(\frac{d_b}{N}\right)^{3},$$

$$(8.47) \qquad \frac{1}{N^8} \sum_{(i,j),(m,n) \in V_b} \sum_{(k,\ell),(p,q) \in V_b} \mathbb{E}\left[(A_{ij}A_{mn} + A_{in}A_{mj})(A_{k\ell}A_{pq} + A_{kq}A_{p\ell})(1 - I_{12})\right] = O\left(\frac{d_b}{N}\right)^{5}..$$

Proof. We first note the estimates (8.46) and (8.47) follow from (8.44) and (8.45), respectively, noting there are $O(N^2)$ pairs (i, j) such that $\alpha = 0$ and $O(N^4)$ pairs of pairs $\{(i, j), (k, \ell)\}$ such that $\beta = 0$.

More generally, for any $\alpha_0 \in [[0, 1]]$ (resp. $\beta_0 \in [[0, 3]]$), we note there are $O(N^{2-s})$ (resp. $O(N^{4-s})$) sets $\{i, j\}$ (resp. $\{i, j, k, \ell\}$) such that $\alpha = \alpha_0$ (resp. $\beta = \beta_0$). Thus, by (8.36) and (8.37), we have

$$(8.48) \qquad \frac{1}{N^2} \sum_{(m,n) \in V_b} \mathbb{E}\left[(A_{ij}A_{mn} + A_{in}A_{mj})(1 - I_1)\right] = O\left(\frac{d_b}{N}\right)^{3-\alpha} + \sum_{\alpha_0=1}^{2} O(N^{-\alpha_0}) \times O\left(\frac{d_b}{N}\right)^{3-\alpha-\alpha_0}$$

$$(8.49) \qquad = O\left(\frac{d_b}{N}\right)^{3-\alpha}.$$

Thus, we derive (8.44). Similarly, we may also deduce (8.45). $\qquad \square$

We now recall the following seminorm for bounded measurable functions f: for $r \geqslant 1$, define

$$(8.50) \qquad \|f\|_r := \left(\mathbb{E}|f(A)|^r\right)^{1/r}.$$

Extending the derivatives, we recall, for $f \in C^k(\mathcal{M}_{\mathbf{e}})$,

$$(8.51) \qquad \|\partial^k f\|_r := \sup_{\theta \in [0,1]^k} \sup_{X \in \mathscr{X}^k} \|\partial_{X_1} \ldots \partial_{X_k} f(A + \theta \cdot X)\|_r;$$

here, we use the notation $X = (X_1, \ldots, X_k) \in \mathscr{X}^k$.

We now derive the following moment estimates coupled with functions of matrices; the bounds in the following lemma follow from the Holder inequality coupled with the the preceding moment estimates. These bounds will be important in bounding the error terms R_1 and R_2.

Lemma 8.8. *Fix $\varepsilon > 0$ and suppose $r = r(\varepsilon) \gg 1$ is sufficiently large depending on the the parameter ε. Define the parameters α, β by the equations*

$$(8.52) \qquad \#\{i, j\} = 2 - \alpha, \quad \#\{i, j, k, \ell\} = 4 - \beta.$$

For any $f \in C^0(\mathscr{M}_{\mathbf{e},0})$ with $\|f\|_r \leqslant 1$, we have the following estimates:

$$(8.53) \qquad \frac{1}{N^4} \sum_{(i,j),(m,n)\in V_b} \mathbb{E}\left[f(A)A_{ij}A_{mn}\overline{I_1}\right] = O\left(\frac{d_b}{N}\right)^{3-\varepsilon},$$

$$(8.54) \qquad \frac{1}{N^4} \sum_{(i,j),(m,n)\in V_b} \mathbb{E}\left[f(A)A_{ij}A_{mn}\right] = O\left(\frac{d_b}{N}\right)^{2-\lfloor \alpha/2 \rfloor-\varepsilon},$$

$$(8.55) \qquad \frac{1}{Nd_w} \sum_{(m,n)\in V_b} \mathbb{E}\left[f(A)\left(A_{ij}A_{mn}-A_{in}A_{mj}\right)\overline{I_1}\right] = O\left(\frac{d_b}{N}\right)^{2-\alpha-\varepsilon},$$

$$(8.56) \qquad \frac{1}{Nd_w} \sum_{(m,n)\in V_b} \mathbb{E}\left[f(A)A_{ij}A_{mn}\right] = O\left(\frac{d_b}{N}\right)^{1-\varepsilon},$$

$$(8.57) \qquad \frac{1}{(Nd_w)^2} \sum_{(m,n),(p,q)\in V_b} \mathbb{E}\left[f(A)\left(A_{ij}A_{mn}-A_{in}A_{mj}\right)\left(A_{k\ell}A_{pq}-A_{kq}A_{p\ell}\right)\overline{I_{12}}\right] = O\left(\frac{d_b}{N}\right)^{3-\beta-\varepsilon},$$

$$(8.58) \qquad \frac{1}{(Nd_w)^2} \sum_{(m,n),(p,q)\in V_b} \mathbb{E}\left[f(A)\left(A_{ij}A_{mn}-A_{in}A_{mj}\right)\left(A_{k\ell}A_{pq}-A_{kq}A_{p\ell}\right)\right] = O\left(\frac{d_b}{N}\right)^{4-\lfloor \beta/2 \rfloor-\varepsilon}.$$

Here, for an indicator random variable χ corresponding to an event Ξ, we denote by $\overline{\chi} = 1 - \chi$ the indicator random variable of the complement of Ξ.

Proof. We prove (8.54); the other bounds follow analogously. By the Holder inequality with respect to the expectation $\mathbb{E}(\cdot)$, we see

$$(8.59) \qquad \mathbb{E}\left[f(A)A_{ij}A_{mn}\right] \leqslant \|f\|_r \left(\mathbb{E}\left[A_{ij}A_{mn}\right]\right)^{1/q} \leqslant \left(\mathbb{E}\,A_{ij}A_{mn}\right)^{1/q},$$

where $q^{-1} = 1 - r^{-1}$. Again by the Holder inequality with respect to the summation over indices $(i,j), (m,n) \in V_b$, we see

$$(8.60) \qquad \frac{1}{N^4} \sum_{(i,j),(m,n)} \left(\mathbb{E}\,A_{ij}A_{mn}\right)^{1/q} \leqslant \frac{1}{N^4}(MN)^{2r}\left(\sum_{(i,j),(m,n)\in V_b} \mathbb{E}\,A_{ij}A_{mn}\right)^{1/q}$$

$$(8.61) \qquad = \left(\sum_{(i,j),(m,n)\in V_b} \frac{1}{N^4}\mathbb{E}\,A_{ij}A_{mn}\right)^{1-r^{-1}}.$$

By Lemma 8.4, this upper bound is also bounded by the following:

$$(8.62) \qquad O\left(\frac{d_b}{N}\right)^{2-\lfloor \alpha/2 \rfloor-C(\alpha)r^{-1}} = O\left(\frac{d_b}{N}\right)^{2-\lfloor \alpha/2 \rfloor-\varepsilon},$$

where the equality holds by choosing $r(\varepsilon) \gg 1$ and the trivial bound $\alpha = O(1)$. This proves (8.54). $\qquad\square$

We now prove the last estimate bounding expectations of function $f : \Omega \to \mathbb{R}$ with coefficients given by the fluctuation terms $A_{ij} - d_b/N$ in the expansions (8.6) and (8.7). These, along with Lemma 8.8, will be important in bounding the error terms $R_3, \ldots, R_5$.

Proposition 8.9. *Fix any $\varepsilon > 0$, and suppose $r = r(\varepsilon) \gg_\varepsilon 1$ is sufficiently large depending on ε. Moreover, define α, β by the equations*

$$(8.63) \qquad \#\{i,j\} = 2 - \alpha, \quad \#\{i,j,m,n\} = 4 - \beta.$$

For any $f \in C^0(\mathscr{M}_{\mathbf{e},0})$, we have

$$(8.64) \qquad \mathbb{E}\left[f(A)\left(A_{ij}-\frac{d_b}{N}\right)\right] = O\left(\frac{d_b}{N}\right)^{1-\varepsilon}\|\partial f\|_r + O\left(\frac{d_b}{N}\right)^{2-\alpha-\varepsilon}\|f\|_r,$$

$$(8.65) \qquad \mathbb{E}\left[f(A)\left(A_{ij}-\frac{d_b}{N}\right)\left(A_{mn}-\frac{d_b}{N}\right)\right] = O\left(\frac{d_b}{N}\right)^{2-\varepsilon}\|\partial^2 f\|_r + O\left(\frac{d_b}{N}\right)^{3-\beta-\varepsilon}\|f\|_r.$$

Proof. We employ the following identity, which follows from the averaging identities $\sum_{m,n} A_{mn} = 2Nd_w$, $\sum_n A_{in} = d_b$ and $\sum_m A_{mj} = d_w$.

$$(8.66) \qquad f(A)\left(A_{ij} - \frac{d_b}{N}\right) = \frac{1}{2Nd_w} f(A) \sum_{(m,n)\in V_b} (A_{ij}A_{mn} - A_{in}A_{mj}).$$

In particular, writing $1 = I_1 + \overline{I_1}$, we have, by Lemma 8.8,

$$(8.67) \qquad \mathbb{E}\left[f(A)\left(A_{ij} - \frac{d_b}{N}\right)\right] = \frac{1}{2Nd_w}\mathbb{E}\left[f(A)\sum_{(m,n)\in V_b}(A_{ij}A_{mn} - A_{in}A_{mj})I_1\right] + O\left(\frac{d_b}{N}\right)^{2-\alpha-\varepsilon}\|f\|_r.$$

Thus, it suffices to bound the first average containing the factor I_1. We first note we may assume $\alpha = 0$ else the first average vanishes. Thus, we see $I_1(A) = I_1(T_S(A))$ with S the subgraph whose edges are given by $\{ij\}$ and $\{mn\}$. With this and the invariance of the uniform measure under $T_S(A)$, we see

$$(8.68) \qquad \mathbb{E}\left[f(A)\left(A_{ij} - \frac{d_b}{N}\right)\right] = \frac{1}{2Nd_w}\sum_{(m,n)\in V_b}\mathbb{E}\left[\left(f(A) - f(A - \xi_{ij}^{mn})\right)A_{ij}A_{mn}I_1\right].$$

With the following Taylor estimate:

$$(8.69) \qquad \left|f(A) - f(A - \xi_{ij}^{mn})\right| \leqslant \sup_{\theta\in[0,1]}\sup_{X\in\mathscr{X}}|\partial_X f(A + \theta X)|,$$

as well as Lemma 8.8, we deduce (8.64). To prove (8.65), we appeal to the identity

$$(8.70) \qquad f(A)\left(A_{ij} - \frac{d_b}{N}\right)\left(A_{k\ell} - \frac{d_b}{N}\right) = \frac{1}{(2Nd_w)^2}f(A)\sum_{(m,n),(p,q)\in V_b}(A_{ij}A_{mn} - A_{in}A_{mj})(A_{k\ell}A_{pq} - A_{kq}A_{p\ell}).$$

Writing $1 = I_{12} + \overline{I}_{12}$, we have, by Lemma 8.8,

$$(8.71) \qquad \mathbb{E}\left[f(A)\left(A_{ij} - \frac{d_b}{N}\right)\left(A_{k\ell} - \frac{d_b}{N}\right)\right] = \frac{1}{(2Nd_w)^2}\sum_{(m,n),(p,q)\in V_b}\mathbb{E}\left[\widetilde{G}_{ij,mn,k\ell,pq}I_{12}\right] + O\left(\frac{d_b}{N}\right)^{3-\beta-\varepsilon}\|f\|_r,$$

where we define

$$(8.72) \qquad \widetilde{G}_{ij,mn,k\ell,pq} = f(A)(A_{ij}A_{mn} - A_{in}A_{mj})(A_{k\ell}A_{pq} - A_{kq}A_{p\ell}).$$

Similarly, we may assume $\beta = 0$, else the first averaging term containing the factor I_{12} vanishes. Again, appealing to the invariance of the uniform measure under $T_S(A)$ and the identities $I_{12}(A) = I_{12}(T_{S_1,S_2}(A))$ with S_1, S_2 the subgraphs containing the vertices $\{ij, mn\}$ and $\{k\ell, pq\}$ respectively, we see

$$(8.73) \qquad \mathbb{E}\left[f(A)\left(A_{ij} - \frac{d_b}{N}\right)\left(A_{k\ell} - \frac{d_b}{N}\right)\right] = \frac{1}{(2Nd_w)^2}\sum_{(m,n),(p,q)\in V_b}\mathbb{E}\left[G_{ij,mn,k\ell,pq}(A)A_{ij}A_{mn}I_{12}\right],$$

where we define

$$(8.74) \qquad G_{ij,mn,k\ell,pq}(A) := f(A) - f(A - \xi_{ij}^{mn}) - f(A - \xi_{k\ell}^{pq}) + f(A - \xi_{k\ell}^{pq} - \xi_{ij}^{mn}).$$

Appealing to the Taylor estimate

$$(8.75) \qquad |G_{ij,mn,k\ell,pq}(A)| \leqslant \sup_{\theta_1,\theta_2\in[0,1]}\sup_{X_1,X_2\in\mathscr{X}}|\partial_{X_1}\partial_{X_2}f(A + \theta X_1 + \theta X_2)|,$$

as well as Lemma 8.8, we also deduce (8.65). $\qquad\square$

We now prove Proposition 8.1. By Lemma 8.8 and Proposition 8.9, we have

$$\text{(8.76)} \qquad \mathbb{E}\, R_1 \;=\; O\left(\frac{d_b}{N}\right)^{2-\varepsilon} \|\partial f\|_r,$$

$$\text{(8.77)} \qquad \mathbb{E}\, R_2 \;=\; O\left(\frac{d_b}{N}\right)^{1-\varepsilon} \|\partial^3 f\|_r,$$

$$\text{(8.78)} \qquad \mathbb{E}\, R_3 \;=\; O\left(\frac{d_b}{N}\right)^{1-\varepsilon} \|\partial^3 f\|_r,$$

$$\text{(8.79)} \qquad \mathbb{E}\, R_4 \;=\; O\left(\frac{d_b}{N}\right)^{2-\varepsilon} \|\partial^2 f\|_r \;+\; O\left(\frac{d_b}{N}\right)^{1-\varepsilon} \|\partial^3 f\|_r,$$

$$\text{(8.80)} \qquad \mathbb{E}\, R_5 \;=\; O\left(\frac{d_b}{N}\right)^{2-\varepsilon} \|\partial^2 f\|_r \;+\; O\left(\frac{d_b}{N}\right)^{1-\varepsilon} \|\partial^4 f\|_r.$$

Thus, by definition of the parameter D and the change of variables $f(A) = F(H)$ with

$$\text{(8.81)} \qquad \partial^k f \;=\; d_b^{-k/2}\partial^k F,$$

we have

$$\text{(8.82)} \qquad \sum_{i=1}^{5} \mathbb{E}\, R_i \;=\; O\left(D^{-1/2}N^{-1+\varepsilon}\right) \sum_{i=1}^{4} \|\partial^i F\|_{r(\varepsilon),0}.$$

This concludes the proof of Proposition 8.1. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\;\;\square$

III. Short Time Stability of Eigenvalues and Eigenvectors

The goal of this discussion will be to use the short-time estimate given in Theorem 5.2 to deduce the stability of eigenvalue statistics after short times. As done in Chapter I of this thesis, in order to obtain understanding of the local eigenvalue statistics of the evolved matrix $X(t)$, our strategy is to first show the stability of its Stieltjes transform. We emphasize here that, only in this discussion do we directly consider the covariance matrix $X_*(t)$ and not the linearization $X(t)$. This allows us to exploit phenomena such as eigenvector delocalization in studying the spectral data of $X_*(t)$ through its Stieltjes transform, whereas delocalization does not hold a priori for the linearization. We refer back to Chapter I for details.

Before we discuss the stability of the Stieltjes transform for short times, we need to introduce two preceding discussions. The first is a general discussion on notions of stochastic inequalities. The second is a discussion of the free convolution of a deterministic matrix with a Gaussian perturbation, which will allow us to interpolate between the initial data $X_*(0)$ and evolved matrices $X_*(t)$ for short times.

9.1. **Aside on Stochastic Inequalities.** We now introduce the following two ubiquitous notions of stochastic inequalities. The second is more important for our purposes, so we emphasize its utility now.

Definition 9.1. Suppose Ω is a probability space, and let $\Xi \subseteq \Omega$ be an event.

- We say Ξ holds with *high probability* if for every $\zeta > 0$, there exists $N_0(\zeta) > 0$ such that for all $N \geqslant N_0(\zeta)$, we have $\mathbb{P}(\Xi^C) \leqslant N^{-\zeta}$.
- Suppose A, B are two nonnegative random variables. We say that B *stochastically dominates* the random variable A if for any $\zeta > 0$, there exists a $N_0(\zeta) > 0$ such that for all $N \geqslant N_0(\zeta)$, we have

$$(9.1) \qquad \mathbb{P}\left(A > N^{1/\zeta}B\right) \leqslant N^{-\zeta}.$$

  In this case, we will adopt the notation $A \prec B$ or $A = O_{\prec}B = O_{\prec}(B)$.

We conclude this short discussion on stochastic inequalities with the following lemma whose proof is a straightforward application of the definition of stochastic domination.

Lemma 9.2. *Suppose I is an indexing set of size $|I| \leqslant N^{O(1)}$, and suppose $\{A_i\}_{i\in I}, \{B_i\}_{i\in I}$ are two families of nonnegative random variables such that for each $i \in I$, we have $A_i \prec B_i$. Then for any nonnegative random variables $\{c_i\}_{i\in I}$, we have*

$$(9.2) \qquad \sum_{i \in I} c_i A_i \ \prec \ \sum_{i \in I} c_i B_i.$$

9.2. **The Free Convolution.** This discussion on the free convolution is most important in Chapter III of this thesis, where the free convolution measure allows us to interpolate between the eigenvalue data of the initial data $X_*(0)$ and the evolved data $X_*(t)$. To define the measure, we first define a Stieltjes transform through the following functional equation:

$$(9.3) \qquad m_{\mathrm{lin},t}(z) \ = \ \frac{1}{2N} \sum_{i=1}^{N} \frac{1}{V_i - z - T m_{\mathrm{lin},t}(z)} + \frac{1}{-V_i - z - T m_{\mathrm{lin},t}(z)}.$$

Here, the terms $\pm V_i$ correspond to the nonzero eigenvalues of the linearization $X(t)$, i.e. those eigenvalues λ such that λ^2 is an eigenvalue of the covariance matrix $X_*(t)$. It is known that the above fixed-point equation admits a unique solution $m_{\mathrm{lin},t}(z)$ that maps the upper-half plane to itself; we refer to [5] for details.

We now define the Stieltjes transform of the free convolution measure, which we denote by $m_{\mathrm{fc},t}$, by the following equation:

$$(9.4) \qquad m_{\mathrm{fc},t}(z) \ = \ \frac{1}{\sqrt{z}} m_{\mathrm{lin},t}(\sqrt{z}),$$

where the square root is chosen with the principal branch of the logarithm on the upper-half plane.

Remark 9.3. The measure ϱ_{fc} here denotes the free convolution measure in the covariance matrix ensemble, whereas in Chapter III, the subscript fc will denote the free convolution measure in the linearization ensemble.

The Radon-Nikodym density ϱ_{fc} of the free convolution measure with respect to Lebesgue measure is defined by taking the Stieltjes inversion of $m_{\mathrm{fc},t}(z)$. We now state a central result of Chapter III that will help us study the short-time stability of the Stieltjes transform $s_t(z)$. The proof resembles that of the local law in Section 7 of [17] and is delegated to Chapter III. First, we will establish the following notation for the partial Stieltjes transform of the linearization $X(t)$:

$$(9.5) \qquad s_{\mathrm{lin},t}(z) \; := \; \frac{1}{2N} \sum_{i=M+1}^{M+N} \left(X(t) - z\right)_{ii}^{-1}.$$

Theorem 9.4. *Fix any $\zeta > 0$. Then for any time $t \in [0, N^{-\zeta}]$ we have with high probability, uniformly over the domain $U_{\varepsilon,\delta}$ defined shortly,*

$$(9.6) \qquad |s_{\mathrm{lin},t}(z) - m_{\mathrm{lin},t}(z)| \; \prec \; \frac{1}{\sqrt{N\eta}}.$$

We now use this result to establish a similar weak local law for the Stieltjes transform $s_t(z)$ of the covariance matrix $X_*(t)$. We first note the following relation:

$$(9.7) \qquad s_t(z) \; = \; \frac{1}{\sqrt{z}} s_{\mathrm{lin},t}(\sqrt{z}),$$

where we again take the principal branch of the logarithm in defining the square root on the upper-half plane. Indeed, this follows from the spectral correspondence between $X_*(t)$ and $X(t)$ discussed in Chapter I.

Because $z \in U_{\varepsilon,\delta}$ implies $z^2 \in U_{\varepsilon',\delta'}$ for some $\varepsilon', \delta' > 0$, and because $|z|^{-1} = O(1)$ for all $z \in U_{\varepsilon,\delta}$, we deduce the following weak local law for the Stieltjes transform of the covariance matrix for any time $t \in [0, N^{-\zeta}]$:

$$(9.8) \qquad \sup_{z \in U_{\varepsilon,\delta}} |s_t(z) - m_{\mathrm{fc},t}(z)| \; \prec \; \frac{1}{\sqrt{N\eta}}.$$

9.3. **Short-Time Estimate for the Stieltjes Transform.** We briefly recall the main estimate from Chapter I. We let $m(z)$ denote the Stieltjes transform of the Marchenko-Pastur law, and we let $s_t(z)$ denote the Stieltjes transform of $X_*(t)$:

$$(9.9) \qquad m(z) \; = \; \frac{1 - \gamma - z + i\sqrt{((1+\sqrt{\gamma})^2 - z)(z - (1-\sqrt{\gamma})^2)}}{2\gamma z},$$

$$(9.10) \qquad s_t(z) \; = \; \frac{1}{N} \operatorname{Tr} G_t(z) \; := \; \frac{1}{N} \operatorname{Tr} \left(X_*(t) - z\right)^{-1}.$$

Here, we have implicitly stipulated the notation for the Green's function $G_t(z)$ of $X_*(t)$. We will take $z \in \mathbb{C}_+$ so that the functions $m(z)$ and $s_t(z)$ define functions from the upper-half plane $\mathbb{C}_+$ to itself.

Theorem 9.5. *Uniformly over the following domain:*

$$(9.11) \qquad U_{\varepsilon,\delta} \; = \; \left\{z = E + i\eta \in \mathbb{C}_+ : \; |E| > \delta > 0, \; \eta \gg N^{-1+\varepsilon}\right\}$$

where $\delta, \varepsilon > 0$ are fixed (small) constants, we have the local Marchenko-Pastur law for the covariance matrix $X_*(0)$ induced by the bipartite graph $X(0)$:

$$(9.12) \qquad |s_0(z) - m(z)| \; \prec \; B + \frac{1}{(N\eta)^{1/4}}$$

for some $B \leqslant N^{-\zeta}$ with a fixed $\zeta > 0$.

We now extend this local law to Stieltjes transforms $s_t(z)$ for times $t = o(1)$. First, we define the subdomain of $U_{\varepsilon,\delta}$:

$$(9.13) \qquad U_{\varepsilon,\delta}^1 \; = \; \left\{z = E + i\eta \in U_{\varepsilon,\delta} : \; |\eta| \leqslant 1\right\}.$$

The following extension of the local Marchenko-Pastur law is a consequence of the weak local law for covariance matrices given in (9.8), the local law (9.12), and the stability of the self-consistent equation defining $m(z)$ discussed in Chapter I.

Proposition 9.6. *Suppose the local law (9.12) holds uniformly over $z = E + i\eta$ with $1 \geqslant \eta \gg N^{-1-\beta}$ for some $\beta > 0$ and for some $B \leqslant N^{-\zeta}$ with a fixed $\zeta > 0$. Then we have the following local law uniformly over such z and uniformly over all times $t \in [0, B]$:*

$$(9.14) \qquad |s_t(z) - m(z)| \prec B + \frac{1}{(N\eta)^{1/4}}.$$

Proof. Using (9.8) and (9.12), we note it suffices to prove, instead, the estimate

$$(9.15) \qquad |m(z) - m_{\mathrm{fc},t}(z)| \prec B + \frac{1}{(N\eta)^{1/4}}.$$

Moreover, by the same self-improving estimate in Chapter I, we may restrict ourselves to $U^1_{\varepsilon,\delta}$. We now unfold the definition of the Stieltjes transform $m_{\mathrm{fc},t}$ to obtain the following fixed-point equation for the Stieltjes transform:

$$(9.16) \qquad m_{\mathrm{fc},t}(z) = \frac{1}{N} \sum_{i=1}^{N} \frac{1 + t m_{\mathrm{fc},t}(z)}{V_i^2 - z(1 + t m_{\mathrm{fc},t}(z))^2}$$

This now allows us to rewrite the free convolution data $m_{\mathrm{fc},t}$ in terms of the Stieltjes transform $s_0(z)$ for the bipartite graphs as follows:

$$(9.17) \qquad m_{\mathrm{fc},t}(z) = (1 + t m_{\mathrm{fc},t}(z)) s_0(z(1 + t m_{\mathrm{fc},t}(z))^2).$$

Using the ideas and methods from Lemma 7.1 and Lemma 7.2 in [17], we first note

$$(9.18) \qquad m_{\mathrm{fc},t}(z) = O(\log N).$$

Now, we study the difference $s_0 - m_{\mathrm{fc},t}$ as in Chapter I on the local law. In particular, using the initial data representation (9.17) and the self-consistent equation for the bipartite graph data s_0 as derived in Chapter I, we may record the following self-consistent equation for the free convolution data $m_{\mathrm{fc},t}$:

$$(9.19) \quad \gamma z \left(1 + t m_{\mathrm{fc},t}(z)\right)^2 \frac{m_{\mathrm{fc},t}(z)^2}{(1 + t m_{\mathrm{fc},t}(z))^2} + \left(\gamma + z(1 + t m_{\mathrm{fc},t}(z))^2 - 1\right) \frac{m_{\mathrm{fc},t}(z)}{1 + t m_{\mathrm{fc},t}(z)} - 1 = (1 + |z|) o_{N \to \infty}(1)$$

We rewrite the first term as $\gamma z m_{\mathrm{fc},t}(z)^2$. We now focus on the second term. Expanding it, we have the second term is given by the following expression:

$$(9.20) \qquad m_{\mathrm{fc},t}(z) \left[(\gamma - 1) \frac{1}{1 + t m_{\mathrm{fc},t}(z)} + z(1 + t m_{\mathrm{fc},t}(z)) \right] = m_{\mathrm{fc},t}(z) \left[(\gamma + z - 1) + \mathscr{E}_1 + \mathscr{E}_2 \right],$$

where the error terms are given by

$$(9.21) \qquad \mathscr{E}_1 = (\gamma - 1) \frac{t m_{\mathrm{fc},t}(z)}{1 + t m_{\mathrm{fc},t}(z)}$$

$$(9.22) \qquad \mathscr{E}_2 = t \left(z m_{\mathrm{fc},t}(z) \right).$$

Here, we use $\log(N)$-bound in (9.18) to deduce the following deterministic bound:

$$(9.23) \qquad \sup_{z \in U_{\varepsilon,\delta}} |t m_{\mathrm{fc},t}| = o(1).$$

This allows us to Taylor expand the term $(1 + t m_{\mathrm{fc},t}(z))^{-1}$ via a geometric series. Ultimately, this expansion of the second term gives us the following self-consistent equation

$$(9.24) \qquad \gamma z m_{\mathrm{fc},t}^2 + (\gamma + z - 1) m_{\mathrm{fc},t} - 1 = (1 + |z|) o_{N \to \infty}(1) - m_{\mathrm{fc},t}(\mathscr{E}_1 + \mathscr{E}_2).$$

Because $t \ll N^{-\zeta}$ for sufficiently small $\zeta > 0$, again by (9.18) we deduce the following bound:

$$(9.25) \qquad \sup_{z \in U_{\varepsilon,\delta}} |m_{\mathrm{fc},t}(z)(\mathscr{E}_1 + \mathscr{E}_2)| = (1 + |z|) o_{N \to \infty}(1).$$

Thus, the stability analysis of the self-consistent equation in Chapter I implies the estimate (9.15). This gives the desired estimate for a fixed time $t \in [0, B]$. To extend to all times $t \in [0, B]$, we appeal to a standard stochastic continuity argument for the Stieltjes transform given in Chapter III. This completes the proof of Proposition 9.6. $\qquad \square$

We now briefly discuss the consequences of Proposition 9.6. We first recall the following notion of a *classical location* defined in Chapter I. For any index $i \in [[1, N]]$, we define the i-th classical location γ_i of the Marchenko-Pastur law as the solution to the following quantile formula:

$$
(9.26) \qquad \frac{i}{N} = \int\limits_{-\infty}^{\gamma_i} \varrho_{\mathrm{MP}}(E) \, \mathrm{d}E,
$$

where ϱ_{MP} denotes the density function of the Marchenko-Pastur law. Similarly, the i-th classical location of the empirical distribution of $X_*(t)$ can be thought of as the eigenvalue $\lambda_i(t)$, assuming the eigenvalues are ordered in strictly increasing order. As discussed in Chapter I, a local law implies a rigidity phenomenon away from the edge of the spectrum. In this chapter, we additionally show a local law implies an entry-wise estimate on the Green's function $G_t(z)$ of the covariance matrix $X_*(t)$. We summarize these results in the following proposition. We again assume eigenvalues are ordered in strictly increasing order. We will also introduce the following notation borrowed from Chapter I of this thesis.

Notation 9.7. *Let $G_t(z)$ denote the Green's function of $X_*(t)$, i.e.*

$$
(9.27) \qquad G_t(z) = (X_*(t) - z)^{-1}
$$

for $z \in \mathbb{C}_+$. We will define the following control parameter for the Green's function:

$$
(9.28) \qquad \Gamma_t(z) = \max_{i,j \in [[1,N]]} \left(|G_{ij}(z)| \vee 1 \right).
$$

Proposition 9.8. *For a fixed $\kappa > 0$ independent of N, fix an index $i \in [[\kappa, (1 - \kappa)N]]$. Then*

$$
(9.29) \qquad \sup_{t \in [0, D^{-1/4}]} |\lambda_i(t) - \gamma(i)| \prec D^{-1/4}
$$

where $\lambda_i(t)$ denotes the i-th eigenvalue of $X_*(t)$ and $\gamma(i)$ denotes the i-th classical location of ϱ_{MP}. Moreover, suppose we have an interval $\mathscr{I} \subseteq [\kappa, (1 - \kappa)(1 + \sqrt{\gamma})^2]$ such that $|\mathscr{I}| \asymp N^{-1+\zeta}$ for some fixed small $\zeta > 0$. Then for any time $t \in [0, D^{-1/4}]$,

$$
(9.30) \qquad \int\limits_{\mathscr{I}} \sum_{i=1}^{N} \delta_{\lambda_i(t)} = |\{i : \lambda_i(t) \in \mathscr{I}\}| = O(N^\zeta) = O_\prec(1).
$$

Lastly, for any $z \in \mathbb{C}_+$ we have

$$
(9.31) \qquad \sup_{t \in [0, D^{-1/4}]} \Gamma_t(z) \prec 1 + \frac{1}{N\eta}.
$$

Proof. The proof of the weak rigidity estimate (9.29) follows from the local law in Proposition 9.6 exactly as in Chapter I of this thesis. The statement of accumulating eigenvalues (9.30) follows from (9.29) and regularity of the Marchenko-Pastur density inside its bulk.

It now remains to prove the second estimate (9.31). Before we do so, we first note a proof using a dyadic decomposition of the scale η is given in Proposition 5.1 in [3] and an eigenvector delocalization result proven in the next section. We provide a different argument using the estimate (9.30). We first appeal to the spectral representation of the Green's function $G_t(z)$ as follows: for any fixed indices i, j, we have

$$
(9.32) \qquad |[G_t(z)]_{ij}| = \left| \sum_{k=1}^{N} \frac{\mathbf{v}_k(i)\mathbf{v}_k(j)}{\lambda_k(t) - z} \right| \leqslant \left(\sup_{k,i,j \in [[1,N]]} |\mathbf{v}_k(i)\mathbf{v}_k(j)| \right) \sum_{k=1}^{N} \frac{1}{|\lambda_k(t) - z|}
$$

$$
(9.33) \qquad \prec \frac{1}{N} \sum_{k=1}^{N} \frac{1}{|\lambda_k(t) - z|}.
$$

Here, we are using the following delocalization result:

$$
(9.34) \qquad \sup_{k,i \in [[1,N]]} |\mathbf{v}_k(i)| \prec \frac{1}{\sqrt{N}}
$$

where the supremum is taken over all ℓ^2-normalized eigenvectors. This result will be proven in a subsequent discussion concerning the stability of eigenvalues; the result for the bipartite graphs at $t = 0$, however, is given in Chapter I. For now, we take the delocalization for granted. Instead of the dyadic decomposition used in [3], we now differentiate in η:

(9.35)
$$\partial_\eta \frac{1}{\sqrt{(\lambda_k(t) - E)^2 + \eta^2}} = -\frac{\eta}{[(\lambda_\alpha - E)^2 + \eta^2]^{3/2}}.$$

Thus by the bound (9.33) we see, writing $z = E + i\eta_0$,

(9.36)
$$|[G_t(z)]_{ij}| \prec 1 + \frac{1}{N} \sum_{k=1}^N \int_{\eta_0}^1 \frac{\eta}{[(\lambda_\alpha - E)^2 + \eta^2]^{3/2}} \, \mathrm{d}\eta.$$

It is now our goal to bound the integral on the RHS of (9.36). To do so, we first define the following sets of eigenvalues $\lambda_k(t)$ of $X_*(t)$: for a fixed, small $\zeta > 0$,

(9.37)
$$\mathscr{I}_1 = \left\{ k : |\lambda_k - E| \leqslant N^{-1+\zeta} \right\}, \quad \mathscr{I}_2 = \left\{ k : |\lambda_k - E| > N^{-1+\zeta} \right\}.$$

We first compute an upper bound on the integral on the RHS of (9.36) by restricting to those eigenvalues λ_k with $k \in \mathscr{I}_2$. In this case, we have the following estimates:

$$\int_{\eta_0}^1 \frac{\eta}{[(\lambda_\alpha - E)^2 + \eta^2]^{3/2}} \mathrm{d}\eta \leqslant \int_{\eta_0}^1 \frac{\eta}{N^{-3+3\zeta} \left[1 + (N^{1-\zeta}\eta)^2\right]^{3/2}} \mathrm{d}\eta$$

$$= N^{1-\zeta} \int_{\eta_0}^1 \frac{\eta}{\left[1 + (N^{1-\zeta}\eta)^2\right]^{3/2}} \mathrm{d}\left(N^{1-\zeta}\eta\right)$$

(9.38)
$$= N^{1-\zeta} \int_{N^{1-\zeta}\eta_0}^{N^{1-\zeta}} \frac{\underline{\eta}}{\left[1 + \underline{\eta}^2\right]^{3/2}} \mathrm{d}\underline{\eta},$$

where we employed a change of variables $\underline{\eta} = N^{1-\zeta}\eta$. We note the integrand in (9.38) is increasing in η; to see this, we compute the derivative as follows:

(9.39)
$$\partial_\eta \left[\frac{\eta}{[1 + \eta^2]^{3/2}} \right] = \frac{(1 + \eta^2)^{3/2} + 3\eta^2 \left[1 + \eta^2\right]^{-5/2}}{(1 + \eta^2)^3} \geqslant 0.$$

Thus, the integral in (9.38) is bounded above by

(9.40)
$$O(N^{2-\zeta}) \frac{N^{1-\zeta}}{[1 + N^{2-2\zeta}]^{3/2}} = O\left(\frac{N^{3-3\zeta}}{N^{3-3\zeta}}\right) = O(1)$$

for $N \gg 1$. This implies

(9.41)
$$N^{-1} \sum_{k \in \mathscr{I}_2} \int_{\eta_0}^1 \frac{\eta}{[(\lambda_\alpha - E)^2 + \eta^2]^{3/2}} \mathrm{d}\eta \prec N^{-1} \sum_{k \in \mathscr{I}_2} 1 \prec 1.$$

We now study the eigenvalues λ_k with the index $k \in \mathscr{I}_1$. To control this term, we simply compute the following estimate:

(9.42)
$$\frac{1}{N} \sum_{k \in \mathscr{I}_1} \int_{\eta_0}^1 \frac{\eta}{[|\lambda_k - E|^2 + \eta^2]^{3/2}} \mathrm{d}\eta \prec \int_{\eta_0}^1 \frac{\eta}{\eta^3} \mathrm{d}\eta \leqslant \frac{1}{N} + \frac{1}{N\eta},$$

where we used (9.30) to estimate the size $|\mathscr{I}_1| \prec 1$. This completes the proof of Proposition 9.8. □

Remark 9.9. We conclude the discussion of the local Marchenko-Pastur law for $X_*(t)$ by remarking on the eigenvector delocalization that was important in the proof of Proposition 9.8. In particular, the delocalization was obtained in Chapter I as a consequence of a local law for the diagonal of the Green's function. In this discussion of time-evolved matrices $X_*(t)$, an estimate on the diagonal entries is unavailable to us. This is also the case in the ensemble of d-regular graphs studied in [3], where other methods become necessary.

Remark 9.10. Instead of partitioning the spectrum of $X_*(t)$ into the sets $\mathscr{I}_1$ and $\mathscr{I}_2$, we may also appeal to a weak level repulsion estimate proved in Chapter III of this thesis.

We now focus our attention towards eigenvector stability under the matrix-valued Ornstein-Uhlenbeck process on $\mathcal{M}_{\mathbf{e}}$. As in the discussion concerning stability of the Stieltjes transform, we recall the result on eigenvector delocalization for bipartite graphs. First, we adopt the following notation.

Notation 10.1. *For the adjacency matrix $X = (H, H^*)$ of a bipartite graph, consider the covariance matrix $X_* = H^* H$. For any given eigenvalue λ_α of X_*, we will denote the corresponding eigenvector by $\mathbf{v}_\alpha$.*

We may now state the eigenvector delocalization result for covariance matrices X_* given by a bipartite graph $E \in \Omega$. For any eigenvector $\mathbf{v}_\alpha$ corresponding to an eigenvalue λ_α of X_*, we have

$$(10.1) \qquad \|\mathbf{v}_\alpha\|_{\ell^\infty} \prec \frac{1}{\sqrt{N}} \|\mathbf{v}_\alpha\|_{\ell^2}.$$

We now extend the delocalization to *all* times. The ideas in this section closely resemble those of Section 4 in [3], relying on eigenvector dynamics studied in [6]. The main goal of this section is to prove the following eigenvector delocalization result for time-evolved covariance matrices. In order to state the result, we introduce notation for the eigenvectors of $X_*(t)$.

Notation 10.2. *We let $\mathbf{v}_{t,\alpha}$ denote the eigenvector of $X_*(t)$ corresponding to the eigenvalue $\lambda_\alpha(t)$.*

Proposition 10.3. *Let $\mathbf{q} \in V$ be a vector such that for some fixed $B > 0$, we have the a priori estimate*

$$(10.2) \qquad \max_\alpha |\mathbf{q} \cdot \mathbf{v}_{0,\alpha}| \leqslant B.$$

Then, for any $t \geqslant 0$, we have

$$(10.3) \qquad \max_\alpha |\mathbf{q} \cdot \mathbf{v}_{t,\alpha}| \prec B.$$

The strategy of proving Proposition 10.3 is as follows. Using the Chebyshev inequality, it suffices to bound the moments of the dot product $\mathbf{q} \cdot \mathbf{v}_{t,\alpha}$. To bound these moments, we define the following pseudo-moment generating function for this dot product for any vector $\eta = (\eta_i)_{i=1}^N$ as follows:

$$(10.4) \qquad f_t(\eta; \mathbf{q}) = \mathbb{E}_{\lambda_t} \left[\prod_{i=1}^N \frac{1}{(2\eta_i - 1)!!} (\mathbf{q} \cdot \mathbf{v}_{t,i})^{2\eta_i} \right],$$

Here, the subscript λ_t in the expectation denotes an expectation conditioning on the eigenvalue process $\{\lambda_\alpha(t)\}_\alpha$. We also use the notation $(n+1)!! = (n+1)(n-1)\ldots 1$ for any odd integer $n \in \mathbb{Z}$, and use the convention $(-1)! = 1$.

Thus, our goal will be to estimate this pseudo-MGF. To study this pseudo-MGF $f_t(\eta; \mathbf{q})$, we study it in context of a particle random walk on the lattice $[[1, N]]$ defined through a generator. The construction and study of this p-particle random walk in the context of eigenvector delocalization is given in more context and detail in the paper [6]. To describe this random walk, we first define its configuration space Ω_p for a fixed $p = O(1)$ independent of N:

$$(10.5) \qquad \Omega = \Omega_p := \left\{ \eta \in \mathbb{Z}_{\geqslant 0}^N : \sum_{i=1}^N \eta_i = p \right\}.$$

In words, the space Ω_p is the space of all configurations of p particles on the lattice $[[1, N]]$. We now define the random walk process; to do so, we first give the following definition.

Notation 10.4. *For any configuration $\eta \in \Omega_p$ and any two fixed sites $i \neq j \in [[1, N]]$, we define $\eta^{ij} \in \Omega_p$ as the configuration obtained by*

$$(10.6) \qquad \eta_k^{ij} = \begin{cases} \eta_k & k \neq i, j \\ \eta_i - 1 & k = i, \ \eta(i) \neq 0 \\ \eta_i & k = i, \ \eta(i) = 0 \\ \eta_j + 1 & k = j, \ \eta(i) \neq 0 \\ \eta_j & k = j, \ \eta(i) = 0 \end{cases}.$$

In words, the configuration η^{ij} is the configuration obtained from the configuration η by moving one particle at site i to site j; if, in the configuration, the site i is void of any particles, then $\eta^{ij} = \eta$, i.e. the configuration η is unchanged. This definition now allows us to define a continuous-time p-particle jump process on $[[1, N]]$ through the following generator:

$$(10.7) \qquad \mathscr{L}_t(f) = \sum_{i \neq j} d_{ij}(t) 2\eta_i(1 + 2\eta_j) \left[f(\eta^{ij}) - f(\eta) \right],$$

where the weight $d_{ij}(t) 2\eta_i(1 + 2\eta_j)$ determines the particle jump rate. In our situation, we are interested in the weights

$$(10.8) \qquad d_{ij}(t) = \frac{\lambda_i(t) + \lambda_j(t)}{N(\lambda_i(t) - \lambda_j(t))^2}.$$

Here, the eigenvalues $\lambda_i(t), \lambda_j(t)$ are the eigenvalues of the time-evolved covariance matrix $X_*(t)$. With these weights, we now give the corresponding Kolmogorov forward equation:

$$(10.9) \qquad \partial_t f_t(\eta) = \mathscr{L}_t f_t(\eta).$$

This ODE is well-posed because Ω is finite. Moreover, its solution for some initial condition f_0 is given by the pseudo-MGF $f_t(\eta; \mathbf{q})$ so long as the spectrum of the initial data $X(0)$ is simple. A proof of this is given by a direct calculation using the following SDE for the eigenvector dynamics:

$$(10.10) \qquad d\mathbf{v}_{t,\alpha} = \frac{1}{\sqrt{N}} \sum_{\beta \neq \alpha} \frac{\sqrt{\lambda_\alpha + \lambda_\beta}}{\lambda_\alpha - \lambda_\beta} \mathbf{v}_{t,\beta} dB_{\alpha\beta}^{(s)} - \frac{1}{2N} \sum_{\beta \neq \alpha} \frac{\lambda_\alpha + \lambda_\beta}{(\lambda_\alpha - \lambda_\beta)^2} \mathbf{v}_{t,\alpha} dt.$$

For details on this SDE, see [6].

We now exploit study the generator $\mathscr{L}_t$ of this p-particle random walk by exploiting its contraction property on any $\ell^r(\Omega)$ space. This follows from an application of the Duhamel formula and semigroup theory.
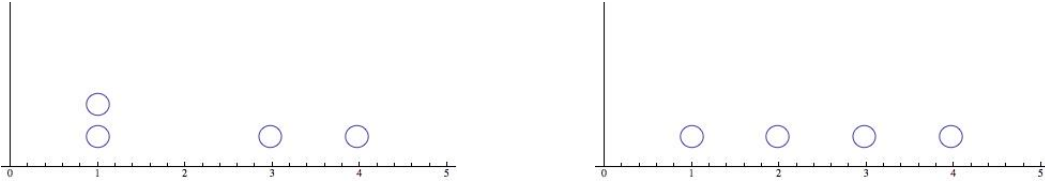
FIGURE 10. We illustrate the particle random walk on the lattice with $N = 5$ and $p = 4$. The LHS corresponds to a configuration vector $\eta = (0, 2, 0, 1, 1, 0)$ and the RHS corresponds to a configuration vector $\eta = (0, 1, 1, 1, 1, 0)$. The RHS is obtained from the LHS by a particle transition from site 1 to site 2 on the lattice.

10.1. **Proof of Proposition 10.3.** Before we begin the proof, we first note the following argument resembles that of delocalization in [3].

Now suppose first that the spectrum of $X(0)$ is simple, in which case the pseudo-MGF $f_t(\eta; \mathbf{q})$ solves the forward equation (10.9). Because $\mathscr{L}_t$ is a contraction on ℓ^∞, we see

$$(10.11) \qquad \|f_t\|_{\ell^\infty(\Omega_p)} \leqslant \|f_0\|_{\ell^\infty(\Omega_p)} \leqslant B^{2p},$$

where for the second inequality, we used delocalization for the initial data and the assumption $\eta \in \Omega_p$ to obtain $\|\eta\|_{\ell^1(\mathbb{Z})} = p$. Thus, for any eigenvector index $i \in [[1, N]]$, we pick the configuration $\eta(i)$ whose components are given by $\eta_j = p\delta_{ij}$, i.e. all p particles are at site i, which gives us

$$(10.12) \qquad f_t(\eta(i)) = \frac{1}{(2p-1)!!} \mathbb{E}\left(\mathbf{q} \cdot \mathbf{v}_{i,t}\right) \leqslant B^{2p}.$$

Using the Chebyshev inequality, we deduce Proposition 10.3 in the case that the spectrum of $X(0)$ is simple. In the case that the spectrum of $X(0)$ is not simple, we note that the eigenvectors $\mathbf{v}_{i,t}$ are uniformly continuous in the eigenvalues $\lambda_i(t)$, and thus a perturbation in the spectrum of $X(0)$ reduces the problem to the calculation for the case of $X(0)$ retaining a simple spectrum. This completes the proof of delocalization.

IV. Short Time Stability for Correlation Functions

11. PRELIMINARY DEFINITIONS AND DISCUSSION

We now aim to prove bulk universality of n-point correlation functions. In the spirit of this thesis we restrict our attention to $\varrho_\infty = \varrho_{\mathrm{MP}}$. We note there that our notation $\varrho_\infty = \varrho_{\mathrm{MP}}$ suppresses the dependence of the ratio $\alpha = M/N$.

We now recall the main theorem, which states that for any energy E in the bulk $\mathscr{I}_{\mathrm{linear},\varepsilon}$, the averaged bulk eigenvalue correlation statistics of the random matrix ensembles $\mathscr{X}(0)$ and $\mathscr{X}(N^{-1-\zeta}D^{1/2})$ agree for a fixed $\zeta > 0$. Here, we define the bulk of ϱ_{MP} to be those energies E a fixed distance from the edges. More precisely, we recall the bulk is defined as

$$(11.1) \qquad \mathscr{I}_{\mathrm{MP},\varepsilon} \;=\; \left[\varepsilon,\; (1-\varepsilon)(1+\sqrt{\gamma})^2\right].$$

Here, $\varepsilon > 0$ is a fixed (small) constant as in the definition of the domains U_ε and $U_{\varepsilon,\delta}$. We also recall the definition $\gamma := \alpha^{-1}$. Similarly, we recall the bulk of the linearization to be

$$(11.2) \qquad \mathscr{I}_{\mathrm{linear},\varepsilon} \;=\; \pm\sqrt{\mathscr{I}_{\mathrm{MP},\varepsilon}} \;=\; \pm\left[\sqrt{\varepsilon},\, \sqrt{1-\varepsilon}(1+\sqrt{\gamma})\right].$$

Here, the sign denotes a reflection of the bulk about the origin.
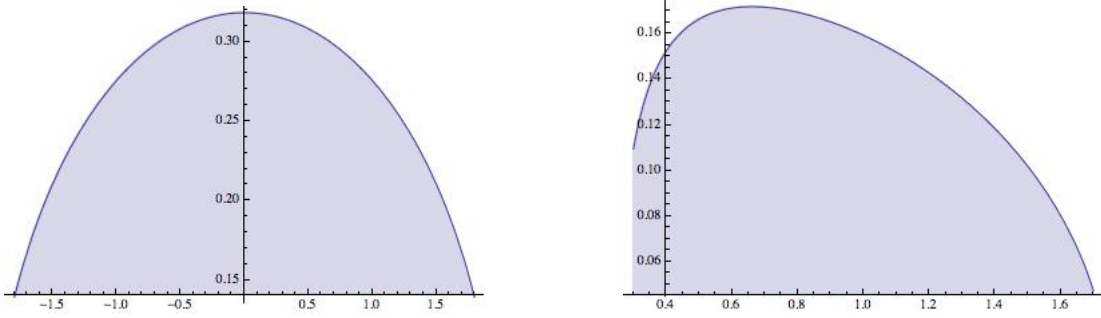
FIGURE 11. The figure on the LHS is the bulk of the semicircle law, which is the linearization of the Marchenko-Pastur law for $\gamma = 1$. Note here the figure includes a neighborhood of the origin, whereas in our ensemble we exclude this neighborhood from the bulk of the linearization. The figure on the RHS is the positive part of the bulk of the linearized Marchenko-Pastur law for $\gamma = 0.75$.

12. PRELIMINARY ESTIMATES

We first record two important estimates that are almost immediate consequences of the short-time stability of the Stieltjes transform and the eigenvector delocalization. These estimates control Green's functions of perturbations of the linearizations.

Lemma 12.1. *Fix a positive integer $n \geqslant 0$. For any spectral parameter $z = E + i\eta \notin \mathbb{R}$ with $|E| > \varepsilon$ for a fixed $\varepsilon > 0$, and a short time $t \leqslant D^{-1/4}$, we have*

$$(12.1) \qquad \sup_{\theta \in [0,1]^n} \sup_{Y \in \mathscr{X}^n} \Gamma\left(z; X(t) + d^{-1/2}\theta \cdot Y\right) \;\prec\; 1 + \frac{1}{N\eta}.$$

Here, the function $\Gamma(z; H)$ denotes the entry-wise maximum of the Green's function $G(z; H) = (H - z)^{-1}$ of the matrix H as used before. Here, we allow either $d = d_b$ or $d = d_w$, without a change to the estimate.

Proof. Before we proceed with any calculations, we reduce the problem to the following three assumptions.

- First, by the following relation which holds for any real matrix H:

$$(12.2) \qquad G(\bar{z}; H) \;=\; \overline{G(z; H)}$$

   where the RHS is the entry-wise complex-conjugate of the Green's function $G(z; H)$, we may assume $\eta > 0$.
- Second, we may assume $|E| \prec 1$ by the $O(1)$ bound on eigenvalues of biregular graphs established in Chapter I and the perturbation inequality for eigenvalues:

$$\lambda(V - W) \leqslant \|V - W\|_\infty,$$

   where V, W are real symmetric matrices. Similarly, we may assume $\eta \prec 1$. Thus, we may assume $|z| \asymp 1$.
- Lastly, by the same bootstrapping method used in the proof of the local law in Chapter I of this thesis, we may work in the regime $\eta \gg N^{-1}$.

Proceeding with the derivation of (12.1), we first compute the following spatial derivatives of $G(z; X_*(t))$:

$$(12.3) \qquad \partial_{X_1 \ldots X_n} G \;=\; (-1)^n \sum_{\sigma \in S_n} G X_{\sigma(1)} G \ldots X_{\sigma(n)} G.$$

Here, S_n denotes the permutation group of n letters. This may be derived via the resolvent identity in Chapter I. We give the calculation for $n = 1$ to illustrate the idea, from which one may differentiate inductively:

$$(12.4) \qquad (X(t) + \theta \cdot Y - z)^{-1} - (X(t) - z)^{-1} \;=\; (X(t) + \theta \cdot Y - z)^{-1} (\theta \cdot Y) (X(t) - z)^{-1}.$$

We note this identity holds because all matrices are in $\mathcal{M}_{\mathbf{e}}$. Dividing by θ and letting $\theta \to 0$ concludes the derivation. With this, we write down the following Taylor estimate:

$$\left| G_{ij}(z; X(t) + d^{-1/2}\theta \cdot Y) \right| \;\leqslant\; |G_{ij}(z; X(t))| + \sup_{\theta \in [0,1]^n} \sup_{X \in \mathscr{X}^n} |\nabla_X G \cdot \theta|$$

$$(12.5) \qquad\qquad\qquad \prec\; |z| \left(1 + \frac{1}{N\eta} \right) + O_n \left(\sup_{X \in \mathscr{X}} \partial_X G \right).$$

To bound the error term above, we give the following estimate:

$$(12.6) \qquad (GYG)_{ij} \;=\; \sum_{k,\ell} G_{ik} Y_{k\ell} G_{\ell j} \;\leqslant\; O\left(\Gamma(z; X_*(t)) \right)^2 \;\prec\; |z|^2 \left(1 + \frac{1}{N\eta} \right)^2 \;\prec\; 1 + \frac{1}{N\eta} + \frac{1}{(N\eta)^2}.$$

We note the implied constant in the big-Oh term may be chosen independent of i, j as any matrix $X \in \mathscr{X}$ has finitely many non-zero terms, which are all bounded. Given the assumption $\eta \succ N^{-1}$, we know

$$(12.7) \qquad \frac{1}{(N\eta)^2} \;\prec\; \frac{1}{N\eta}.$$

Because $|z| \asymp 1$, we also know $|z|^2 = O(|z|)$. With this, we see

$$(12.8) \qquad \left| G_{ij}(z; X(t) + d^{-1/2}\theta \cdot Y) \right| \;\prec\; 1 + \frac{1}{N\eta} + 1 + \frac{1}{N\eta} + \frac{1}{(N\eta)^2} \;\prec\; |z| \left(1 + \frac{1}{N\eta} \right),$$

which completes the derivation of (12.1). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

Before we proceed with the proof of Theorem 4.4, we record the following consequences, which give the same Green's function estimate (12.1) for covariance matrices. Before we do so we introduce the following notation that will only be used in stating the consequences of Lemma 12.1.

Notation 12.2. *Suppose $X(t) = (H(t), H(t)^*)$, so that $X_*(t) = H(t)^* H(t)$. We establish the following notation for the perturbed covariance matrices: for $\theta \in [0,1]^n$ and $Y \in \mathscr{X}^n$, we define*

$$(12.9) \qquad X_*(t; \theta, Y) \;:=\; \left(H(t) + d^{-1/2}\theta \cdot Y \right)^* \left(H(t) + d^{-1/2}\theta \cdot Y \right),$$

$$(12.10) \qquad X^*(t; \theta, Y) \;:=\; \left(H(t) + d^{-1/2}\theta \cdot Y \right) \left(H(t) + d^{-1/2}\theta \cdot Y \right)^*.$$

We record the following consequence for possible future use as it will not be used in this thesis. The proof of the corollary follows immediately from Lemma 12.1, the spectral correspondence between the linearization $X(t)$ and the corresponding covariance matrices, and lastly the spectral representation of the Green's function.

Corollary 12.3. *Assuming the setting of Lemma 12.1, the estimate* (12.1) *holds upon replacing $X(t) + d^{-1/2}\theta \cdot Y$ with $X_*(t; \theta, X)$ and $X^*(t; \theta, X)$.*

13. Proof of Theorem 4.4: Correlation Functions

We are now in a position to prove Theorem 4.4. To do so, we rely on the following result concerned with universality of averaged correlation functions in the bulk. The result is taken as Lemma 5.4 in Section 5 in [3]; for a proof, we refer to Theorem 6.4 in [13].

Theorem 13.1. *Suppose H_1 and H_2 are two random matrix ensembles of equal dimension N, and denote their Green's functions by $G_1(z)$ and $G_2(z)$, respectively.*

Fix a positive integer $n > 0$, and fix a sequence of positive integers $k_1, \ldots, k_n$. Fix a (small) constant $\beta > 0$. For a scale $\eta \in [N^{-1-\beta}, N^{-1}]$, we fix a sequence of complex numbers $z_j^m = E_j^m \pm i\eta$ for $j \in [[1, k_m]]$ and $m \in [[1, n]]$. Here, we stipulate the energies $E_j^m \in \mathscr{I}_{\text{linear},\varepsilon}$ are in the bulk of the Marchenko-Pastur law. Moreover, the signs in the imaginary part of z_j^m are arbitrary.

Let $\varphi \in C_c^\infty(\mathbb{R}^n)$ be a smooth function with compact support such that for any multi-index $\nu = (\nu_1, \ldots, \nu_n)$ with $1 \leqslant |\nu| \leqslant 4$, the following gradients estimates hold for any $\omega > 0$ fixed and sufficiently small:

$$\text{(13.1)} \qquad \sup_{x \in [-N^\omega, N^\omega]} |\partial^\nu \varphi(x)| \leqslant N^{O(\omega)},$$

$$\text{(13.2)} \qquad \sup_{x \in [-N^2, N^2]} |\partial^\nu \varphi(x)| \leqslant N^{O(1)}.$$

Lastly, suppose the following estimate holds:

$$\text{(13.3)} \qquad \left| \mathbb{E}\, \varphi \left(N^{-k_1} \operatorname{Tr} \left(\prod_{j=1}^{k_1} G_1(z_j^1) \right), \ldots, N^{-k_n} \operatorname{Tr} \left(\prod_{j=1}^{k_n} G_1(z_j^n) \right) \right) - \mathbb{E}\, \varphi(G_1 \to G_2) \right| = O\left(N^{-\delta/2 + O(\beta)} \right),$$

where the notation $G_1 \to G_2$ denotes switching all terms depending on G_1 to the corresponding terms depending on G_2, and the implicit constant is allowed to depend on all data in the statement of this theorem except the dimension N. Then, the averaged bulk eigenvalue correlation statistics of H_1 and H_2 agree in the sense of Definition 4.1.

In particular, to prove Theorem 4.4, it will suffice to show that the estimate (13.3) holds in our matrix ensembles $\mathscr{X}(0)$ and $\mathscr{X}(t)$ for short times; this is the content of the following proposition.

Proposition 13.2. *Fix a small constant $\zeta > 0$, and suppose $t \in [0, N^{-1-\zeta} D^{1/2}]$. Assuming the setting of Theorem 13.1 up until the gradient estimates* (13.1) *and* (13.2) *with the following random matrix ensembles of linearized covariance matrices:*

$$\text{(13.4)} \qquad H_1 = \mathscr{X}(0), \quad H_2 = \mathscr{X}(t).$$

Then the estimate (13.3) *holds.*

Before we prove Proposition 13.2 in detail, we give an outline of the argument. Because the Green's functions $G(z; X_*(t))$ and $G(z; X_*(0))$ are functions of the linearizations $X(0)$ and $X(t)$, we may compare the expectation terms in (13.3) with the short-time stability estimate on the generator $\mathscr{L}$ of the matrix-valued diffusion. This is where the restriction on the time interval $t \in [0, N^{-1-\zeta} D^{1/2}]$ comes from.

To implement the short-time stability in Theorem 5.2, we will need to control the derivatives of the expectation term with respect to the switching matrices $\xi_{ij}^{mn} \in \mathscr{X}$. To do so, we will appeal to the derivative formula (12.3) as well as the estimates obtained in Lemma 12.1. The remainder of the proof will then follow directly from straightforward calculations.

Proof. We begin by defining the following function:

$$
(13.5) \qquad F(X(t)) \;=\; \varphi\left(N^{-k_1}\operatorname{Tr}\left(\prod_{j=1}^{k_1} G_1(z_j^1)\right), \ldots, N^{-k_n}\operatorname{Tr}\left(\prod_{j=1}^{k_n} G_1(z_j^n)\right)\right).
$$

This will be treated as a function of $X(t)$ for times $t = 0$ and another time $t \leqslant N^{-1-\zeta}D^{1/2}$. By Theorem 5.2 we have the following short-time stability for the expectation of $F(X(t))$:

$$
(13.6) \qquad \mathbb{E}\,F(X(t)) - \mathbb{E}\,F(X(0)) \;=\; O\left(D^{-1/2}N^{1+\zeta}\max_{1\leqslant i\leqslant 4}\int_0^t \|\partial^i F\|_{r,s}\,\mathrm{d}s\right).
$$

Thus, it will suffice to prove the following gradient estimate on the function F for times $t \leqslant N^{-1-\zeta}D^{1/2}$:

$$
(13.7) \qquad \max_{1\leqslant i\leqslant 4}\|\partial^i F\|_{r,s} \;=\; O\left(N^{\zeta/2 + O(\beta)}\right)
$$

and then choose $\zeta, \beta > 0$ sufficiently small. For simplicity and clarity of presentation, we will focus on the case $n = 1$ and $k_1 = 1$; the argument for general $n > 0$ follows similarly.

We now differentiate the function $F(X)$ from definition. In what follows, the Green's function $G(z)$ will denote the Green's function of a perturbed linearized covariance matrix as in Lemma 12.1. In particular, by the chain-rule we have

$$
(13.8) \qquad \partial_{X_1 \ldots X_i} F\left(X(t) + d^{-1/2}\theta\cdot X\right) \;=\; \partial_{X_1 \ldots X_i}\varphi\left(\frac{1}{N}\operatorname{Tr} G(z)\right)
$$

$$
(13.9) \qquad = \partial_{X_2 \ldots X_i}\left[\partial_{X_1}\varphi\left(\frac{1}{N}\operatorname{Tr} G(z)\right) \times \frac{1}{N}\operatorname{Tr}\left(\partial_{X_1} G\right)\right]
$$

$$
(13.10) \qquad = O_i\left(\max_{1\leqslant k\leqslant i}|\varphi^{(k)}| \times \frac{1}{N}\max_{1\leqslant k\leqslant i}\left|\operatorname{Tr}\left(\partial_{X_{j_1}\ldots X_{j_k}} G(z)\right)\right|\right).
$$

Here, (13.10) follows from a repeated application of the Leibniz rule and chain rule for differentiating along switching matrices $X_j \in \mathscr{X}$. We now use the a priori gradient estimate (13.1) to bound the first term inside the big-Oh term in (13.10) to deduce

$$
(13.11) \qquad \partial_{X_1 \ldots X_i} F\left((X(t) + d^{-1/2}\theta\cdot X\right) \;=\; O_i\left(\frac{N^{O(\omega)}}{N}\max_{1\leqslant k\leqslant i}\left|\operatorname{Tr}\left(\partial_{X_{j_1}\ldots X_{j_k}} G(z; X(t) + d^{-1/2}\theta\cdot X)\right)\right|\right).
$$

We now bound the trace term appealing back to the differentiation identity (12.3) which we rewrite as follows:

$$
(13.12) \qquad \partial_{X_1 \ldots X_i} G \;=\; (-1)^i \sum_{\sigma \in S_i} G X_{\sigma(1)} G \ldots X_{\sigma(i)} G.
$$

Because $i \in [[1,4]]$ and each X_j has at most $O(1)$ non-vanishing entries, we deduce the following straightforward gradient-trace bound, which will help us control the gradient bound (13.10):

$$
(13.13) \qquad \frac{1}{N}\operatorname{Tr}\left(\partial_{X_1 \ldots X_i} G\right) \;=\; O\left(\frac{1}{N}\sum_{j=1}^{N} i!\left[\max_{\sigma \in S_i}\left|GX_{\sigma(1)}\ldots X_{\sigma(i)}G\right|\right]_{jj}\right)
$$

$$
(13.14) \qquad = O_i\left(\Gamma^{O_i(1)}\right).
$$

Combining this estimate with the bounds on perturbed Green's functions (12.1) in Lemma 12.1, we deduce the following bound

$$
(13.15) \qquad \partial_{X_1 \ldots X_i} F\left(X(t) + d^{-1/2}\theta\cdot X\right) \;=\; O_i\left(\left(1 + \frac{1}{N\eta}\right)^{O(1)}\right) \;\prec\; O\left(N^{O(\beta)}\right),
$$

where we used the assumption $\eta \in [N^{-1-\beta}, 1]$ in the last big-Oh estimate. We note this bound holds only with high-probability as the inequality (12.1) in Lemma 12.1 is a stochastic inequality. For the low-probability complement event, we

will go back to the preliminary estimate (13.10) and apply straightforward bounds as follows, instead using (13.2) as opposed to (13.1):

$$(13.16) \qquad \partial_{X_1 \ldots X_i} F\left(X(t) + d^{-1/2}\theta \cdot X\right) \; = \; O_i \left(\max_{1 \leqslant k \leqslant i} |\varphi^{(k)}| \times \frac{1}{N} \max_{1 \leqslant k \leqslant i} \left| \operatorname{Tr} \left(\partial_{X_{j_1} \ldots X_{j_k}} G(z) \right) \right| \right)$$

$$(13.17) \qquad = \; O_i \left(N^{O(1)} \eta^{-C} \right)$$

$$(13.18) \qquad = \; O_i \left(N^{O(1)} \right),$$

where $C = O(1)$ is a positive constant. Thus in taking an expectation in the definition of the $\| - \|_{r,s}$ norm, we have

$$(13.19) \qquad \|\partial^i F\|_{r,t} \; = \; O \left(N^{\zeta + O(\beta)} N^{-\zeta/r + O(1)} \right) \; = \; O \left(N^{\zeta/2 + O(\beta)} \right)$$

upon taking the exponent $r > 0$ suitably small. Here, we drop the subscript i from the big-Oh term because $i \in [[1, 4]]$ is drawn from a set of size $O(1)$. This completes the proof of Proposition 13.2 and thus the proof of Theorem 4.4. $\qquad \square$

Chapter III: Convergence of Local Statistics for Dyson's Brownian Motion for Covariance Matrices

I. Bulk Universality for Deformed Linearized Covariance Matrices

1. A General Underlying Model

As alluded to in the introduction, this chapter focuses on comparing eigenvalue statistics *after* short-time evolution to the equilibrium statistics. As in Chapter II of this thesis, we begin by constructing a Brownian motion on the Hilbert space $\mathscr{M} = \mathbf{M}_{M \times N}(\mathbb{R})$ of $M \times N$ matrices with real entries. We denote this Brownian motion by $\mathbf{B}(t)$.

Remark 1.1. Again, we will assume the following limit of ratios converges:

$$\text{(1.1)} \qquad \lim_{N \to \infty} \frac{M}{N} \geqslant 1,$$

and for all M, N the corresponding ratios exhibit the lower bound $M/N \geqslant 1$. This is a technical assumption and is not necessary, but places the following discussion in the context of Chapter I and Chapter II.

In contrast to Chapter II, we define the following normalized matrix-valued Brownian motion $H(t)$ as opposed to the Ornstein-Uhlenbeck process:

$$\text{(1.2)} \qquad \mathrm{d}H(t) \;=\; \frac{1}{\sqrt{N}} \mathrm{d}\mathbf{B}(t), \quad H(0) \;=\; H.$$

We now define the following matrix space of linearized covariance matrices:

$$\text{(1.3)} \qquad \mathscr{M}_\ell := \left\{ X \;=\; \begin{pmatrix} 0 & H^* \\ H & 0 \end{pmatrix}, \quad H \in \mathscr{M} \right\}.$$

In words, the space $\mathscr{M}_\ell$ is the space of real symmetric matrices whose diagonal blocks trivially vanish and whose off-diagonal blocks are parameterized by the space $\mathscr{M}$. This allows us to define $X(t)$ through the following SDE driven by the Brownian motion $\mathbf{B}(t)$:

$$\text{(1.4)} \qquad \mathrm{d}X(t) \;=\; \begin{pmatrix} 0 & \mathrm{d}H(t) \\ \mathrm{d}H(t)^* & 0 \end{pmatrix}, \quad X(0) \;=\; X \;=\; \begin{pmatrix} 0 & H(0) \\ H(0)^* & 0 \end{pmatrix},$$

where the matrix diffusion $\mathrm{d}H(t)$, with initial condition $H(0)$, is given by (1.2). We note that the matrix $X(t)$ may be parameterized via the following Gaussian perturbation of the initial data:

$$\text{(1.5)} \qquad X(t) \;=\; X(0) \;+\; \sqrt{t} \begin{pmatrix} 0 & W \\ W^* & 0 \end{pmatrix} \;=:\; X \;+\; \sqrt{t} W_\ell,$$

where W is a random matrix whose entries are i.i.d. centered Gaussian random variables with variance $1/N$. In particular, under the notation established in Chapter II of this thesis, we have $W_\ell = (W, W^*)$.

We now address the initial data $H(0)$ and $X(0)$ and define a notion of regularity. First, for notational simplicity and consistency with [17] which studies the DBM statistics for Wigner matrices, we let $V \;=\; H(0)$ denote the initial data. We begin with the following assumption on V.

Assumption 1.2. *The initial data V is $M \times N$ diagonal, i.e. V has the following form:*

(1.6)
$$V = \begin{pmatrix} V_1 & 0 & 0 & \dots \\ 0 & V_2 & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & V_N \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix}.$$

The representation V may be attained from a (time-dependent) singular value decomposition. Indeed, the trace inner product on $\mathbf{M}_{M \times N}(\mathbb{R})$ is invariant under the SVD, and thus so is the Brownian motion $\mathbf{B}(t)$.

Remark 1.3. The main result of this chapter will be deterministic in V, which allows us to take an SVD of the initial data in the overall goal of universality for linearized covariance matrix ensembles.

We now introduce the pair of parameters $(g_N, G_N) = (g, G)$ satisfying the following bounds for some fixed $\varepsilon_1 > 0$:

(1.7)
$$\frac{1}{N} \leqslant g \leqslant N^{-\varepsilon_1}, \quad N^{\varepsilon_1} \ell \leqslant G \leqslant \mathscr{G}_{\varepsilon_1}(N),$$

where we take either $\mathscr{G}_{\varepsilon_1}(N) = N^{-\varepsilon_1}$ or $\mathscr{G}_{\varepsilon_1}(N) \asymp 1$.

For a real number E_0 we define the following interval, or energy window,

(1.8)
$$\mathscr{I}_{E_0, G} = (E_0 \pm G).$$

Lastly, we define the following pseudo-Stieltjes transform for the initial data V as a function of $z = E + i\eta \in \mathbb{C}_+$:

(1.9)
$$m_V(E + i\eta) := \frac{1}{N} \sum_{i=1}^N \frac{1}{V_i - E - i\eta}.$$

We now introduce the following notion of regularity which controls the spectrum and pseudo-Stieltjes transform of the initial data at small scales in the energy window $\mathscr{I}_{E_0, G}$.

Definition 1.4. *The initial potential V is (g, G)-regular at E_0 if the following pseudo-Stieltjes transform bounds hold uniformly over $z = E + i\eta$ with $E \in \mathscr{I}_{E_0, G}$ and $\eta \in [g, 10]$:*

(1.10)
$$c_V \leqslant \operatorname{Im} m_V(E + i\eta) \leqslant C_V$$

for some N-independent constants $C_V, c_V > 0$. Moreover, we require the following spectral bounds:

- For some N-independent constant $B_V > 0$, we have

(1.11)
$$\|V\|_\infty \leqslant N^{B_V},$$

  where the norm $\|\cdot\|_\infty$ is the operator or spectral norm.
- If $M > N$, we require, in addition for some N-independent constant $\varepsilon > 0$, the following lower bound:

(1.12)
$$\inf_{i \in [[1, N]]} |V_i| > \varepsilon > 0.$$

In words, regularity of V amounts to regularity of the spectral data of V through its pseudo-Stieltjes transform in a window around a fixed energy E_0, for suitable scales η. The ℓ^∞-estimate serves to control the growth of V in the large N limit. Lastly, the lower bound on the singular values in the regime $M > N$ serves to regularize otherwise singular dynamics that we will define shortly. We note that for a large class of random matrix ensembles satisfying a local law, the uniform lower bound on the spectrum of V is also satisfied, and is thus a reasonable constraint to impose.

Lastly, we introduce the following a priori delocalization estimate from Chapter II on the eigenvectors of the solution $X(t)$ to the SDE (1.4). For sufficiently regular initial data V, as discussed in Chapter II, the result holds. In the context of this thesis, this includes, with high probability, initial data coming from biregular bipartite graphs.

A Priori Estimate 1.5. *Suppose $X(t)$ solves the SDE (1.4) with $H(0) = V$. Then the eigenvectors of $X(t)$ are delocalized with high probability, i.e. for any growth parameter ξ satisfying $\xi \log \xi \gg \log^2 N$ and time $t \geqslant 0$, we have*

$$(1.13) \qquad \mathbb{P}\left(\sup_{\lambda(t) \in \sigma(X(t))} \sup_{i > M} \left|\mathbf{u}_{\lambda(t)}(i)\right| \geqslant \frac{\xi}{\sqrt{N}}\right) \leqslant e^{-\frac{\xi^2}{2}},$$

where we let $\mathbf{u}_{\lambda(t)}$ denote the ℓ^2-normalized eigenvector of $X(t)$ corresponding to the eigenvalue $\lambda(t)$. Moreover, if $M = N$, then the constraint $i > M$ on the index may be removed.

We briefly remark that although we will use this a priori bound, extending the existing methods in this thesis, coupled with ideas from [18], we may remove this technical assumption. For simplicity, however, we assume it. This will be remarked on with more detail later.

Instead of restricting our initial data to bipartite graphs, as is the subject of this thesis, we work with a much wider class of (g, G)-regular initial data V. By the local laws established in Chapter I of this thesis, the initial data coming from bipartite graphs is (g, G)-regular with high probability in a sense defined in Chapter I.

Lastly, we define the following sets of allowable times. For $\omega, \delta > 0$ to be determined, define

$$(1.14) \qquad \mathscr{T}_{\delta,\omega} := \left\{t : \ell N^{\omega} \leqslant t \leqslant \ell N^{\omega+\delta}\right\},$$

$$(1.15) \qquad \mathscr{T}_{\omega} := \left\{t : \ell N^{\omega} \leqslant t \leqslant G N^{-\omega}\right\}.$$

2. The Linearized Marchenko-Pastur Law, The Semicircle Law, and The Deformed Law

We review the semicircle law, which describes the limiting spectral distribution of Wigner matrices:

$$(2.1) \qquad \varrho_{\mathrm{sc}}(E)\mathrm{d}E = \mathbf{1}_{|E| \leqslant 2} \frac{\sqrt{4 - E^2}}{2\pi} \mathrm{d}E.$$

In particular, the semicircle law describes the eigenvalue statistics of the GOE ensemble, which we recall here.

Definition 2.1. The *Gaussian Orthogonal Ensemble*, or GOE for short, is the following probability density on the space of $N \times N$-real symmetric matrices with respect to flat Lebesgue measure:

$$(2.2) \qquad \frac{1}{Z_{\mathrm{GOE},N}} \exp\left(-\frac{N}{4} \operatorname{Tr} H^2\right) \mathrm{d}H, \quad \mathrm{d}H = \prod_{i \leqslant j} \mathrm{d}H_{ij}.$$

We also review the linearized Marchenko-Pastur law describing spectral statistics of linearized covariance matrices:

$$(2.3) \qquad \varrho(E) = \begin{cases} \frac{\gamma}{(1+\gamma)\pi|E|}\sqrt{(\lambda_+ - E^2)(E^2 - \lambda_-)} & E^2 \in [\lambda_-, \lambda_+] \\ 0 & E^2 \notin [\lambda_-, \lambda_+] \end{cases}.$$

As will soon be made precise, while the macroscopic statistics of linearized covariance matrices follow the linearized Marchenko-Pastur law discussed in Chapter I of this thesis, the *microscopic* statistics follow the classical GOE statistics, which has been historically tied to Wigner matrices and generalized Wigner matrices; for a reference, see [2], [3], [11], [13], [17], and [16].

While the macroscopic eigenvalue statistics of linearized covariance matrices follow the linearized Marchenko-Pastur law given by the density ϱ, the Gaussian perturbation follows a slightly perturbed statistics. We follow the ideas of [17] and [18] and introduce the following interpolation of the spectra of X and $\sqrt{t}W_\ell$ to address the perturbation in eigenvalue statistics.

Definition 2.2. The *free convolution measure* of the Gaussian perturbation $X + \sqrt{t}W_\ell$ is the probability measure corresponding to the Stieltjes transform $m_{\mathrm{fc},t} : \mathbb{C}_+ \to \mathbb{C}_+$, given by the unique solution to the following fixed-point equation:

$$(2.4) \qquad m_{\mathrm{fc},t}(z) = \frac{1}{2N} \sum_{i=1}^{N} \left(\frac{1}{V_i - z - tm_{\mathrm{fc},t}(z)} + \frac{1}{-V_i - z - tm_{\mathrm{fc},t}(z)}\right), \quad z \in \mathbb{C}_+.$$

Here, we take for granted existence and uniqueness of solutions to the above fixed-point equation. We also take for granted the following properties of the solution. For a reference on the free convolution, we cite [5].

- The free convolution measure has a density $\varrho_{\mathrm{fc},t}$ absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$.

- The density $\varrho_{\mathrm{fc},t}$ is compactly supported and analytic on the interior of its support.

We briefly remark that the free convolution density $\varrho_{\mathrm{fc},t}$ and Stieltjes transform $m_{\mathrm{fc},t}$ depend on the matrix parameter N; this will continue to be suppressed from the notation for convenience, but we note that if we take $N \to \infty$, the above qualitative properties of free convolution become rough.

To each law ϱ, ϱ_{fc} and ϱ_{sc}, we introduce the *classical locations* γ_i, $\gamma_{i,t}$, and μ_i with the following quantile formulas:

$$(2.5) \qquad \frac{i}{N} \;=\; \int\limits_{-\infty}^{\gamma_i} \varrho(E)\,\mathrm{d}E \;=\; \int\limits_{-\infty}^{\gamma_{i,t}} \varrho_{\mathrm{fc},t}(E)\,\mathrm{d}E \;=\; \int\limits_{-\infty}^{\mu_i} \varrho_{\mathrm{sc}}(E)\,\mathrm{d}E.$$

In words, the classical locations give, heuristically, a rough expectation for the eigenvalue locations for the corresponding matrix ensemble. These will be necessary in both stating and deriving the universality of eigenvalue statistics.

3. Bulk Universality of Eigenvalue Statistics

We now introduce the two main results in this chapter. The first concerns eigenvalue gap statistics, comparing the gap statistics between adjacent eigenvalues for the perturbed matrix $X(t) = X + \sqrt{t}W_\ell$ to the statistics for the GOE ensemble.

Theorem 3.1. *Suppose $X(t) = X + \sqrt{t}W_\ell$, and assume $X = (V, V^*)$, where V is (g, G)-regular at E_0. Suppose $t \in \mathscr{T}_{\delta,\omega}$ for sufficiently small $\omega, \delta > 0$. Then, for any $O \in C_c^\infty(\mathbb{R}^n)$ and index i such that $\gamma_{i,t} \in \mathscr{I}_{E_0,G/2}$, we have, for any indices $i_1, \ldots, i_n \leqslant N^{c_\omega}$,*

$$\mathbb{E}^{X(t)}\left[O\left(N\varrho_{\mathrm{fc},t}(\gamma_{i,t})(\lambda_i - \lambda_{i+i_1}), \ldots, N\varrho_{\mathrm{fc},t}(\gamma_{i,t})(\lambda_i - \lambda_{i+i_n})\right)\right]$$

$$(3.1) \qquad - \mathbb{E}^{\mathrm{GOE}}\left[O\left(N\varrho_{\mathrm{sc}}(\mu_i)(\lambda_i - \lambda_{i+i_1}), \ldots, N\varrho_{\mathrm{sc}}(\mu_i)(\lambda_i - \lambda_{i+i_n})\right)\right] \leqslant N^{-c_\omega}$$

for some universal constant $c_\omega > 0$.

We note that Theorem 3.1 gives a rough rate of decay explicitly in the gap statistics. We also remark the universality of the constant c_ω follows from the spectral bound on the initial data V and depends on at most the C^1-norm of the test function O. These dependencies will follow from the proof of Theorem 3.1.

From Theorem 3.1, we may obtain the second result of this chapter giving bulk universality of averaged correlation functions for perturbed matrices $X(t)$, which will avoid giving an explicit rate of decay and will instead pass to the limit $N \to \infty$. To state this result, recall the n-point correlation function of a matrix ensemble from Chapter II of this thesis.

Theorem 3.2. *Assume the setting of Theorem 3.1, and for any fixed $c < \omega/2 \wedge \delta/2$, define the parameter $b = N^c/N$. Then for any $O \in C_c^\infty(\mathbb{R}^n)$, we have for any E'' in the interior of the support of ϱ:*

$$\int\limits_{E_0-b}^{E_0+b} \frac{\mathrm{d}E'}{2b} \int\limits_{\mathbb{R}^n} O(\alpha_1, \ldots, \alpha_n)$$

$$(3.2) \qquad \left[\frac{1}{(\varrho_{\mathrm{fc},t}(E_0))^n}\varrho_t^{(n)}\left(E' + \frac{\alpha_1}{N\varrho_{\mathrm{fc},t}(E_0)}, \ldots, E' + \frac{\alpha_n}{N\varrho_{\mathrm{fc},t}(E_0)}\right) - \begin{pmatrix} \varrho_{\mathrm{fc},t} \to \varrho_{\mathrm{sc}} \\ E' \to E'' \\ \varrho_t^{(n)} \to \varrho_{\mathrm{GOE}}^{(n)} \end{pmatrix}\right] \mathrm{d}^n\vec{\alpha} \to_{N\to\infty} 0,$$

where the vector notation denotes the same term but with replacements, e.g. replacing $\varrho_{\mathrm{fc},t}$ with ϱ_{sc}.

The method for deriving the universality of correlation functions from the gap universality may be found in [11] and [17]. We briefly remark here that the core tool is the Helffer-Sjostrand functional calculus, and the fundamental estimate is a *strong local law* comparing local eigenvalue statistics of the GOE and $X(t)$, respectively; we derive the appropriate strong local law later in this chapter.

Before we proceed with an outline of the proof of Theorem 3.1, we briefly remark on the results themselves in Theorem 3.1 and Theorem 3.2. Theorem 3.2 is an extension of previously known results, i.e. those in [11] and [22]. In those two papers, much stronger a priori estimates were required for the initial data V in order to prove the universality of averaged correlation functions. The methods used in [11] to prove such universality involved studying the entropy and time to local ergodicity

motivated by methods for the Wigner ensemble, in contrast to first proving gap universality as in this thesis. Theorem 3.1, however, is a new result in the theory of covariance matrix ensembles, as far as the author is aware. As suggested by the methods and results in this chapter, the reason for this is because previous works, i.e. [1], [21], and [22] studied covariance matrices rather than their linearizations whose eigenvalue dynamics contain only constant coefficient diffusion terms. Moreover, the results in Theorem 3.1 and 3.2 helps provide a bridge between the local eigenvalue statistics of a vast family of covariance matrix ensembles and Wigner matrix ensembles, helping shape the macroscopic picture of the WDGM universality conjecture discussed in the introduction of this thesis.

3.1. **Outline of Theorem 3.1.** We now provide an outline of the proof of Theorem 3.1. We appeal to the ideas from universality of gap statistics for Wigner matrices, first deriving the explicit eigenvalue dynamics known as *Dyson's Brownian Motion*, or DBM for short, along the matrix-valued SDE (1.4) as first performed in [9]. The derivation will resemble that of the Wigner flow, but with modifications to account for the inherently different algebraic structure of linearized covariance matrices. In particular, the SDEs defining the eigenvalue dynamics include logarithmic interaction terms as well. We note this step differs from the approach in Chapter II in studying the DBM.

Next, we appeal to the short-range approximation scheme for DBM in [16] and [17]. This approximation scheme provides a sub-microscopic error after accounting for only the interaction terms from nearby eigenvalues. In [17], this is done with level repulsion, a phenomenon currently unknown for eigenvalues of (linearized) covariance matrices. On the other hand, [16] provides the approximation using only a *rigidity estimate* established in [17], which provides a priori control on the location of eigenvalues up to their average fluctuation with high probability. We will establish the rigidity estimate following the ideas of [17] by deriving a strong local law.

Lastly, accounting for only interaction terms on a microscopic scale allows us to approximate the DBM with a Wigner flow, i.e. the eigenvalue dynamics established in [9]. Following the ideas of [17], this allows us to directly compare the DBM for linearized covariance matrices and the GOE flow. From this, we may deduce Theorem 3.1.

4. Applications to Random Matrix Ensembles

We now aim to apply Theorem 3.2 to the matrix ensemble of bipartite graphs. By the main result in Chapter I of this thesis, we deduce that with high probability the adjacency matrix of a bipartite graph is (g, G)-regular. To be precise, we let Ξ denote an event holding with high probability on which the adjacency matrix is regular. Because Theorem 3.2 is deterministic in the initial data we deduce that on Ξ, the averaged local correlation functions of the bipartite graph, after evolving under the DBM for time $t = N^{-1+\varepsilon}$, coincide with those of the GOE. Meanwhile, by the bound $\mathbb{P}\Xi^C \leqslant N^{-D}$ for all $D > 0$, the contribution from the expectation term in Theorem 3.2 on the complement event Ξ^C is negligible in the limit as $N \to \infty$. We thus deduce that, after time $t = N^{-1+\varepsilon}$, the averaged local correlation functions of bipartite graphs and the GOE coincide. Thus by the results in Chapter II, this completes the proof of universality for averaged bulk local correlation functions for bipartite graphs.

We conclude this introduction by applying Theorem 3.1 and Theorem 3.2 to other covariance matrix ensembles of interest. Applying Theorem 3.2 in the following manners still leaves incomplete answers to important questions, but we provide the following outline to give a broad picture of the universality problem for covariance matrices.

- Concerning the ensemble of sparse covariance matrices as studied in [1], by Theorem 3.2, to prove universality of local correlation functions it suffices to show short-time stability of eigenvalue statistics. To this end, the arguments in [15] applied to linearized covariance matrices should suffice in proving the desired short-time stability.
- Concerning the strongly regular initial data in [22], by Theorem 3.1, to show universality of eigenvalue gaps for the corresponding linearization matrices it now suffices to show short-time stability of gap statistics. This seems to require an optimal level repulsion estimate for linearized covariance matrices, which unfortunately seems to be out of reach for now. We remark on this later in this chapter.

II. A Brownian Motion Model for Linearized Covariance Matrices

5. The Stochastic Differential Equation

We begin by reviewing the following matrix-valued SDEs in the underlying model:

$$
\text{(5.1)} \qquad \mathrm{d}H(t) \;=\; \frac{1}{\sqrt{N}}\mathrm{d}\mathbf{B}(t), \quad H(0) \;=\; H,
$$

$$
\text{(5.2)} \qquad \mathrm{d}X(t) \;=\; \begin{pmatrix} 0 & \mathrm{d}H(t) \\ \mathrm{d}H(t)^* & 0 \end{pmatrix}, \quad X(0) \;=\; X \;=\; \begin{pmatrix} 0 & H(0) \\ H(0)^* & 0 \end{pmatrix}.
$$

As in the breakthrough paper [9], our goal is now to compute the associated eigenvalue dynamics of the SDE (5.2). To derive these dynamics systematically, we introduce the following spectral set which captures the trivial eigenvalues of the time-parameterized family of matrices $X(t)$:

$$
\text{(5.3)} \qquad \zeta_t(X) \;:=\; \left\{ \lambda(t) \in \sigma(X(t)) : \; \lambda^2(t) \notin \sigma(H(t)^* H(t)) \right\}.
$$

We now describe the spectral set $\zeta_t(X)$ in words. If $M \geqslant N$, for any eigenvalue $\lambda^2 \in \sigma(H(t)^* H(t))$, we obtain a pair of eigenvalues $\pm \lambda \in \sigma(X(t))$ in the spectrum of the linearization $X(t)$. The trivial eigenvalues consist of those eigenvalues $\lambda \in \sigma(X(t))$ not obtained from this spectral procedure. This is the set of eigenvalues $\zeta_t(X)$. As discussed in Chapter I, we deduce any $\lambda(t) \in \zeta_t(X)$ is equal to 0, and $\zeta_t(X)$ is empty exactly when $M = N$.

Remark 5.1. We briefly remark here why we define the set $\zeta_t(X)$, as the eigenvalues in this set are trivially 0. In particular, the corresponding eigenvalue dynamics of the matrix-valued SDE (5.2) will involve terms depending on the repulsion between eigenvalues, including repulsion between eigenvalues both contained in and not contained in the set $\zeta_t(X)$. Thus, we will want to interpret the trivial eigenvalues as honest eigenvalues and not fixed real numbers.

We now introduce an important assumption on which regularization of the eigenvalue dynamics heavily depends. The assumption takes form of an a priori estimate on the repulsion of nontrivial eigenvalues from 0 and resembles the notion of regularity for our initial potential V.

Assumption 5.2. *If $M > N$, then for any fixed time scale $T \in \mathcal{T}_{\delta,\omega}$ and $t \in [0, T]$*

$$
\text{(5.4)} \qquad \inf_{\lambda_\alpha \notin \zeta(X)} |\lambda_\alpha| \;>\; \varepsilon \;>\; 0,
$$

where we assume $\varepsilon > 0$ is independent of M, N.

We note that Assumption 5.2 with time scale T follows with high probability if the following hold:

- Assumption 5.2 holds with time scale $T = 0$.
- The time scale T satisfies $T = o_{N \to 1}(0)$.

This is an immediate consequence of the perturbation inequality for any real symmetric matrices A, B:

$$
\text{(5.5)} \qquad \sup_{\lambda \in \sigma(A-B)} |\lambda| \;\leqslant\; \|A - B\|_\infty
$$

where the norm on the RHS is the operator norm. Indeed, we parameterize solutions to (5.2) as follows for time $t \geqslant 0$:

$$
\text{(5.6)} \qquad X(t) \;=\; \begin{pmatrix} 0 & H + \sqrt{t}W \\ H^* + \sqrt{t}W^* & 0 \end{pmatrix},
$$

where $W \in \mathcal{M}$ is sampled from the standard Gaussian measure on $\mathcal{M}$. Letting $A = X$ and $B = X(t)$, coupled with the high-probability estimate on the operator norm $\|W\|_\infty$ from the high-probability estimate on centered Gaussian random variables, this gives sufficiency of the conditions on the initial data and the time scale.

We may now introduce the corresponding eigenvalue dynamics. To this end we define the following simplex:

$$(5.7) \qquad \Delta_{M+N} := \{(x_1, \ldots, x_{M+N}) : x_1 < x_2 < \ldots < x_{M+N}\}.$$

The eigenvalue dynamics will live on the simplex Δ_{M+N}, with the following deterministic equation for $\lambda \in \zeta_t(X)$:

$$(5.8) \qquad d\lambda(t) = 0, \quad \lambda(0) = 0.$$

The SDEs for nontrivial eigenvalues are given as the main result in the following theorem.

Theorem 5.3. *Suppose $X(t)$ solves the SDE in (5.2). Let $\{\lambda_\alpha(t)\}_\alpha$ denote the eigenvalues of $X(t)$ realized as correlated paths on the simplex Δ_{M+N}. Moreover, we assume $N \gg 1$ is sufficiently large so that the following equations are nonsingular.*

(1) *If $M = N$, then the eigenvalues $\{\lambda_\alpha(t)\}_\alpha$ solve the following system of SDEs, known as* Dyson's Brownian Motion:

$$(5.9) \qquad d\lambda_\alpha(t) = \frac{1}{\sqrt{N}} dB_\alpha(t) + \frac{1}{2N} \sum_{\beta \neq \pm\alpha} \frac{1}{\lambda_\alpha - \lambda_\beta} \, dt,$$

where $\{B_\alpha(t)\}_\alpha$ denote independent standard one-dimensional Brownian motions with the constraint $B_\alpha(t) = -B_{-\alpha}(t)$.
Here, $B_{-\alpha}(t)$ denotes the Brownian motion driving the process defining $-\lambda_\alpha(t)$.

(2) *Conditioning on Assumption 5.2, if $M > N$, then the eigenvalues $\{\lambda_\alpha(t)\}_{\alpha \notin \zeta_t(X)}$ solve the following system of SDEs:*

$$(5.10) \qquad d\lambda_\alpha(t) = \frac{1}{\sqrt{N}} dB_\alpha(t) + \frac{1}{2N} \left(\sum_{\beta \notin \zeta_t(X)}^{(\pm\alpha)} \frac{1}{\lambda_\alpha - \lambda_\beta} + \frac{M-N}{\lambda_\alpha} \right) dt$$

$$(5.11) \qquad = \frac{1}{\sqrt{N}} dB_\alpha(t) + \frac{1}{2N} \left(\sum_{\beta \notin \zeta_t(X)}^{(\pm\alpha)} \frac{1}{\lambda_\alpha - \lambda_\beta} + \sum_{\gamma \in \zeta_t(X)} \frac{1}{\lambda_\alpha - \lambda_\gamma} \right) dt$$

where $(\pm\alpha)$ over the summation indicates an omission of terms $\beta = \pm\alpha$. Again, the $\{B_\alpha(t)\}_\alpha$ are independent standard one-dimensional Brownian motions except for the constraint $B_\alpha(t) = -B_{-\alpha}(t)$, where we retain the notation for $B_{-\alpha}(t)$.

Remark 5.4. The regime $M = N$ will be referred to as *square DBM*, and the regime $M > N$ as *rectangular DBM*.
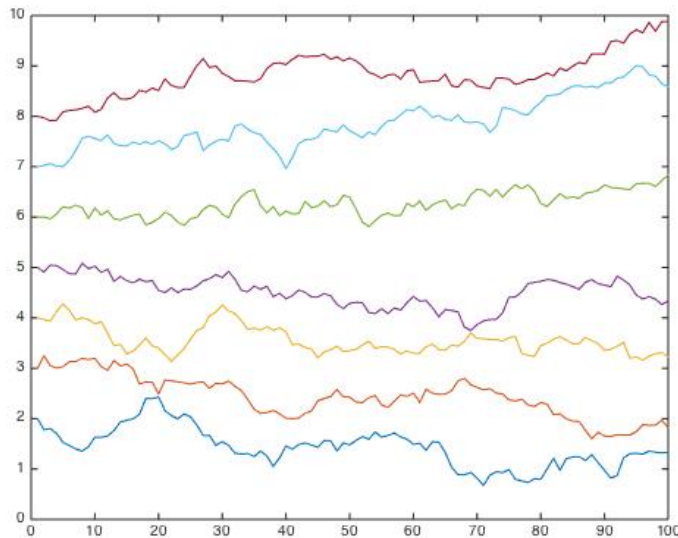

FIGURE 12. Simulation for the DBM equations driving positive eigenvalues for $M = N = 7$.

Before we derive the above systems of SDEs, we emphasize here that the eigenvalue SDEs are autonomous; outside possibly the initial data X_0, the dynamics do not depend on the matrix entries. Moreover, the eigenvalue equations are decoupled from
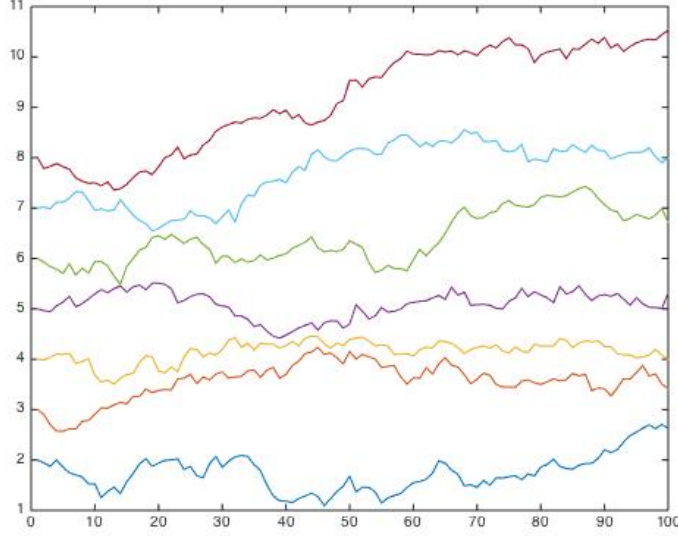
FIGURE 13. Simulation for the DBM equations driving positive eigenvalues for $M > N$ and $\alpha = 3$.

the eigenvectors. This autonomous behavior of the SDEs is an important ingredient in studying the relaxation properties and ergodicity of the spectral dynamics.

We now provide a brief outline for the derivation of the square and rectangular DBMs; the derivation resembles that of the DBM analog for Wigner matrices; see [2], for example. The underlying computational tool will be the Ito formula. This reduces the problem of deriving the DBMs to computing derivatives of eigenvalues along the matrix entries given by the *first-* and *second-order perturbation formulas* from quantum mechanics; we will give a derivation of these identities as well.

Upon applying the Ito formula, the drift terms will follow from straightforward calculation with the eigenvectors and matrix entries of $X(t)$. To study the martingale term, we will compute its generator and identify it as a (scaled) one-dimensional Brownian motion with uniqueness results from semigroup theory.

We note, however, that the drift term includes a singularity at collisions of eigenvalues with distinct indices. In particular, a rigorous derivation of this identity includes an analysis of the collisions of distinct eigenvalues, when viewing eigenvalues with distinct indices as distinct particles on the real line. We will not address this issue in full detail, instead citing a result for covariance matrices and Assumption 5.2 coupled with global eigenvalue control for Gaussian matrices, which will imply eigenvalue intersections occur with low probability.

The Ornstein-Uhlenbeck Variant. We now briefly discuss a slight variant of the matrix dynamics (5.1) and (5.2). In particular, instead of a classical Brownian motion, we consider the following matrix-valued *Ornstein-Uhlenbeck* (OU) equation:

$$(5.12) \qquad \mathrm{d}H(t) \;=\; \frac{1}{\sqrt{N}}\mathrm{d}\mathbf{B}(t) \;-\; \frac{1}{2}H(t)\mathrm{d}t, \quad H(0) \;=\; H.$$

Similarly, we define the Ornstein-Uhlenbeck equation on the space $\mathscr{M}_\ell$ as follows:

$$(5.13) \qquad \mathrm{d}X(t) \;=\; \begin{pmatrix} 0 & \mathrm{d}H(t) \\ \mathrm{d}H(t)^* & 0 \end{pmatrix}, \quad X(0) \;=\; X \;=\; \begin{pmatrix} 0 & H \\ H^* & 0 \end{pmatrix}.$$

In particular, the equation (5.13) is the matrix-valued SDE studied in detail in Chapter II of this thesis. On the simplex Δ_{M+N}, we may derive a similar system of SDEs driving the eigenvalue dynamics corresponding to (5.13). Before we state this result, we briefly remark on the qualitative similarities and differences when adding the drift term in (5.12).

The presence of a drift term in the SDE (5.13), in contrast to (5.2) is motivated by ideas in [11] in which the corresponding eigenvalue dynamics follow Langevin dynamics with respect to a uniformly convex quadratic potential. Because the eigenvalue statistics of covariance matrices should coincide regardless of the distribution of independent matrix entries so long as certain technical assumptions are satisfied such as uniformly sub-exponential decay (see, for example, [1] and [22]), the local eigenvalue statistics should coincide. Thus, the inclusion of the drift term in (5.13) should not change local eigenvalue statistics, but we include this brief discussion for possible future use and for its own sake as well.

As for the corresponding eigenvalue SDEs for the equation (5.13), by the Ito formula one should expect only a change in the drift term. This is, indeed, true, and we summarize the changes in the following theorem.

Theorem 5.5. *Suppose $X(t)$ solves the SDE in (5.13). Let $\{\lambda_\alpha(t)\}_\alpha$ denote the eigenvalues of $X(t)$ realized as correlated paths on the simplex Δ_{M+N}.*

(1) *If $M = N$, then, the eigenvalues $\{\lambda_\alpha(t)\}_\alpha$ solve the following system of SDEs:*

$$(5.14) \qquad d\lambda_\alpha(t) \;=\; \frac{1}{\sqrt{N}}dB_\alpha(t) \;+\; \left(\frac{1}{2N} \sum_{\beta \neq \pm\alpha} \frac{1}{\lambda_\alpha - \lambda_\beta} - \frac{\lambda_\alpha(t)}{2} \right) dt,$$

where $\{B_\alpha(t)\}_\alpha$ denote independent standard one-dimensional Brownian motions with the constraint $B_\alpha(t) = -B_{-\alpha}(t)$. Here, $B_{-\alpha}(t)$ denotes the Brownian motion driving the process defining $-\lambda_\alpha(t)$.

(2) *Conditioning on Assumption 5.2, if $M > N$, then the eigenvalues $\{\lambda_\alpha(t)\}_{\alpha \notin \zeta_t(X)}$ solve the following system of SDEs:*

$$(5.15) \qquad d\lambda_\alpha(t) \;=\; \frac{1}{\sqrt{N}}dB_\alpha(t) \;+\; \left(\frac{1}{2N} \sum_{\beta \notin \zeta_t(X)}^{(\pm\alpha)} \frac{1}{\lambda_\alpha - \lambda_\beta} + \frac{M-N}{2N\lambda_\alpha} - \frac{\lambda_\alpha(t)}{2} \right) dt$$

$$(5.16) \qquad =\; \frac{1}{\sqrt{N}}dB_\alpha(t) \;+\; \left(\frac{1}{2N} \sum_{\beta \notin \zeta_t(X)}^{(\pm\alpha)} \frac{1}{\lambda_\alpha - \lambda_\beta} + \frac{1}{2N} \sum_{\gamma \in \zeta_t(X)} \frac{1}{\lambda_\alpha - \lambda_\gamma} - \frac{\lambda_\alpha(t)}{2} \right) dt$$

where the notation for the superscript $(\pm\alpha)$ over the summation indicates an omission of terms β with $\beta = \pm\alpha$. Again, the $\{B_\alpha(t)\}_\alpha$ denote independent standard one-dimensional Brownian motions except for the constraint $B_\alpha(t) = -B_{-\alpha}(t)$, where we retain the same notation for $B_{-\alpha}(t)$.

6. Perturbation Formulas from Quantum Mechanics

We now direct our focus towards computing the first- and second-order derivative formulas for the eigenvalues of $X(t)$. For our purposes, the matrix $X(t)$ solves a matrix-valued SDE, but this will not contribute towards the derivative formulas. Before we state the result, we introduce some notation.

Notation 6.1. *Suppose $\pm\lambda_\alpha \notin \zeta(X)$ are an eigenvalue pair of X, so in particular $-\lambda_\alpha = \lambda_{-\alpha}$. If $\mathbf{v}_\alpha$ denotes the eigenvector of X corresponding to the eigenvalue λ_α, then we let $\mathbf{v}_{-\alpha}$ denote the eigenvector of X corresponding to the eigenvalue $\lambda_{-\alpha}$.*

As discussed in Chapter I of this thesis, we know

$$(6.1) \qquad \mathbf{v}_{-\alpha}(i) \;=\; \begin{cases} \mathbf{v}_\alpha(i) & i \leqslant M \\ -\mathbf{v}_\alpha(i) & i > M \end{cases}.$$

We now give the eigenvalue formulas below, as well as the first-order derivative formulas for the eigenvectors (corresponding to nontrivial eigenvalues, i.e. those not contained in $\zeta(X)$). The eigenvector formulas are important in deriving the second-order formulas for the eigenvalues.

Lemma 6.2. *Suppose $X \in \mathscr{M}_\ell$, and let $\{\lambda_\alpha\}_\alpha$ denote the eigenvalues of X. The following formulas hold with high probability in context of the DBM. Writing $X = (X_{ij})$, we have, for $\lambda_\alpha(t) \notin \zeta(X)$,*

$$\text{(6.2)} \qquad \partial_{X_{ij}} \lambda_\alpha = \mathbf{v}_\alpha(i)\mathbf{v}_\alpha(j),$$

$$\text{(6.3)} \qquad \partial_{X_{ij}X_{k\ell}} \lambda_\alpha = \sum_{\beta \neq \pm\alpha} \frac{\mathbf{v}_\beta(k)\mathbf{v}_\alpha(\ell) + \mathbf{v}_\beta(\ell)\mathbf{v}_\alpha(k)}{\lambda_\alpha - \lambda_\beta} \left[\mathbf{v}_\beta(i)\mathbf{v}_\alpha(j) + \mathbf{v}_\beta(j)\mathbf{v}_\alpha(i)\right],$$

where the notation for the summation index denotes a sum over eigenvalues $\lambda_\beta \neq \pm\lambda_\alpha$. Moreover, the corresponding eigenvectors $\mathbf{v}_\alpha$ satisfy the following derivative formula:

$$\text{(6.4)} \qquad \partial_{X_{ij}} \mathbf{v}_\alpha = \sum_{\beta \neq \pm\alpha} \frac{\mathbf{v}_\beta^* \left[\partial_{X_{ij}} X\right] \mathbf{v}_\alpha}{\lambda_\alpha - \lambda_\beta} \mathbf{v}_\beta.$$

We reemphasize that the second-order formula for the eigenvalues and the first-order formula for the eigenvectors are coupled. Moreover, in the regime $M > N$, these formula are well-posed in part because of Assumption 5.2, i.e. the terms in the summation corresponding to $\lambda_\beta \in \zeta(X)$ do not introduce singularities into the derivatives.

Proof. We begin with the following eigenvalue-eigenvector equation:

$$\text{(6.5)} \qquad X\mathbf{v}_\alpha = \lambda_\alpha \mathbf{v}_\alpha.$$

We now differentiate both sides with respect to X_{ij} and, using the dot notation for the derivative, we obtain

$$\text{(6.6)} \qquad \dot{X}\mathbf{v}_\alpha + X\dot{\mathbf{v}}_\alpha = \dot{\lambda}_\alpha \mathbf{v}_\alpha + \lambda_\alpha \dot{\mathbf{v}}_\alpha.$$

From here, we may proceed in two different fashions. First, we begin by pairing both sides with $\mathbf{v}_\alpha$ to derive one equation; to derive another set of equations, we pair both sides with $\mathbf{v}_\beta$ for all $\beta \neq \pm\alpha$, obtaining the following system of equations:

$$\text{(6.7)} \qquad \mathbf{v}_\alpha^* \dot{X}\mathbf{v}_\alpha = \dot{\lambda}_\alpha,$$

$$\text{(6.8)} \qquad \mathbf{v}_\beta^* \dot{X}\mathbf{v}_\alpha + \lambda_\beta \mathbf{v}_\beta^* \dot{\mathbf{v}}_\alpha = \lambda_\alpha \mathbf{v}_\beta^* \dot{\mathbf{v}}_\alpha.$$

To derive the first equation, we differentiate the equation $\mathbf{v}_\alpha^* \mathbf{v}_\alpha = 1$ to deduce $\mathbf{v}_\alpha$ is orthogonal to its derivative. We also note the first equation gives us the first-order formula for the eigenvalue λ_α since $\dot{X}$ is the matrix whose entries are all 0, except the (i, j) and (j, i) entries are both equal to 1.

We now look at the second equation (6.8), and deduce, for $\lambda_\alpha \neq \pm\lambda_\beta$,

$$\text{(6.9)} \qquad \mathbf{v}_\beta^* \dot{\mathbf{v}}_\alpha = \frac{\mathbf{v}_\beta^* \dot{X}\mathbf{v}_\alpha}{\lambda_\alpha - \lambda_\beta}.$$

We note the above expression holds for $\lambda_\alpha \neq \pm\lambda_\beta$ as one can show the event $\lambda_\alpha(t) = \lambda_\beta(t)$ does not hold for any $t \geqslant 0$ with high probability in the context of DBM; for a reference, we cite [2]. On the other hand, (6.1) implies

$$\text{(6.10)} \qquad \mathbf{v}_{-\alpha}^* \dot{\mathbf{v}}_\alpha = 0.$$

Thus, we deduce by orthonormality of the eigenvectors,

$$\text{(6.11)} \qquad \dot{\mathbf{v}}_\alpha = \sum_{\beta \neq \pm\alpha} \frac{\mathbf{v}_\beta^* \dot{X}\mathbf{v}_\alpha}{\lambda_\alpha - \lambda_\beta} \mathbf{v}_\beta.$$

This proves the first-order perturbation formula for the eigenvectors $\dot{\mathbf{v}}_\alpha$. We use this formula to derive the second-order formula for the eigenvalues by differentiating the first-order formula as follows:

$$\text{(6.12)} \qquad \ddot{\lambda}_\alpha = \partial_{X_{k\ell}}\left[\mathbf{v}_\alpha(i)\mathbf{v}_\alpha(j)\right] = \dot{\mathbf{v}}_\alpha(i)\mathbf{v}_\alpha(j) + \mathbf{v}_\alpha(i)\dot{\mathbf{v}}_\alpha(j)$$

$$\text{(6.13)} \qquad = \sum_{\beta \neq \pm\alpha} \frac{\mathbf{v}_\beta(k)\mathbf{v}_\alpha(\ell) + \mathbf{v}_\beta(\ell)\mathbf{v}_\alpha(k)}{\lambda_\alpha - \lambda_\beta} \left[\mathbf{v}_\beta(i)\mathbf{v}_\alpha(j) + \mathbf{v}_\beta(j)\mathbf{v}_\alpha(j)\right],$$

which completes the derivation of the second-order formula for the eigenvalues λ_α. $\qquad\square$

7. Derivation of the DBM

We now use the perturbation formulas given in Lemma 6.2 to derive the DBM in Theorem 5.3 and Theorem 5.5. We begin by deriving the square DBM in Theorem 5.3, thus allowing us to forget about the trivial eigenvalues in $\zeta(X)$. From here, we then derive the rectangular DBM in Theorem 5.3, focusing on the contribution from $\zeta(X)$. Lastly, we prove Theorem 5.5, redirecting our focus to the contribution from the additional drift term in the Ornstein-Uhlenbeck equation (5.12).

We first introduce the following result, whose proof may be found in, for example, [2]. This result shows that the event on which the nontrivial eigenvalues of $X(t)$ intersect occurs with probability 0.

Proposition 7.1. *For each nontrivial eigenvalue $\lambda_\alpha(t) \notin \zeta(X(t))$, define the event*

$$(7.1) \qquad E_\alpha(t) := \bigcup_{\beta \neq \alpha: \lambda_\beta \notin \zeta(X)} \{\lambda_\alpha(t) = \lambda_\beta(t)\},$$

where $X(t)$ solves the stochastic matrix dynamics (5.2) and (5.13). Then for any fixed time $T \geqslant 0$,

$$(7.2) \qquad \mathbb{P}\left(\bigcap_{t \leqslant T} \bigcap_{\lambda_\alpha \notin \zeta(X)} E_\alpha(t)\right) = 0.$$

We briefly remark that the proof of Proposition 7.1 follows from standard martingale stopping time techniques in stochastic analysis. We take Proposition 7.1 for granted, however. In particular, the following calculations with the potential singularity hold *rigorously* with high probability given the high probability event on which Assumption 5.2 holds. Lastly, we remark that the proof in [2] is an argument along the DBM for *Wigner* matrices, not covariance matrices. The proof, however, applies with minor modifications to the SDEs (5.2) and (5.13).

Derivation of the Square DBM. By Ito's formula, we have

$$(7.3) \qquad \mathrm{d}\lambda_\alpha = \sum_{X_{ij} \in X} \partial_{X_{ij}} \lambda_\alpha \, \mathrm{d}X_{ij} + \frac{1}{2} \sum_{X_{ij}, X_{k\ell} \in X} \partial_{X_{ij} X_{k\ell}} \lambda_\alpha \, \mathrm{d}\langle X_{ij}, X_{k\ell}\rangle$$

$$(7.4) \qquad = \frac{1}{\sqrt{N}} \sum_{X_{ij} \in X} \partial_{X_{ij}} \lambda_\alpha \, \mathrm{d}B_{ij} + \frac{1}{2}\left(\sum_{X_{ij}, X_{k\ell} \in X} \partial_{X_{ij} X_{k\ell}} \lambda_\alpha \, \mathrm{d}\langle X_{ij}, X_{k\ell}\rangle\right).$$

In the above equation, we suppress from all processes the dependence on time. We first address the martingale term, which we denote by $\mathscr{X}_{\mathrm{mgle}}(t)$. Using the first-order perturbation formula for the eigenvalue λ_α, we have

$$(7.5) \qquad \mathscr{X}_{\mathrm{mgle}}(t) = \frac{1}{\sqrt{N}}\left(\sum_{i=1}^{M}\sum_{j=M+1}^{M+N} + \sum_{i=M+1}^{M+N}\sum_{j=1}^{N}\right) \mathbf{v}_\alpha(i)\mathbf{v}_\alpha(j) \, \mathrm{d}B_{ij}.$$

We now let $\mathscr{X}_{\mathrm{mgle},1}(t)$ denote the first sum in $\mathscr{X}_{\mathrm{mgle}}(t)$ and $\mathscr{X}_{\mathrm{mgle},2}(t)$ denote the second sum. We now identify these two terms as Brownian motions by computing their generators.

Fix a time $t_0 \geqslant 0$, and let $\mathbb{E}_{t_0}$ denote conditional expectation conditioning on events occurring up to time t_0. Because $\mathscr{X}_{\mathrm{mgle},i}(t)$ are centered processes, it remains to compute their conditional quadratic covariations. In particular, we compute

$$(7.6) \qquad \mathbb{E}_{t_0}\langle \mathscr{X}_{\mathrm{mgle},1}(t), \mathscr{X}_{\mathrm{mgle},1}(t)\rangle = \frac{1}{N}\sum_{i=1}^{M}\sum_{j=M+1}^{M+N} |\mathbf{v}_\alpha(i)|^2 |\mathbf{v}_\alpha(j)|^2 \langle \mathrm{d}B_{ij}, \mathrm{d}B_{ij}\rangle$$

$$(7.7) \qquad = \frac{1}{4N}\mathrm{d}t.$$

This follows from the matrix structure of $X(t)$: for indices (i,j) with $|j - i| \geqslant M$, the Brownian motions $\mathrm{d}B_{ij}$ are statistically independent. Similarly,

$$(7.8) \qquad \mathbb{E}_{t_0}\langle \mathscr{X}_{\mathrm{mgle},2}(t), \mathscr{X}_{\mathrm{mgle},2}(t)\rangle = \frac{1}{N}\sum_{i=M+1}^{M+N}\sum_{j=1}^{M} |\mathbf{v}_\alpha(i)|^2 |\mathbf{v}_\alpha(j)|^2 \langle \mathrm{d}B_{ij}, \mathrm{d}B_{ij}\rangle$$

$$(7.9) \qquad = \frac{1}{4N}\mathrm{d}t.$$

Lastly, to compute the covariation, we have from the same calculation

$$(7.10) \qquad \mathbb{E}_{t_0} \langle \mathscr{X}_{\mathrm{mgle},1}(t), \mathscr{X}_{\mathrm{mgle},2}(t) \rangle \;=\; \frac{1}{N} \sum_{i=1}^{M} \sum_{j=1}^{M} |\mathbf{v}_\alpha(i)|^2 |\mathbf{v}_\alpha(j)|^2 \, \mathrm{d}t$$

$$(7.11) \qquad\qquad\qquad\qquad\qquad\qquad =\; \frac{1}{4N} \mathrm{d}t.$$

Thus, we see that $\mathscr{X}_{\mathrm{mgle}}(t)$ is a centered Gaussian process with quadratic variation given by

$$(7.12) \qquad d\langle \mathscr{X}_{\mathrm{mgle}}(t), \mathscr{X}_{\mathrm{mgle}}(t) \rangle \;=\; \left(\frac{1}{4N} + \frac{1}{4N} + \frac{2}{4N} \right) \mathrm{d}t \;=\; \frac{1}{N} \mathrm{d}t.$$

This implies, by the Kolmogorov equations, that $\mathscr{X}_{\mathrm{mgle}}(t)$ is a scaled Brownian motion, i.e.

$$(7.13) \qquad \mathscr{X}_{\mathrm{mgle}}(t) \;=\; \frac{1}{\sqrt{N}} dB_\alpha(t),$$

where $B_\alpha(t)$ is a standard one-dimensional Brownian motion on $\mathbb{R}$. To show the Brownian motions $B_\alpha(t)$ are independent, it suffices to show they are statistically uncorrelated because they are Gaussian random variables. This follows from orthonormality of the eigenvectors $\{\mathbf{v}_\alpha\}_\alpha$ and the representation of $\mathscr{X}_{\mathrm{mgle}}(t)$ in terms of the eigenvectors given in Lemma 6.2. The relation $B_\alpha(t) = -B_{-\alpha}(t)$ follows immediately from $\dot{\lambda}_\alpha = -\dot{\lambda}_{-\alpha}$.

We now address the drift term; because the entries $X_{ij}, X_{k\ell}$ are driven by i.i.d. Brownian motions $B_{ij}(t), B_{k\ell}(t)$, we have

$$(7.14) \qquad \mathrm{d}\langle X_{ij}(t), X_{k\ell}(t) \rangle \;=\; \frac{1}{N} \left(\delta_{ik}\delta_{\ell j} + \delta_{i\ell}\delta_{jk} \right) \mathrm{d}t$$

given the symmetric structure of the matrix X_t. Thus, the drift term may be written as

$$(7.15) \qquad \mathscr{X}_{\mathrm{drift}}(t) \;=\; \frac{1}{2} \left(\frac{1}{N} \sum_{X_{ij}=X_{k\ell}\in X} \left(\partial_{X_{ij}X_{k\ell}} \lambda_\alpha \right) \right) \mathrm{d}t$$

$$(7.16) \qquad\qquad =\; \frac{1}{2N} \left(\sum_{i=1}^{M} \sum_{j=M+1}^{M+N} \sum_{\beta\neq\pm\alpha} \frac{\mathbf{v}_\beta(i)\mathbf{v}_\alpha(j) + \mathbf{v}_\beta(j)\mathbf{v}_\alpha(i)}{\lambda_\alpha - \lambda_\beta} \left[\mathbf{v}_\beta(i)\mathbf{v}_\alpha(j) + \mathbf{v}_\beta(j)\mathbf{v}_\alpha(i) \right] \right) \mathrm{d}t.$$

For each term in the sum, we expand the eigenvector terms and group according to indices α, β as follows:

$$\left[\mathbf{v}_\beta(i)\mathbf{v}_\alpha(j) + \mathbf{v}_\beta(j)\mathbf{v}_\alpha(i) \right] \left[\mathbf{v}_\beta(i)\mathbf{v}_\alpha(j) + \mathbf{v}_\beta(j)\mathbf{v}_\alpha(i) \right] \;=\; |\mathbf{v}_\beta(i)|^2 |\mathbf{v}_\alpha(j)|^2 + |\mathbf{v}_\beta(j)|^2 \mathbf{v}_\alpha(i)|^2$$

$$(7.17) \qquad\qquad\qquad\qquad\qquad\qquad\qquad + 2\mathbf{v}_\beta(i)\mathbf{v}_\alpha(i)\mathbf{v}_\beta(j)\mathbf{v}_\alpha(j).$$

By the spectral correspondence between $X(t)$ and the covariance matrices $X_*(t) = H(t)^* H(t)$ and $X_{*,+}(t) = H(t)H(t)^*$ discussed in Chapter I of this thesis, we see

$$(7.18) \qquad \sum_{i=1}^{M} \sum_{j=M+1}^{M+N} |\mathbf{v}_\beta(i)|^2 |\mathbf{v}_\alpha(j)|^2 \;=\; \sum_{i=1}^{M} \sum_{j=M+1}^{M+N} |\mathbf{v}_\beta(j)|^2 |\mathbf{v}_\alpha(i)|^2 \;=\; \frac{1}{2},$$

$$(7.19) \qquad \sum_{i=1}^{M} \mathbf{v}_\beta(i)\mathbf{v}_\alpha(i) \sum_{j=M+1}^{M+N} \mathbf{v}_\beta(j)\mathbf{v}_\alpha(j) \;=\; 0,$$

where the second equation follows from orthonormality of different eigenvectors. Thus, we finally deduce

$$(7.20) \qquad \mathscr{X}_{\mathrm{drift}}(t) \;=\; \frac{1}{2N} \sum_{\beta\neq\pm\alpha} \frac{1}{\lambda_\alpha - \lambda_\beta} \, \mathrm{d}t,$$

which completes the formal derivation of the square DBM.

7.1. Derivation of Rectangular DBM.

With the same argument as in the derivation of the square DBM, we have

$$(7.21) \qquad d\lambda_\alpha = \frac{1}{\sqrt{N}} dB_\alpha(t) + \frac{1}{2N} \sum_{i=1}^{M} \sum_{j=M+1}^{M+N} \sum_{\beta \neq \pm\alpha} \partial_{X_{ij}}^{(2)} \lambda_\alpha \, dt,$$

where the Brownian motions $B_\alpha(t)$ are independent. We now expand the drift term as follows:

$$(7.22) \qquad \mathscr{X}_{\text{drift}}(t) = \frac{1}{2N} \sum_{i=1}^{M} \sum_{j=M+1}^{M+N} \sum_{\beta \neq \pm\alpha} \partial_{X_{ij}}^{(2)} \lambda_\alpha \, dt$$

$$(7.23) \qquad = \frac{1}{2N} \sum_{\beta \neq \pm\alpha} \sum_{i=1}^{M} \sum_{j=M+1}^{M+N} \frac{\mathbf{v}_\beta(i)\mathbf{v}_\alpha(j) + \mathbf{v}_\beta(j)\mathbf{v}_\alpha(i)}{\lambda_\alpha - \lambda_\beta} \left[\mathbf{v}_\beta(i)\mathbf{v}_\alpha(j) + \mathbf{v}_\beta(j)\mathbf{v}_\alpha(i) \right] dt$$

$$(7.24) \qquad = \frac{1}{2N} \sum_{\beta \neq \pm\alpha} \frac{1}{\lambda_\alpha - \lambda_\beta} \, dt.$$

Here, we appeal to Chapter I and use that for any $\mathbf{v}_\beta$ with $\lambda_\beta \in \zeta(X)$, we have

$$(7.25) \qquad \sum_{i=1}^{M} |\mathbf{v}_\beta(i)|^2 = 1$$

as well as the following vanishing identity for any index $j > M$:

$$(7.26) \qquad \mathbf{v}_\beta(j) = 0.$$

We now split the sum over eigenvalues λ_β into those in and not in $\zeta(X)$, respectively:

$$(7.27) \qquad \mathscr{X}_{\text{drift}}(t) = \frac{1}{2N} \sum_{\beta \notin \zeta(X)}^{(\pm\alpha)} \frac{1}{\lambda_\alpha - \lambda_\beta} + \frac{M-N}{2N\lambda_\alpha}.$$

Here, we used that any $\lambda \in \zeta(X)$ is equal to 0 and the size of $\zeta(X)$ is equal to $M - N$. Noting that by Assumption 5.2 the eigenvalue λ_α is bounded uniformly away from 0, the second term coming from those $\lambda_\beta \in \zeta(X)$ does not diverge, completing the derivation of the rectangular DBM.

7.2. Derivation of Theorem 5.5.

We proceed as in the derivation of Theorem 5.3 and use the Ito formula, giving us the following SDE:

$$(7.28) \qquad d\lambda_\alpha = \sum_{X_{ij} \in X} \partial_{X_{ij}} \lambda_\alpha dX_{ij} + \frac{1}{2} \sum_{X_{ij}, X_{k\ell} \in X} \partial_{X_{ij} X_{k\ell}} \lambda_\alpha \, d\langle X_{ij}, X_{k\ell} \rangle.$$

Proceeding as in the derivation of the square DBM and the rectangular DBM, and using the Ornstein-Uhlenbeck definition of dX_{ij} as given in (5.12), we now have

$$d\lambda_\alpha = \frac{1}{\sqrt{N}} dB_\alpha(t) + \left(\frac{1}{2N} \sum_{\beta \notin \zeta(X)}^{(\pm\alpha)} \frac{1}{\lambda_\alpha - \lambda_\beta} + \frac{1}{2N} \sum_{\gamma \in \zeta(X)} \frac{1}{\lambda_\alpha - \lambda_\gamma} \right) dt$$

$$(7.29) \qquad - \frac{1}{2} \sum_{X_{ij} \in X} X_{ij} \partial_{X_{ij}} \lambda_\alpha \, dt,$$

where the contribution from the additional drift term in (5.12) is given by the term in the second line. By the first-order perturbation formula for the eigenvalue λ_α, we see the contribution from this drift term is given by

$$(7.30) \qquad \frac{1}{2} \sum_{X_{ij} \in X} X_{ij} \partial_{X_{ij}} \lambda_\alpha = \frac{1}{2} \sum_{i,j} \mathbf{v}_\alpha(i) X_{ij} \mathbf{v}_\alpha(j) = \frac{1}{2} \lambda_\alpha$$

since the first term on the RHS is exactly equal to the standard dot product of the eigenvector $\mathbf{v}_\alpha$ with $X\mathbf{v}_\alpha = \lambda_\alpha \mathbf{v}_\alpha$. This completes the derivation of Theorem 5.5.

We conclude the derivation of DBM for linearized covariance matrices by discussing the derivation of DBM for the eigenvalues of the honest $N \times N$-dimensional covariance matrix $X_*(t)$, where $H(t)$ solves the matrix-valued SDE (5.1). In that case, we have the following corollary of Theorem 5.3.

Corollary 8.1. *In the context of Theorem 5.3, let $\{\xi_\alpha(t)\}_\alpha$ denote the eigenvalues of $X_*(t)$ realized as correlated paths on the simplex Δ_N. Then the eigenvalues $\{\xi_\alpha(t)\}_\alpha$ solve the following system of SDEs:*

$$(8.1) \qquad \mathrm{d}\xi_\alpha(t) = \frac{2\sqrt{\xi_\alpha(t)}}{\sqrt{N}}\mathrm{d}B_\alpha(t) + \left(\frac{1}{N} \sum_{\beta \neq \pm\alpha} \frac{\xi_\alpha + \xi_\beta}{\xi_\alpha - \xi_\beta} + \frac{M}{N} \right) \mathrm{d}t.$$

Remark 8.2. Historically, the DBM (8.1) was computed before the DBMs given in Theorem 5.3 and 5.5, e.g. in [7].

Before we proceed with a proof of Corollary 8.1, we first note a similar result holds with an additional drift term if we instead let $H(t)$ solve a matrix-valued Ornstein-Uhlenbeck equation. We omit the details as it will not be important for our discussion of universality, but we make note of it here for the interested reader.

Proof. Taking Theorem 5.3 for granted, by the spectral correspondence for covariance matrices in Chapter I we derive the SDE by applying the Ito formula to the smooth function $f(x) = x^2$. This gives

$$(8.2) \qquad \mathrm{d}\xi_\alpha(t) = \mathrm{d}(\lambda_\alpha(t))^2 = 2\lambda_\alpha(t)\mathrm{d}\lambda_\alpha(t)$$

$$(8.3) \qquad = \frac{2\sqrt{\xi_\alpha(t)}}{\sqrt{N}}\mathrm{d}B_\alpha(t) + \left(\frac{\lambda_\alpha}{N} \sum_{0 < \beta \neq \pm\alpha} \frac{1}{\lambda_\alpha - \lambda_\beta} + \frac{1}{\lambda_\alpha + \lambda_\beta} \right) \mathrm{d}t + \left(\frac{M}{N} - 1 \right) \mathrm{d}t.$$

We note the quadratic covariation term vanishes by noting the eigenvalue $\xi_\alpha(t)$ is symmetric, as a function, in the variables $\pm\lambda_\alpha(t)$. We now study the drift term, noting in particular the following identities:

$$(8.4) \qquad \frac{\lambda_\alpha}{N} \sum_{0 < \beta \neq \pm\alpha} \frac{1}{\lambda_\alpha - \lambda_\beta} + \frac{1}{\lambda_\alpha - \lambda_\beta} = \frac{1}{N} \sum_{|\beta| \neq |\alpha|} \frac{2\xi_\alpha}{\xi_\alpha - \xi_\beta},$$

$$(8.5) \qquad -1 = \frac{1}{N} \sum_{|\beta| \neq |\alpha|} \frac{\xi_\beta - \xi_\alpha}{\xi_\beta - \xi_\alpha}.$$

Thus we see the drift term is given by

$$(8.6) \qquad \left(\frac{\lambda_\alpha}{N} \sum_{0 < \beta \neq \pm\alpha} \frac{1}{\lambda_\alpha - \lambda_\beta} + \frac{1}{\lambda_\alpha + \lambda_\beta} \right) \mathrm{d}t + \left(\frac{M}{N} - 1 \right) \mathrm{d}t = \left(\frac{1}{N} \sum_{|\beta| \neq |\alpha|} \frac{\xi_\alpha + \xi_\beta}{\xi_\alpha - \xi_\beta} + \frac{M}{N} \right) \mathrm{d}t.$$

This completes the derivation of (8.1), again appealing to [2] to show the potential singularities occur with probability 0. □

We conclude our treatment of the derivation of DBM for the random covariance matrix $X_*(t)$ by discussing the upshot of studying first the DBM for the linearization $X(t)$. We first note the traditional derivation of the DBM of $X_*(t)$, i.e. those given in [1] and [7], involve nontrivial gymnastics in analyzing the nonlinearity in the covariance matrix $X_*(t)$.

The second upshot (and final one in this remark) of studying the eigenvalue dynamics along the linearized stochastics is that we may use those equations given in Theorem 5.3 to deduce the eigenvalue dynamics in Corollary 8.1. As illustrated in the proof above, this is a consequence of the Ito formula applied to the smooth function $f(x) = x^2$. On the other hand, to deduce the SDEs in Theorem 5.3 from those in Corollary 8.1 in a similar fashion, we would need to apply the Ito formula to the function $f(x) = \sqrt{x}$, for either choice of square root. This function is not even differentiable in a neighborhood of $x = 0$, so we need an a priori repulsion estimate as in Assumption 5.2. For our presentation, we only require this estimate for the rectangular regime $M > N$; in fact, for many matrix ensembles, such an estimate does not hold for $M = N$. This illustrates the weaker nature of the equations in Corollary 8.1 compared to the equations in Theorem 5.3.

III. A Strong Local Law for Deformed Linearized Covariance Matrices

We now discuss the rigidity estimates and local law necessary in the short-range approximation for DBM. To this end we introduce the following notion of high probability similar to the notions of high probability used in Chapters I and II.

Definition 9.1. We say an event Ω holds with (ξ, ν)-*high probability* if

$$(9.1) \qquad \mathbb{P}\left(\Omega^C\right) \;\leqslant\; e^{-\nu \log^{\xi} N}.$$

We say an event Ω_1 holds with (ξ, ν)-high probability *on* Ω_2 if

$$(9.2) \qquad \mathbb{P}\left(\Omega_1^C \cap \Omega_2\right) \;\leqslant\; e^{-\nu \log^{\xi} N}.$$

Lastly, before we state the local law for $X(t)$, we define the following spectral domains:

$$(9.3) \qquad \mathbf{D}_1 \;=\; \left\{ z = E + i\eta : \; E \in \mathscr{I}_{E_0, qG}, \quad \eta \in [N^{-1}\varphi^L, 10] \right\},$$

$$(9.4) \qquad \mathbf{D}_2 \;=\; \left\{ z = E + i\eta : \; |E| \leqslant N^{5B_V}, \quad \eta \in [10, N^{10B_V + 1}] \right\}.$$

Gluing together these domains, we define

$$(9.5) \qquad \mathbf{D}_{L,q} \;=\; \mathbf{D}_1 \cup \mathbf{D}_2.$$

From now on, we assume $X(t)$ solves the matrix-valued Brownian motion equation. We recall this gives the following parameterization $X(t) = X + \sqrt{t}W_\ell$. Moreover, we define the following *partial* Stieltjes transform of $X(t)$:

$$(9.6) \qquad m_N(z; t) \;=\; \frac{1}{2N} \sum_{i=1}^{N} \left(\frac{1}{\lambda_i - z} + \frac{1}{-\lambda_i - z} \right),$$

where the sum is taken over eigenvalues $\pm\lambda_i$ such that $|\lambda_i|^2 \in \sigma(X(t)^* X(t))$. We briefly remark that the partial Stieltjes transform was an object of primary interest in Chapter I of this thesis.

Theorem 9.2. *Suppose V is (g, G)-regular at E_0. Let*

$$(9.7) \qquad \xi \;=\; \frac{A_0 + o(1)}{2} \log \log N$$

and fix $q < 1$ and $L \geqslant 40\xi$. For any $z \in \mathbf{D}_{L,q}$, we have

$$(9.8) \qquad \sup_{t \in \mathscr{T}_\omega} |m_N(z; t) - m_{\mathrm{fc}, t}(z)| \;\leqslant\; \frac{\varphi^{C_1 \xi}}{N\eta}$$

with (ξ, ν)-high probability and M, N sufficiently large. Here, ν and C_1 are constants depending on all data involved in the statement of this result except the dimension M, N. The notion of M, N sufficiently large is in the sense of depending on all other data involved in the statement of this theorem.

From the local law in Theorem 9.2, we deduce an optimal rigidity estimate on the location of eigenvalues relative to their "expectation". To state this estimate, for any time t and any eigenvalue index i, first recall the classical eigenvalue locations, denoted $\gamma_{i,t}$ for $\rho_{\mathrm{fc},t}$. We also establish the following notation: for a constant $0 < q < 1$ and a time t we define

$$(9.9) \qquad \mathscr{A}_{q,t} \;=\; \left\{ i : \; \gamma_{i,t} \in \mathscr{I}_{E_0, qG} \right\}$$

for some fixed energy E_0. Letting $\lambda_{i,t}$ denote the i-th largest eigenvalue of $X(t)$, we may record the following rigidity estimate comparing $\gamma_{i,t}$ and $\lambda_{i,t}$. This rigidity estimate is a consequence of Theorem 9.2 by standard methods in Helffer-Sjostrand functional calculus; for details, we refer to [17].

Theorem 9.3. *For some constants $\nu, c_4 > 0$ depending on the data in Theorem 9.2, for $i \in \mathscr{A}_{q,t}$, we have*

$$
(9.10) \qquad \sup_{t \in \mathscr{T}_\omega} |\lambda_{i,t} - \gamma_{i,t}| \leqslant \frac{\varphi^{c_4 \xi}}{N}
$$

with (ξ, ν)-high probability. In particular, if V is (g, G)-regular for $G \asymp C$ sufficiently large, and if $\|V\|_\infty = O(1)$, then the rigidity estimate (9.10) holds for all indices.

We note here the extra assumption stated at the end of Theorem 9.3 is a simple consequence of the definition of $\mathscr{A}_{q,t}$. However, we will later consider this regime for extremely regular initial data V while performing calculations in establishing bulk universality. We discuss this assumption in more detail while studying the short-range approximation scheme.

Before we provide an outline of Theorem 9.2, we encourage the reader interested in studying the DBM equations to skip straight through this discussion of Theorem 9.2 and to the proof of gap universality. The ideas and methods towards the strong local law and rigidity estimates will not be important in analyzing the DBM equations in the rest of this chapter.

9.1. **Outline of Theorem 9.2.** We now provide an outline of the proof of Theorem 9.2. We follow the work of [17] and begin by using the Schur complement formula to derive a self-consistent equation for the Stieltjes transform m_N mimicking the self-consistent equation defining the Stieltjes transform $m_{\mathrm{fc},t}$ up to a sufficiently small error. Using the stability of this equation under small perturbations, we derive a *weak* local law, proving instead the following weaker estimate with high probability:

$$
(9.11) \qquad |m_N(z; t) - m_{\mathrm{fc},t}(z)| \leqslant \frac{\varphi^{C_1 \xi}}{\sqrt{N \eta}}.
$$

The method of studying stability of the self-consistent equation will be a standard bootstrapping from $\eta = 1$ to smaller scales with a self-improving estimate. We remark this bootstrapping technique *will not* resemble that used in Chapter I of this thesis. For a reference, we refer to [17]. From here, we introduce a result known as the *fluctuation averaging estimate*, also adapted from [17], to improve the weak local law to the strong local law. To obtain the estimate for all times, we compute the SDE driving the Stieltjes transform m_N and appeal to a standard stochastic continuity argument. This last result on uniformity in time is not necessary for our goal in deriving universality of local eigenvalue statistics, but we will use it for convenience, and we record and derive it also for its own sake and possible future use.

10. THE SELF-CONSISTENT EQUATION

The following matrix identity will be crucial in deriving a fixed-point equation for the Stieltjes transform $m_N(z; t)$. Before we give the statement of the result, we first recall the following notation for any set of indices $\mathbb{T} \subset \{1, \ldots, M + N\}$.

Notation 10.1. *We let $X^{(\mathbb{T})}$ denote the matrix obtained from removing the rows and columns indexed by elements of $\mathbb{T}$. We let $G^{(\mathbb{T})}$ denote the Green's function of $X^{(\mathbb{T})}$.*

Lemma 10.2. *Retaining the definition of $X(t)$, for any index $i \in [[1, N]]$, we have with $\mathbb{T} = (i, M + i)$,*

$$
(10.1) \qquad \begin{pmatrix} G_{ii} & G_{i,M+i} \\ G_{M+i,i} & G_{M+i,M+i} \end{pmatrix} = \left(\begin{pmatrix} -z & V_i + \sqrt{t} W_{ii} \\ V_i + \sqrt{t} W_{ii} & -z \end{pmatrix} - t v^* G^{(\mathbb{T})} v \right)^{-1}
$$

where

$$
(10.2) \qquad v^* = \begin{pmatrix} 0 & 0 & \ldots & 0 & W_{12} & W_{13} & \ldots & W_{1N} \\ W_{21} & W_{31} & \ldots & W_{M1} & 0 & 0 & \ldots & 0 \end{pmatrix}.
$$

Indeed, Lemma 10.2 follows from the Schur Complement Formula, which we record as follows for sake of completeness.

Proposition 10.3. *Suppose H is a Hermitian matrix of dimension N, so H may be written as*

$$
(10.3) \qquad H = \begin{pmatrix} A & B^* \\ B & C \end{pmatrix},
$$

where A is an $m \times m$ matrix, B is an $(N-m) \times m$ matrix, and C is an $(N-m) \times (N-m)$ matrix, for some $1 \leqslant m \leqslant N$. Here, B^* denotes the adjoint of B. If H is invertible, then for any set $\mathbb{T}$ of removed labels, we have

$$(10.4) \qquad \left[\left(H^{\mathbb{T}}\right)^{-1}\right]_{ij} = \left[\left((A - B^*CB)^{\mathbb{T}}\right)^{-1}\right]_{ij}$$

for any $i, j \notin \mathbb{T}$.

Lemma 10.2 now follows from removing all rows and columns from $X(t)$ except those corresponding to the indices $i, M+i$.

Remark 10.4. Alternatively, to derive the identity in Lemma 10.2, suppose $i = 1$; the case for other indices follows from a permutation of the matrix indices. We begin with following identity whose proof is a straightforward consequence of the Schur Complement Formula:

$$
\begin{aligned}
&G_{11}\left(-z - t\sum_{k,\ell=1}^{N-1} W_{1,k+1}G^{(\mathbb{T})}_{M-1+k,M-1+\ell}W_{1,\ell+1}\right) \\
(10.5) \qquad &+ G_{1,M+1}\left(V_1 + \sqrt{t}W_{11} - t\sum_{k=1}^{M-1}\sum_{\ell=1}^{N-1} W_{k+1,1}G^{(\mathbb{T})}_{k,M-1+\ell}W_{1,\ell+1}\right) = 1.
\end{aligned}
$$

From this identity, we deduce the $(1,1)$-entry of the following matrix product is 1:

$$(10.6) \qquad \begin{pmatrix} G_{11} & G_{1,M+1} \\ G_{M+1,1} & G_{M+1,M+1} \end{pmatrix} \left(\begin{pmatrix} -z & V_1 + \sqrt{t}W_{11} \\ V_1 + \sqrt{t}W_{11} & -z \end{pmatrix} - tv^*G^{(\mathbb{T})}v\right).$$

To compute the other entries of the above 2×2-matrix, we similarly expand the matrix product in terms of the entries of the matrices $G, G^{(\mathbb{T})}, W$, and the parameter z and apply the entrywise formulation of the Schur complement formula. From this calculation we deduce the desired matrix equation (10.1).

We now focus on using (10.1) to derive a fixed-point equation for $m_N(z;t)$. First, we establish the following two pieces of notation simply for convenience of presentation.

Notation 10.5. *For the index $k \in [[1, N]]$, we let G_{kk} denote the Green's function entry $G_{M+k,M+k}$. For any other index, i.e. those not denoted by k, we retain the usual matrix entry notation. For the following calculations, we let $k \leftrightsquigarrow M+1$ correspond to the first index larger than M.*

Notation 10.6. *We denote the entries of the matrix $v^*G^{(\mathbb{T})}v$ as in Lemma 10.2 by M_{ij} for $i, j = 1, 2$. Thus, we have the following formulas defining M_{ij}:*

$$(10.7) \qquad M_{11} = \sum_{k,\ell=1}^{N-1} W_{1,k+1}G^{(\mathbb{T})}_{M-1+k,M-1+\ell}W_{1,\ell+1},$$

$$(10.8) \qquad M_{12} = \sum_{k=1}^{M-1}\sum_{\ell=1}^{N-1} W_{k+1,1}G^{(\mathbb{T})}_{k,M-1+\ell}W_{1,\ell+1},$$

$$(10.9) \qquad M_{21} = \sum_{k=1}^{N-1}\sum_{\ell=1}^{M-1} W_{1,k+1}G^{(\mathbb{T})}_{M-1+k,\ell}W_{\ell+1,1},$$

$$(10.10) \qquad M_{22} = \sum_{k,\ell=1}^{M-1} W_{k+1,1}G^{(\mathbb{T})}_{k\ell}W_{\ell+1,1}.$$

We note that we are still focusing on the index $i = 1$.

We begin with the following representation of G_{kk} by taking the representation of the RHS of (10.1) in terms of minors, giving the following formula:

$$(10.11) \qquad G_{kk} = \frac{-z - tM_{11}}{(-z - tM_{11})(-z - tM_{22}) - (V_1 + \sqrt{t}W_{11} - tM_{12})(V_1 + \sqrt{t}W_{11} - tM_{21})}$$

$$(10.12) \qquad = \frac{-z - tM_{11}}{(-z - tM_{11})(-z - tM_{22}) - \left|V_1 + \sqrt{t}W_{11} - tM_{12}\right|^2},$$

where the second equation holds by the Hermitian property of the Green's function, which follows immediately from the spectral representation of the Green's function discussed in Chapter I.

We now expand the above representation of G_{kk}. To do so, we first define the following control parameters, most of which will serve as error terms in the desired self-consistent equation for $m_N(z;t)$. However, the reader is invited to skip straight to the next result and refer back to these control parameters upon reading the derivation of the self-consistent equation.

We first introduce the following terms which will serve as the main terms in the self-consistent equations:

$$(10.13) \qquad g_1^{\pm} = \frac{1}{\pm V_1 - z - tm_N(z;t)},$$

$$(10.14) \qquad \psi^{(\mathbb{T})} = m_N(z;t) - m_N^{(\mathbb{T})}(z;t).$$

With these main terms, we define the following eventual error terms:

$$(10.15) \qquad F_1 = t^2 \left(\frac{M_{11} - M_{22}}{2}\right)^2,$$

$$(10.16) \qquad F_2 = \frac{M_{11} + M_{22}}{2} - m_N^{(\mathbb{T})}(z;t),$$

$$(10.17) \qquad F_3^{\pm} = g_1^{\pm}\left(\pm\sqrt{t}W_{11} \mp tA_{12} - tF_2 + t\psi^{(\mathbb{T})}\right).$$

Lastly, we define the following accumulated error terms which will collect all error terms in the self-consistent equation:

$$(10.18) \qquad \mathscr{E}_1 = F_3^+ + F_3^- + F_3^+ F_3^- - g_1^+ g_1^- F_1,$$

$$(10.19) \qquad \mathscr{K}_1 = \frac{-t[(M_{11} - m_N^{(\mathbb{T})}(z;t)] - \psi^{(\mathbb{T})}}{-z - tm_N(z;t)}.$$

For general indices, the error parameters are defined by changing $1 \to i$ in all of the definitions. We define, for clarity's sake, the following accumulated error terms for general index i to emphasize which parameters depend on the index:

$$(10.20) \qquad \mathscr{E}_i = F_3^+(i) + F_3^-(i) + F_3^+(i)F_3^-(i) - g_i^+ g_i^- F_1(i),$$

$$(10.21) \qquad \mathscr{K}_i = \frac{-T[M_{11}(i) - m_N^{(\mathbb{T})}(z;T)] - \psi^{(\mathbb{T})}}{-z - Tm_N(z;T)}.$$

We now state the following result, from which we deduce a fixed-point equation for $m_N(z;t)$ up to a controllable error term. The proof of this result amounts to a straightforward expansion of the representation (10.12) in terms of the main terms $g_k^{\pm}$ and $\psi^{(\mathbb{T})}$ and the error terms defined above.

Proposition 10.7. *In the setting above, for indices $k = M+1, \ldots, M+N$, defining $i = k - M$, we have*

$$(10.22) \qquad G_{kk} = \frac{1}{2}\left(g_i^+ + g_i^-\right)\frac{1 + \mathscr{E}_i}{1 + \mathscr{K}_i}.$$

Proof. We focus on the case $k = M+1$ for simplicity; the case for other indices follows from a permutation of the rows and entries of $X(t)$. We first note the following identity controlling the first denominator term in (10.12):

$$(10.23) \qquad (-z - tM_{11})(-z - tM_{22}) = \left[-z - \frac{t}{2}(M_{11} + M_{22})\right]^2 - F_1.$$

We rewrite the representation (10.12) using the above identity and expand the denominator as follows:

$$(10.24) \qquad G_{kk} = \frac{-z - tM_{11}}{\left(-z - \frac{t}{2}(M_{11} + M_{22})\right)^2 - \left(V_1 + \sqrt{t}W_{11} - tM_{12}\right)^2 - F_1}$$

$$(10.25) \qquad = \frac{1}{\left(-z - tm_N^{(\mathbb{T})} + V_1 + \sqrt{t}W_{11} - tM_{12} - tF_2\right)\left(-z - tm_N^{(\mathbb{T})} - V_1 - \sqrt{t}W_{11} + tM_{12} - tF_2\right) - F_1}.$$

Here, we recall we take the index set $\mathbb{T} = \{1, M + 1\}$. Moreover, we may rewrite the above equation as follows introducing the term $\psi^{(\mathbb{T})}$ and the error terms $F_3^{\pm}$:

$$(10.26) \qquad G_{kk} = \frac{-z - tM_{11}}{\left(-z - tm_N + V_1\right)\left(1 + F_3^+\right)\left(-z - tm_N - V_1\right)\left(1 + F_3^-\right) - F_1}.$$

The proof of Proposition 10.7 now follows from a straightforward calculation and the definition of $\mathscr{E}_i, \mathscr{K}_i$. $\qquad\square$

If without the presence of the error terms $(1 + \mathscr{E}_i)(1 + \mathscr{K}_i)^{-1}$ in the refolded expansion (10.22), we would be able to simply average over indices k and obtain the same fixed-point equation for the free convolution Stieltjes transform $m_{\mathrm{fc},t}$. To deal with the error terms, we appeal to an a priori estimate which we will derive under suitable conditions. With this control on the error terms, we derive a perturbed form of the fixed-point equation for $m_{\mathrm{fc},t}$, from which we deduce a local law.

A Priori Estimate 10.8. *The following uniform estimate holds with with (ξ, ν)-high probability:*

$$(10.27) \qquad \frac{1 + \mathscr{E}_i}{1 + \mathscr{K}_i} = 1 + O\left(\sqrt{\frac{N^\varepsilon}{N\eta}}\right).$$

for $z = E + i\eta \in \mathbf{D}_{L,q}$ and for some constant C and any $\varepsilon > 0$.

With this a priori estimate on the error terms, we deduce the following perturbed self-consistent equation for m_N.

Corollary 10.9. *Assuming the upper bound (10.27), we have the following equation for all $z \in \mathbf{D}_{L,q}$ and $t \in \mathscr{T}_\omega$:*

$$(10.28) \qquad m_N(z;t) = \frac{1}{2N}\sum_{i=1}^N \left(\frac{1}{V_i - z - tm_N(z;t)} + \frac{1}{-V_i - z - tm_N(z;t)}\right) + O\left(\frac{N^\varepsilon}{N\eta}\right).$$

Here, $\varepsilon > 0$ is an arbitrarily small, N-independent constant.

Indeed, upon averaging the identity (10.22) over indices k and using the a priori bound (10.27), we deduce

$$(10.29) \qquad m_N(z;t) = \frac{1}{2N}\sum_{i=1}^N \left(\frac{1}{V_i - z - tm_N(z;t)} + \frac{1}{-V_i - z - tm_N(z;t)}\right)\left(1 + O\left(\sqrt{\frac{N^\varepsilon}{N\eta}}\right)\right).$$

The equation (10.28) now follows from the following short-time stability estimate particular to the matrix-valued Brownian motion $X(t)$. The proof of this result mirrors the arguments of Lemma 7.5 in [17] without introducing additional technical difficulties, so we omit it. In that paper, the estimate is derived for $m_{\mathrm{fc},t}$ instead of m_N, but since we assume the initial data V is also (g, G)-regular, the same proof holds.

Lemma 10.10. *Uniformly over $z \in \mathbf{D}_{L,q}$, we have*

$$(10.30) \qquad \frac{1}{2N}\sum_{i=1}^N \left(\frac{1}{|V_i - z - tm_N(z;t)|} + \frac{1}{|-V_i - z - tm_N(z;t)|}\right) = O(\log N).$$

Using Lemma 10.10, we deduce the following elementary bound on the big-Oh term upon possibly redefining $\varepsilon > 0$:

$$(10.31) \qquad \left|\frac{1}{2N}\sum_{i=1}^N \frac{O\left(\sqrt{\frac{N^\varepsilon}{N\eta}}\right)}{\pm V_i - z - tm_N(z;t)}\right| \leqslant O\left(\sqrt{\frac{N^\varepsilon}{N\eta}}\log N\right) \leqslant O\left(\sqrt{\frac{N^\varepsilon}{N\eta}}\right).$$

The signs on the initial data terms V_i indicate a summation over signs as well. This completes the derivation of the self-consistent equation (10.28).

We conclude this preliminary derivation of the self-consistent equation by recapping the arguments and calculations presented above. First, we deduce the equation (10.28) assuming the a priori bound (10.27). The derivation of this upper bound will follow from an a priori estimate on the following difference term we ultimately want to control:

$$(10.32) \qquad \Lambda(z) := |m_N(z;t) - m_{\mathrm{fc},t}(z)|.$$

Before we make this precise, we introduce another a priori estimate of crucial importance in exploiting the analytic structure of the self-consistent equation (10.28). This estimate is another stability estimate for short times $t \in \mathscr{T}_\omega$.

A Priori Estimate 10.11. *For any $t \in \mathscr{T}_\omega$, we have the following lower bound for any small $\delta > 0$:*

$$(10.33) \qquad \left| 1 - \frac{t}{2N} \sum_{i=1}^N \frac{1}{(\pm V_i - z - tm_N(z;t))(\pm V_i - z - tm_{\mathrm{fc},t}(z))} \right| \geqslant N^{-\delta}.$$

Here, the signs $\pm$ are chosen to be the same; in particular, only two possible pairs of signs are allowed.

We now make precise the use of controlling $\Lambda(z)$ to derive the bounds (10.27) and (10.33). This result will also allow us to begin our bootstrapping scheme at the scale $\eta = 1$.

Proposition 10.12. *For any fixed $z \in \mathbf{D}_{L,q}$, suppose either $\eta \geqslant 1$ or the following upper bound holds:*

$$(10.34) \qquad \Lambda(z) = O\left(\sqrt{\frac{N^\varepsilon}{N\eta}} \right).$$

Then with (ξ, ν)-high probability, the bound (10.27) holds uniformly over all indices k, and the bound (10.33) also holds.

Proof. For the proof of deducing (10.27), we refer to the proof of Lemma 7.9 in [17] which again mirrors the proof of (10.27). We briefly remark here that the necessary ingredients for the proof involve the Ward identity from Chapter I. Moreover, we require the concentration of the Gaussian terms appearing in $\mathscr{E}_i$ and $\mathscr{K}_i$. Lastly, we require the Cauchy Interlacing Lemma, which controls $\psi^{(\mathbb{T})}$ for any index set $\mathbb{T}$ of bounded size. For a proof of the Cauchy Interlacing Lemma, we refer to [2].

To derive the bound (10.33), we first note it suffices to provide a $O(N^{-\delta})$ upper bound for the following term:

$$(10.35) \qquad \frac{t}{2N} \sum_{i=1}^N \frac{1}{(\pm V_i - z - tm_N(z;t))(\pm V_i - z - tm_{\mathrm{fc},t}(z))}.$$

To derive this upper bound, we begin by rewriting each term in the summation as follows:

$$(10.36) \qquad \frac{1}{(V_i - z - tm_N(z;t))(V_i - z - tm_{\mathrm{fc},t}(z))} = \frac{1}{(V_i - z - tm_{\mathrm{fc},t}(z))^2} + \frac{t(m_N(z;t) - m_{\mathrm{fc},t}(z))}{(V_i - z - tm_N(z;t))(V_i - z - tm_{\mathrm{fc},t}(z))^2}.$$

We handle the first term on the RHS as follows. We cite (7.10) in Lemma 7.2 and (7.26) in Lemma 7.3 in [17] to obtain the following $O(1)$-bound:

$$(10.37) \qquad 1 - \frac{t}{2N} \sum_{i=1}^N \frac{1}{(\pm V_i - z - tm_{\mathrm{fc},t}(z))^2} = O(1)$$

We briefly remark here that [17] is concerned with the free convolution for the Wigner matrix ensemble – however, these arguments apply to proving the bound (10.37) as the only necessary ingredients are the regularity on the initial data and the qualitative properties of the free convolution that hold in both the Wigner matrix ensemble and our matrix ensemble of linearized covariance matrices. Again, the proofs mirror each other.

It now remains to bound the following term:

$$(10.38) \qquad \frac{t^2(m_N(z;t) - m_{\mathrm{fc},t}(z))}{2N} \sum_{i=1}^N \frac{1}{(\pm V_i - z - tm_N(z;t))(\pm V_i - z - tm_{\mathrm{fc},t}(z))^2}.$$

This follows from the Schwarz inequality and the following estimate uniform all indices $i \in [[1, N]]$:

$$(10.39) \qquad t \lesssim |V_i - z - Tm_{\mathrm{fc},t}(z)| \lesssim 1,$$

which, in turn, follow from the estimate $\mathrm{Im}\, m_{\mathrm{fc},t}(z) \asymp 1$. These estimates hold uniformly over $z \in \mathbf{D}_{L,q}$ and $t \in \mathscr{T}_\omega$, which completes the proof of Proposition 10.12. $\qquad\square$

11. Analysis of the Self-Consistent Equation

We begin our analysis of the self-consistent equation by establishing the following formula which will be the backbone of our bootstrapping method to obtain local laws at smaller scales η. The derivation of this formula will be a consequence of comparing the fixed-point equation satisfied by $m_{\mathrm{fc},t}$ and the perturbed equation (10.28) satisfied by $m_N(z;t)$.

Lemma 11.1. *Suppose $z \in \mathbf{D}_{L,q}$ and $t \in \mathscr{T}_\omega$. The following estimates hold with (ξ, ν)-high probability. Assuming the a priori bound (10.27), we have*

$$(11.1) \qquad (m_N(z;t) - m_{\mathrm{fc},t}(z))\left(1 - \frac{t}{2N}\sum_{i=1}^{N} \frac{1}{(\pm V_i - z - tm_N(z;t))(V_i - z - tm_{\mathrm{fc},t}(z))}\right) = O\left(\sqrt{\frac{N^\varepsilon}{N\eta}}\right).$$

In particular, assuming the bound (10.33) in addition, we have, for a possibly different $\varepsilon > 0$,

$$(11.2) \qquad \Lambda(z) := O\left(\sqrt{\frac{N^\varepsilon}{N\eta}}\right),$$

where the implied constant depends only on the equation (11.1).

Remark 11.2. The estimate (11.2) is known as a *weak local law* in contrast to the strong local law in Theorem 9.2.

Proof. We note it suffices to prove the first estimate, as the second estimate follows from the first and the lower bound in (10.33). We take the difference between the fixed-point equation for $m_{\mathrm{fc},t}$ and (10.28) to obtain the following perturbed equation:

$$(11.3) \qquad m_N(z;t) - m_{\mathrm{fc},t}(z) = \frac{1}{2N}\sum_{i=1}^{N}\left(\frac{1}{\pm V_i - z - tm_N(z;t)} - \frac{1}{\pm V_i - z - tm_{\mathrm{fc},t}(z)}\right) + O\left(\sqrt{\frac{N^\varepsilon}{N\eta}}\right).$$

Here, we appealed to the a priori bound (10.27) to derive the above equation. We now rewrite each term in the summation on the RHS as follows:

$$(11.4) \qquad \frac{1}{\pm V_i - z - tm_N(z;t)} - \frac{1}{\pm V_i - z - tm_{\mathrm{fc},t}(z)} = \frac{t\,(m_N(z;t) - m_{\mathrm{fc},t}(z))}{(\pm V_i - z - tm_N(z;t))(\pm V_i - z - tm_{\mathrm{fc},t}(z))}.$$

Having rewritten the summation term, we subtract it from both sides and obtain a $O\left(\sqrt{\frac{N^\varepsilon}{N\eta}}\right)$ bound on the following term:

$$(11.5) \quad m_N(z;t) - m_{\mathrm{fc},t}(z) - \frac{t(m_N(z;t) - m_{\mathrm{fc},t}(z))}{2N}\sum_{i=1}^{N}\frac{1}{(\pm V_i - z - tm_N(z;t))(\pm V_i - z - tm_{\mathrm{fc},t}(z))}$$

$$(11.6) \qquad = (m_N(z;t) - m_{\mathrm{fc},t}(z))\left(1 - \frac{t}{2N}\sum_{i=1}^{N}\frac{1}{(\pm V_i - z - tm_N(z;t))(\pm V_i - z - tm_{\mathrm{fc},t}(z))}\right).$$

This completes the proof of Lemma 11.1. $\qquad\square$

11.1. **The Bootstrapping Scheme.** We may now begin the bootstrapping estimates. Before we begin, we briefly outline and remark on the method behind obtaining a weak local law from Lemma 11.1. The key idea will be to use the local-Lipschitz property of $\Lambda(z)$ to extend the local law at a scale η to the scale $\eta - N^{-4}$. This provides an a priori estimate on $\Lambda(z)$, with which we may deduce a local law at the scale $\eta - N^{-4}$ by Lemma 11.1. We now remark that the local law obtained via Lemma 11.1 may be taken independent of the number of iterations of this local-Lipschitz extension. Moreover, iterating this scheme $N^{O(1)}$ times still yields a result that holds with (ξ, ν)-high probability. To obtain a weak local law uniformly over the domain $\mathbf{D}_{L,q}$, we use the local-Lipschitz property as in Chapter I. These last two points rely on the following result concerning (ξ, ν)-high probability events.

Lemma 11.3. *Suppose $\Xi_1, \ldots, \Xi_{N^C}$ is a collection of events, each holding with (ξ, ν)-high probability; here $C = O(1)$ is a fixed constant. Then the intersection of these events holds with (ξ, ν)-high probability for a possibly different pair of parameters (ξ, ν), i.e.*

$$(11.7) \qquad \mathbb{P}\left(\bigcup_{i=1}^{N^C} \Xi_i^C \right) \leqslant e^{-\nu \log^\xi N}.$$

Moreover, if $\xi \gg 1$ in the limit of large N, then the adjusted parameter ξ also diverges as N grows.

As an immediate consequence of this general result it now suffices to obtain a pointwise estimate. We now state the desired estimate precisely in the following proposition.

Proposition 11.4. *Assume the setting of Theorem 9.2. Then for any $z \in \mathbf{D}_{L,q}$ and any $\varepsilon > 0$, we have*

$$(11.8) \qquad \Lambda(z) = O\left(\sqrt{\frac{N^\varepsilon}{N\eta}} \right).$$

Proof. For a fixed $z \in \mathbf{D}_{L,q}$, we define $z_k = E + i\eta_k$. Here, we have defined the parameters as follows:

$$(11.9) \qquad E = \mathrm{Re}(z), \quad \eta_k = 1 - kN^{-4}.$$

Lastly, we let K denote the maximal integer such that $z_K \in \mathbf{D}_{L,q}$; in particular we have the bound $K = N^{O(1)}$. By N^2-Lipschitz continuity of the Stieltjes transforms, and thus of Λ, to derive the weak local law at z, it suffices to derive the weak local law at z_k for all $k = 1, \ldots, K$. We do so inductively in k.

We note the weak local law at z_0 holds with (ξ, ν)-high probability by Proposition 10.12. This allows us to obtain the following estimate for z_1:

$$(11.10) \qquad \Lambda(z_1) \leqslant |\Lambda(z_1) - \Lambda(z)| + \Lambda(z) \leqslant N^2 |z_1 - z| + O\left(\sqrt{\frac{N^\varepsilon}{N\eta}} \right)$$

$$(11.11) \qquad \leqslant N^{-4} + O\left(\sqrt{\frac{N^\varepsilon}{N\eta}} \right).$$

By Proposition 10.12 and Lemma 11.1, this implies

$$(11.12) \qquad \Lambda(z_1) = O\left(\sqrt{\frac{N^\varepsilon}{N\eta_1}} \right).$$

Continuing inductively, we see for any $k \in [[1, K]]$ that the weak local law holds:

$$(11.13) \qquad \Lambda(z_k) = O\left(\sqrt{\frac{N^\varepsilon}{N\eta}} \right),$$

where the implied constant is independent of the index k. This completes the proof. $\qquad \square$

With a similar bootstrapping method, we may also obtain an estimate of similar type for the off-diagonal entries of the Green's function G. To state this estimate, we first define the following control parameter as a function on $\mathbf{D}_{L,q}$:

$$(11.14) \qquad \Lambda_o(z) \;=\; \max_{i \neq j > M} |G_{ij}(z)|.$$

In words, the parameter Λ_o is a pseudo-maximal function for the off-diagonal entries. To control this term, we appeal to the following result from [18], which allows us to control Λ_o in terms of the weak local law.

Lemma 11.5. *Assume the following a priori estimate:*

$$(11.15) \qquad \Lambda_o(z) + \Lambda(z) \;\leqslant\; \varphi^{-2\xi}$$

where we retain the definition of the control parameters φ, ξ from the introduction of the underlying model. Then we have

$$(11.16) \qquad \sup_{z \in \mathbf{D}_{L,q}} \Lambda_o(z) \;=\; O\left(\varphi^\xi \sqrt{\frac{\Lambda(z) + O(1)}{N\eta}} \right)$$

with (ξ, ν)-high probability, where the implied constant is independent of z.

The result in Lemma 11.5 is a parallel result to Lemma 3.8 in [18]. The method of proof is analogous between the two matrix ensembles, so we refer to the details of Lemma 3.8 in [18] for the proof of Lemma 11.5.

We may perform a similar bootstrapping scheme to estimate Λ_o, with the initial step at $z = E + iN^C$ for any $C \geqslant 0$. In effect, we obtain the following consequence of Lemma 11.5.

Corollary 11.6. *In the setting of Theorem 9.2, we have the following estimate uniformly over $z \in \mathbf{D}_{L,q}$ with (ξ, ν)-high probability:*

$$(11.17) \qquad \sup_{z \in \mathbf{D}_{L,q}} \Lambda_o(z) \;=\; O\left(\varphi^\xi \sqrt{\frac{\Lambda(z) + O(1)}{N\eta}} \right).$$

The last weak result we discuss here is the following estimates on diagonal entries of the Green's function. In this spirit, we similarly define a pseudo-maximal function controlling the diagonal entries as follows:

$$(11.18) \qquad \Lambda_d(z) \;=\; \max_{M+1 \leqslant k \leqslant M+N} |G_{kk}(z)|.$$

On one hand, by eigenvector delocalization and the weak local law, we may bound the diagonal entries by the Stieltjes transform up to a factor of N^ε for any small $\varepsilon > 0$, and we may compare the Stieltjes transform of G to the Stieltjes transform of the free convolution by the weak local law. Ultimately, this argument provides the following bound for any small $\varepsilon > 0$:

$$(11.19) \qquad \Lambda_d(z) \;=\; O(N^\varepsilon).$$

Remark 11.7. We briefly remark the above estimate on Λ_d may be obtained with a similar bootstrapping scheme which allows for a sharper estimate comparing the diagonal entries to the free convolution m_{fc}. This more precise estimate will not be necessary in this thesis, and it requires a nontrivial extension of the bootstrapping scheme coupled with ideas from [18], so we omit the details and refer the reader to [18].

12. The Fluctuation Averaging Lemma and the Strong Local Law

We now aim to conclude the strong local law in Theorem 9.2 from the weak local law in Proposition 11.4. To do so, we need to appeal to a result controlling fluctuations of Green's function entries; this is called the *fluctuation averaging lemma.* This result is ubiquitous in deriving strong local laws, as it will help us control the nonlinear terms appearing in the matrix equation (10.1), i.e. those nonlinear terms in the Schur complement formula. For instance, we refer to [17]. To this end, we establish the following notation.

Notation 12.1. *For a set of indices* $\mathbb{T}$*, we define the following conditional expectation and fluctuation of a random variable* X *defined as a function of the matrix entries of* $X(t)$*:*

$$(12.1) \qquad \mathbb{E}_{\mathbb{T}} X \; := \; \mathbb{E}\left(X|X(t)^{\mathbb{T}}\right), \quad Q_{\mathbb{T}} X \; := \; X - \mathbb{E}_{\mathbb{T}} X.$$

Alternatively, the conditional expectation with respect to $\mathbb{T}$ may be thought of as a conditional expectation with respect to those matrix entries of $X(t)$ whose indices are not contained in $\mathbb{T}$.

We now define the fluctuation terms that we aim to control in the fluctuation averaging lemma. These terms are defined as the following fluctuations of nonlinear functions of the Green's function:

$$(12.2) \qquad Z_{11}^{(1)} \; = \; Q_{(1)}\left(\frac{G_{M+1,M+1}}{G_{11}G_{M+1,M+1} - G_{1,M+1}G_{M+1,1}}\right),$$

$$(12.3) \qquad Z_{12}^{(1)} \; = \; Q_{(1)}\left(\frac{G_{M+1,1}}{G_{11}G_{M+1,M+1} - G_{1,M+1}G_{M+1,1}}\right).$$

We may define the more general fluctuation term replacing the labels 1 on the RHS above with any index $i \in [[1, M]]$. Secondly, we define another pair of fluctuation terms; although these terms are not important in the statement of the fluctuation averaging lemma, we use them later in deriving the strong local law:

$$(12.4) \qquad Z_{21}^{(1)} \; = \; Q_{(1)}\left(\frac{G_{1,M+1}}{G_{11}G_{M+1,M+1} - G_{1,M+1}G_{M+1,1}}\right),$$

$$(12.5) \qquad Z_{22}^{(1)} \; = \; Q_{(1)}\left(\frac{G_1}{G_{11}G_{M+1,M+1} - G_{1,M+1}G_{M+1,1}}\right).$$

Again, we define $Z_{k\ell}^{(i)}$ similarly by replacing the 1-indices on the RHS with $i \in [[1, M]]$.

We now state the fluctuation averaging lemma.

Proposition 12.2. *Suppose* $a_k \asymp 1$ *are fixed constants, and assume the following a priori bounds for any fixed* $\varepsilon > 0$*:*

$$(12.6) \qquad \sup_{\eta \geqslant N^{-1+\varepsilon}} \sup_{i \neq j > M} |G_{ij}(E + i\eta)| \; \leqslant \; O\left(\sqrt{\frac{N^{\varepsilon}}{N\eta}}\right), \qquad \sup_{\eta \geqslant N^{-1+\varepsilon}} \sup_{M+1 \leqslant k \leqslant M+N} |G_{kk}(E + i\eta)| \; = \; O(N^{\varepsilon}).$$

Then, we have, for $z = E + i\eta$ *with* $\eta \geqslant N^{-1+\varepsilon}$*,*

$$(12.7) \qquad \frac{1}{N} \sum_{i=1}^{N} a_k Z_{11}^{(k)} \; = \; O\left(\frac{N^{\varepsilon}}{N\eta}\right),$$

$$(12.8) \qquad \frac{1}{N} \sum_{i=1}^{N} a_k Z_{12}^{(k)} \; = \; O\left(\frac{N^{e}}{N\eta}\right)$$

with (ξ, ν)*-high probability, where* $k = M + i$*.*

Remark 12.3. The a priori bounds (12.6) assumed in deriving the fluctuation averaging lemma are fulfilled in our context.

Proof. We give an outline of the proof; for details, we refer to Lemma 7.15 in [17]. In particular, Proposition 12.2 follows an application of the Chebyshev inequality combined with the following moment bounds:

$$(12.9) \qquad \mathbb{E}\left(\frac{1}{N^{2p}} \sum_{i_1,\ldots,i_{2p}=1}^{N} \prod_{j=1}^{2p} a_{k_j} Z_{11}^{(i_j)}\right) \; = \; O\left(\frac{N^{\varepsilon}}{N\eta}\right),$$

and the analogous estimate replacing Z_{11} with Z_{12}. This moment bound is quite technical and requires a combinatorial analysis of words on indices. For this reason we omit the proof, but we encourage the reader to look at [17] for details. $\qquad \square$

12.1. **Proof of the Strong Local Law.** We now use Proposition 12.2 to derive the strong local law uniformly over $z \in \mathbf{D}_{L,q}$, but at *fixed* times. To this end, we define the following error terms that will arise naturally from studying the matrix equation (10.1):

$$(12.10) \qquad R_{11}^{(1)} = \frac{1}{N} G_{M+1,M+1} + \frac{1}{N} \sum_{i=1}^{N} G_{M+1,k} G_{k,M+1},$$

$$(12.11) \qquad R_{22}^{(1)} = \frac{1}{N} G_{11} + \frac{1}{N} \sum_{i=1}^{N} G_{1,k} G_{k,1}.$$

Again, we retain the notation $k = M + i$. We similarly define $R_{jj}^{(\ell)}$ upon replacing $\ell = 1$ on the RHS above with any appropriate index $\ell \in [[1, N]]$. Lastly, before we proceed directly with the proof of the strong local law, we record the following consequence of the fluctuation averaging lemma:

$$(12.12) \qquad \sup_{z \in \mathbf{D}_{L,q}} \sup_{i \in [[1,M]]} \sup_{j=1,2} |R_{jj}^{(i)}| \leqslant \frac{N^{\varepsilon}}{N\eta}.$$

Having defined the error terms $R_{jj}^{(i)}$, we now rewrite the matrix equation (10.1) in the following manner, in which we extract these error terms, along with the fluctuations Z:

$$\begin{pmatrix} G_{ii} & G_{i,M+i} \\ G_{M+i,i} & G_{M+i,M+i} \end{pmatrix}^{-1} = \begin{pmatrix} -z - tm_{\mathrm{fc},t}(z) & V_i \\ V_i & -z - tm_{\mathrm{fc},t}(z) \end{pmatrix}$$

$$+ \begin{pmatrix} t\left(m_N(z;t) - m_{\mathrm{fc},t}(z)\right) & 0 \\ 0 & t\left(m_N(z;t) - m_{\mathrm{fc},t}(z)\right) \end{pmatrix}$$

$$(12.13) \qquad\qquad + t\mathbf{Z}^{(i)} + t\mathbf{R}^{(i)}.$$

Here, the matrices $\mathbf{Z}$ and $\mathbf{R}$ are defined by the terms $Z_{k\ell}^{(i)}$ and $R_{k\ell}^{(i)}$. For clarity's sake, we note $\mathbf{R}^{(i)}$ is diagonal. For convenience, we will introduce the following notation for the matrices on the RHS above:

$$(12.14) \qquad \mathscr{Y}_{V_i} = \begin{pmatrix} -z - tm_{\mathrm{fc},t}(z) & V_i \\ V_i & -z - tm_{\mathrm{fc},t}(z) \end{pmatrix},$$

$$(12.15) \qquad \mathscr{Y}_d = \begin{pmatrix} t\left(m_N(z) - m_{\mathrm{fc},t}(z)\right) & 0 \\ 0 & t\left(m_N(z) - m_{\mathrm{fc},t}(z)\right) \end{pmatrix}.$$

We briefly explain the decomposition given above. The matrix $\mathscr{Y}_{V_i}$ detects the main terms which yield the Stieltjes transform of the free convolution measure, containing the ingredients of the fixed-point equation defining the free convolution. The matrix $\mathscr{Y}_{d,i}$ denotes the error term we want to control. The remaining matrices $\mathbf{Z}^{(i)}$ and $\mathbf{R}^{(i)}$ contain the error terms that we may control by appealing to the concentration of sub-Gaussian random variables.

We now proceed with the analysis of the rewritten matrix equation (12.13). By the resolvent identity discussed in Chapter I of this thesis, we have the following uniform bound for any small $\varepsilon > 0$:

$$(12.16) \qquad \sup_{i \in [[1,N]]} \left\| \frac{1}{N} \sum_{i=1}^{N} \left[\left(\mathscr{Y}_{V_i} - \mathscr{Y}_d + t\mathbf{Z}^{(i)} + t\mathbf{R}^{(i)} \right)^{-1} - \left(\mathscr{Y}_{V_i} - \mathscr{Y}_d \right)^{-1} \right] \right\|_{\infty} \leqslant C \frac{N^{\varepsilon}}{N\eta}.$$

This is a consequence of a straightforward application of the diagonal estimates on $\mathbf{R}^{(i)}$ we obtained as well as the fluctuation averaging lemma, both coupled with the a priori bounds in the statement of Proposition 12.2. This allows us to Taylor expand resolvents of matrices in (12.13) to obtain the following main term-error term decomposition of the LHS of (12.13):

$$\frac{1}{N} \sum_{i=1}^{N} \begin{pmatrix} G_{ii} & G_{ik} \\ G_{ki} & G_{kk} \end{pmatrix} = \frac{1}{N} \sum_{i=1}^{N} \left(\mathscr{Y}_{V_i} - \mathscr{Y}_d \right)^{-1} + O\left(\frac{N^{\varepsilon}}{N\eta} \right)$$

$$(12.17) \qquad\qquad = \frac{1}{N} \sum_{i=1}^{N} \mathscr{Y}_{V_i}^{-1} - \frac{1}{N} \sum_{i=1}^{N} \mathscr{Y}_{V_i}^{-1} \mathscr{Y}_d \mathscr{Y}_{V_i}^{-1} + O\left(\frac{N^{\varepsilon}}{N\eta} \right)$$

where the big-Oh terms denote matrices with the corresponding operator bound. The second line follows from the geometric series expansion of the inverse of a matrix noting the $o(1)$ estimates on Λ, Λ_o provided by the weak local law.

Remark 12.4. We note that if an inverse written above does not exist, a perturbation ε in the entries can be introduced and then taken to $\varepsilon \to 0$ in the end of this calculation. We do not fully illustrate this in detail for convenience and clarity of the presentation of the calculation.

We now take a partial trace of both sides of the last matrix equation (12.17). In particular, we sum over diagonal entries with index larger than M. This gives us the following equation:

$$\frac{1}{N} \sum_{i=1}^{N} G_{kk} = \frac{1}{N} \sum_{i=1}^{N} \frac{-z - t m_{\text{fc},t}(z)}{(-z - t m_{\text{fc},t}(z))^2 - V_i^2}$$

$$- \left(m_N(z;t) - m_{\text{fc},t}(z) \right) \frac{t}{N} \sum_{i=1}^{N} \operatorname{Tr} \mathscr{Y}_{V_i}^{-2}$$

(12.18)
$$+ O\left(\frac{N^\varepsilon}{N\eta} \right).$$

We now study the trace equation (12.18), noting the term on the LHS is exactly m_N, and the first term on the RHS is exactly $m_{\text{fc},t}$. A straightforward rearrangement of the above trace equation now implies the following estimate:

(12.19)
$$\left(m_N(z;t) - m_{\text{fc},t}(z) \right) \left(1 + \mathscr{E}_t \right) = O\left(\frac{N^\varepsilon}{N\eta} \right),$$

where we defined

(12.20)
$$\mathscr{E}_t = \frac{t}{N} \sum_{i=1}^{N} \operatorname{Tr} \mathscr{Y}_{V_i}^{-2}.$$

Thus, to derive the strong local law for a fixed point $z \in \mathbf{D}_{L,q}$, it suffices to derive a constant *lower* bound on $1 + \mathscr{E}_t$. To this end we appeal to (7.10) in Lemma 7.2 and (7.26) in Lemma 7.3, both in [17]. This completes the proof of the strong local law.

13. Stochastic Continuity of the Strong Local Law

We conclude this discussion of the strong local law with a stochastic continuity argument that will allow us to deduce a high-probability result uniformly in time. We reiterate that this result is unnecessary in studying universality, but may be of interest and potential use in its own right.

13.1. **Stochastic Dynamics of the Stieltjes Transform.** The backbone of this discussion rests on the following calculation with the Ito formula. In effect, it computes the stochastic dynamics of the Stieltjes transform under the matrix-valued Brownian motion dynamics on $X(t)$. Before we give the result, we recall that the eigenvalue dynamics under the matrix-valued Brownian dynamics are given by the following system of SDEs:

(13.1)
$$d\lambda_\alpha(t) = \frac{1}{\sqrt{N}} dB_\alpha(t) + \frac{1}{2N} \left(\sum_{\beta \notin \zeta_t(X)}^{(\pm\alpha)} \frac{1}{\lambda_\alpha - \lambda_\beta} + \sum_{\gamma \in \zeta_t(X)} \frac{1}{\lambda_\alpha - \lambda_\gamma} \right) dt$$

We also recall the Brownian motions $B_\alpha(t)$ are jointly independent. With these eigenvalue dynamics, an application of the Ito formula will give the following Stieltjes transform dynamics. To suggest the suitable interpretation of these dynamics, we write the integral equation instead of the differential equation.

Proposition 13.1. *Suppose $\{\lambda_\alpha(t)\}_\alpha$ solves the system of SDEs given in (13.1), and let $m_N(z;t)$ denote the Stieltjes transform of $X(t)$. Moreover, suppose $\eta = \mathrm{Im}(z) > 0$. Then, for any times $0 \leqslant t_0 \leqslant t_1$, we have*

$$
m_N(z; t_1) - m_N(z; t_0) = \sum_{\lambda_\alpha \not\equiv 0} \frac{1}{\sqrt{N}} \int_{t_0}^{t_1} \frac{1}{(\lambda_\alpha - z)^2} \, \mathrm{d}B_\alpha(t)
$$

(13.2)
$$
+ \left(\frac{1}{N} \sum_{\lambda_\alpha \not\equiv 0} \frac{1}{(\lambda_\alpha - z)^3} + \frac{1}{2N} \sum_{\lambda_\alpha \not\equiv \pm \lambda_\beta} \frac{\lambda_\alpha + \lambda_\beta - 2z}{(\lambda_\alpha - z)^2 (\lambda_\beta - z)^2} + \frac{M - N}{\lambda_\alpha} \right) \mathrm{d}t.
$$

Proof. Because the Brownian motions $B_\alpha(t)$ are jointly independent, they are also orthogonal, i.e.

(13.3)
$$
\mathrm{d}\langle \lambda_\alpha(t), \lambda_\beta(t) \rangle = \frac{\delta_{\alpha\beta}}{N} \mathrm{d}t.
$$

From this, by the Ito formula, at least formally we have

(13.4)
$$
m_N(z; t_1) - m_N(z; t_0) = \sum_{\lambda_\alpha \not\equiv 0} \int_{t_0}^{t_1} \partial_{\lambda_\alpha} m_N(z;t) \mathrm{d}\lambda_\alpha + \frac{1}{2N} \sum_{\lambda_\alpha \not\equiv 0} \partial^{(2)}_{\lambda_\alpha} m_N(z;t) \, \mathrm{d}t.
$$

From here, it suffices to calculate the derivatives of m_N with respect to the nontrivial eigenvalues $\lambda_\alpha \not\equiv 0$; because m_N is nonsingular at $z = E + i\eta \in \mathbb{C}_+$, the following derivative identities hold rigorously as well:

(13.5)
$$
\partial_{\lambda_\alpha} m_N(z;t) = \sum_{\lambda_\alpha} -\frac{1}{(\lambda_\alpha - z)^2},
$$

(13.6)
$$
\partial^{(2)}_{\lambda_\alpha} = \frac{1}{2} \sum_{\lambda_\alpha} \frac{1}{(\lambda_\alpha - z)^3}.
$$

Lastly, we remark that the calculation with Ito's formula holds rigorously at $z = E + i\eta \in \mathbb{C}_+$, a point at which m_N is smooth as a function of the real eigenvalues λ_α. □

We remark that the equation (13.2) removes the direct dependence on the eigenvalue gaps for the dynamics of the Stieltjes transform. This is a consequence of averaging over eigenvalues in the spectral representation of the Stieltjes transform; this observation is crucial in deriving sufficient estimates to run a stochastic continuity argument.

Remark 13.2. As will be noted later in this thesis in a more detailed context and fashion, the problem of tracking eigenvalue gaps along the SDE (13.1) is a difficult problem. Moreover, this thesis will only address a partial result along these lines, as the optimal result on eigenvalue gaps *for linearized covariance matrices* is, for now, out of reach.

We now study the SDE (13.2) and extract a short-time stability estimate. This will conclude our analysis of the SDE in this thesis. The proof of this short-time stability estimate will require a general result in stochastic analysis that controls the concentration of Ito integrals of bounded functions. To state this result, we recall that a sub-exponential random variable X is one whose decay at infinity is bounded as follows:

(13.7)
$$
\mathbb{P}(|X| \geqslant \Lambda\sigma) \leqslant C_0 e^{-\Lambda^{1/\theta}}.
$$

In particular, sub-exponential random variables satisfy the following high-probability concentration phenomenon.

Proposition 13.3. *Suppose $(a_i)_{i \in I}$ and $(b_i)_{i \in I}$ are independent families of centered random variables with variance σ^2 and sub-exponential decay in the following sense:*

(13.8)
$$
\mathbb{P}(|a_i| \geqslant x\sigma) + \mathbb{P}(|b_i| \geqslant x\sigma) \leqslant C_0 e^{-x^{1/\theta}},
$$

where $C_0 > 0$ and $\theta \geqslant 1$ are constants uniform over $i \in I$. Suppose $A_i, B_{ij} \in \mathbb{C}$ are deterministic constants. Then, for some constants $a_0 > 1$, $A_0 \geqslant 10$, and $C_1 \geqslant 1$ depending on θ, C_0, for any parameter $a_0 \leqslant \xi \leqslant A_0 \log\log N$ and $\varphi = \log^{C_1} N$, we

have the following concentration inequalities:

$$(13.9) \qquad \mathbb{P}\left(\left| \sum_{i=1}^{N} A_i a_i \right| \geqslant \varphi^{\xi} \sigma \left(\sum_{i=1}^{N} |A_i|^2 \right)^{1/2} \right) \leqslant e^{-\log^{\xi} N},$$

$$(13.10) \qquad \mathbb{P}\left(\left| \sum_{i=1}^{N} \overline{a_i} B_{ii} a_i - \sum_{i=1}^{N} \sigma^2 B_{ii} \right| \geqslant \varphi^{\xi} \sigma \left(\sum_{i=1}^{N} |B_{ii}|^2 \right)^{1/2} \right) \leqslant e^{-\log^{\xi} N},$$

$$(13.11) \qquad \mathbb{P}\left(\left| \sum_{i \neq j}^{N} \overline{a_i} B_{ij} a_j \right| \geqslant \varphi^{\xi} \sigma \left(\sum_{i \neq j} |B_{ij}|^2 \right)^{1/2} \right) \leqslant e^{-\log^{\xi} N},$$

$$(13.12) \qquad \mathbb{P}\left(\left| \sum_{i,j=1}^{N} \overline{a_i} B_{ij} b_j \right| \geqslant \varphi^{\xi} \sigma \left(\sum_{i,j=1}^{N} |B_{ij}|^2 \right)^{1/2} \right) \leqslant e^{-\log^{\xi} N}.$$

We now present a general result in stochastic calculus which states that Ito integrals of bounded functions are indeed sub-exponential.

Proposition 13.4. *Let $t > 0$ be a positive time and assume $f(x) \in C^2(\mathbb{R})$ is bounded, adapted and measurable. Suppose X_t is given by the following Ito integral:*

$$(13.13) \qquad X_t = \int_0^t f(X_s) \, \mathrm{d}B_s.$$

Then $\sup_{t \leqslant T} X_t$ is sub-exponential, and thus with (ξ, ν)-high probability

$$(13.14) \qquad \sup_{t \leqslant T} |X_t| \leqslant C \sigma_T$$

for some constant $C = O(1)$. Here, σ_T is the standard deviation of X_T.

Remark 13.5. We warn the reader that this stability estimate will deteriorate linearly in time for long-time evolutions and infinitesimally small scales η, though this will not present itself as an issue in our application.

Roughly speaking, the proof of Proposition 13.4 is a consequence of the Ito formula to obtain bounds on the even moments of the Ito integral X_t. The sub-exponential decay is then deduced from these moment bounds and Chernoff's inequality.

Proposition 13.4 allows us to control the martingale term in the Stieltjes transform dynamics. The precise result is given as follows.

Lemma 13.6. *Suppose $0 \leqslant t_0 \leqslant t_1$ are two times. Then we have uniformly over $z \in \mathbf{D}_{L,q}$ the following bound with (ξ, ν)-high probability:*

$$(13.15) \qquad \sup_{t \in [t_0, t_1]} |m_N(z;t) - m_N(z;t_0)| = O\left((t_1 - t_0) \left(\frac{1}{\eta^3} + \frac{N^{B_V}}{\eta^4} + N^{5B_V} + 1 + \frac{1}{\eta^4} \right) \right).$$

Proof. We first consider the drift term in the SDE (13.2). Suppose $z \in \mathbf{D}_1$, so that the drift term is bounded as follows:

$$(13.16) \qquad \left| \frac{1}{N} \sum_{\lambda_\alpha \neq 0} \frac{1}{(\lambda_\alpha - z)^3} + \frac{1}{2N} \sum_{\lambda_\alpha \neq \pm \lambda_\beta} \frac{\lambda_\alpha + \lambda_\beta - 2z}{(\lambda_\alpha - z)^2 (\lambda_\beta - z)^2} + \frac{M - N}{\lambda_\alpha} \right| = O\left(\frac{1}{\eta^3} + \frac{N^{B_V}}{\eta^4} + 1 \right).$$

For $z \in \mathbf{D}_2$, we note the first two drift terms are decreasing as η increases, and the third drift term is z-independent. Thus, we may assume $\eta = 10$, in which case we have the bound

$$(13.17) \qquad \left| \frac{1}{N} \sum_{\lambda_\alpha \neq 0} \frac{1}{(\lambda_\alpha - z)^3} \right| + \left| \frac{1}{2N} \sum_{\lambda_\alpha \neq \pm \lambda_\beta} \frac{\lambda_\alpha + \lambda_\beta - 2z}{(\lambda_\alpha - z)^2 (\lambda_\beta - z)^2} \right| + \left| \frac{M - N}{\lambda_\alpha} \right| \leqslant O\left(1 + N^{5B_V} \right).$$

Here, we use the a priori repulsion from the origin in bounding the singular eigenvalue term $\frac{M-N}{\lambda_\alpha}$. This provides the appropriate bound for the drift terms.

It remains to bound the diffusion terms with high probability. Because the diffusion terms are bounded, adapted and measurable, by Proposition 13.3 we have the following estimate uniformly over $z \in \mathbf{D}_{L,q}$ with (ξ, ν)-high probability:

$$(13.18) \qquad \sup_{t \in [t_0, t_1]} \left| \sum_{\lambda_\alpha \neq 0} \frac{1}{\sqrt{N}} \int_{t_0}^{t} \frac{1}{(\lambda_\alpha - z)^2} \, \mathrm{d}B_\alpha(t) \right| \leqslant C \, \mathbb{E} \left(\int_{t_0}^{t_1} \frac{1}{(\lambda_\alpha - z)^2} \, \mathrm{d}B_\alpha(t) \right)^2.$$

We compute the RHS with Ito's L^2-isometry to obtain the following bound for the martingale term:

$$(13.19) \qquad \sup_{t \in [t_0, t_1]} \left| \sum_{\lambda_\alpha \neq 0} \frac{1}{\sqrt{N}} \int_{t_0}^{t} \frac{1}{(\lambda_\alpha - z)^2} \, \mathrm{d}B_\alpha(t) \right| = O\left(\frac{t_1 - t_0}{\eta^4} \right).$$

This completes the proof of Lemma 13.6. $\qquad \square$

13.2. **Free Convolution Estimates and Stochastic Continuity.**
Before we may implement a stochastic continuity argument, we need an estimate on the short-time evolution of the Stieltjes transform of the free convolution. This result is a consequence of either straightforward estimates from classical calculus or the bootstrapping iteration scheme used for a fixed-point equation with a perturbation in time. Nevertheless, the methods are straightforward, so we omit the details for the proof.

Proposition 13.7. *Fix $\delta > 0$ and suppose $0 < t_1 < t_2$ are times satisfying*

$$(13.20) \qquad t_1 + t_2 \leqslant C N^{-\delta}$$

for some fixed constant $C > 0$. Suppose V is (g, G)-regular at E_0, and let $0 < q < 1$. Then, uniformly over $z \in \mathbf{D}_{L,q}$, we have

$$(13.21) \qquad |m_{\mathrm{fc}, t_2}(z) - m_{\mathrm{fc}, t_1}(z)| = O\left((t_2 - t_1) \left(\frac{1}{\eta} + \frac{1}{\eta^2} \right) \right).$$

Proof. We briefly note here that the Stieltjes transform solves the following PDE:

$$(13.22) \qquad \partial_t \, m_{\mathrm{fc}, t}(z) = \frac{1}{2} \partial_z \left[m_{\mathrm{fc}, t}(z) \left(m_{\mathrm{fc}, t}(z) + z \right) \right].$$

$\qquad \square$

With Lemma 13.6 and Proposition 13.7, we may now deduce a local law uniformly in time. We first partition the time interval $\mathscr{T}_\omega$ into a set of time intervals with small gaps as follows:

$$(13.23) \qquad \mathscr{T}_\omega = \mathscr{T}_{\omega,1} \cup \ldots \cup \mathscr{T}_{\omega,K} : \quad \sup_{1 \leqslant j \leqslant K} |\mathscr{T}_{\omega,j}| \leqslant N^{-5B_V - 4}, \quad \sup \mathscr{T}_{\omega,i} \leqslant \inf \mathscr{T}_{\omega,i+1}.$$

We note that, if B_V is fixed, that $K = O(1)$. With (ξ, ν)-high probability, by the pointwise (in time) local law in Theorem 9.2, the strong local law holds for the $N^{O(1)}$ set of times $\{\inf \mathscr{T}_{\omega,j}\}$. For notational convenience, we establish the following:

$$(13.24) \qquad t_j := \inf \mathscr{T}_{\omega,j}.$$

On the other hand, by Lemma 13.6 and Proposition 13.7, we have the following estimates for any fixed $j \in [[1, K]]$:

$$\sup_{t \in \mathscr{T}_{\omega,j}} \sup_{z \in \mathbf{D}_{L,q}} |m_N(z; t) - m_{\mathrm{fc}, t}(z)| \leqslant \sup_{t \in \mathscr{T}_{\omega,j}} \sup_{z \in \mathscr{D}_{L,q}} |m_N(z; t) - m_N(z; t_j)|$$

$$+ \sup_{t \in \mathscr{T}_{\omega,j}} \sup_{z \in \mathscr{D}_{L,q}} \left| m_N(z; t_j) - m_{\mathrm{fc}, t_j}(z) \right|$$

$$(13.25) \qquad + \sup_{t \in \mathscr{T}_{\omega,j}} \sup_{z \in \mathscr{D}_{L,q}} \left| m_{\mathrm{fc}, t_j}(z) - m_{\mathrm{fc}, t}(z) \right|$$

$$(13.26) \qquad \leqslant O\left(N^{-4} \left(1 + \frac{1}{\eta} + \frac{1}{\eta^2} + \frac{1}{\eta^3} + \frac{1}{\eta^4} \right) + \frac{N^\varepsilon}{N\eta} \right).$$

This completes the stochastic continuity argument and in particular implies that on the event where the strong local law holds for the times $\{t_j\}$, the strong local law also holds uniformly in time.

IV. Short-Range Approximation and Gap Universality

14. Short-Range Approximation

We now exploit the local structure of the DBM equations by introducing a scheme known as *short-range approximation*. To provide a brief summary, the short-range approximation is composed of the following steps:

- We begin letting the DBM equations run for a short-time t_0 to local equilibrium. At this point, the corresponding Stieltjes transforms of the particles $\{\lambda_i(t_0)\}$ exhibit a strong local law and rigidity phenomena.
- We then introduce cutoffs for the interaction terms. This will approximate the diffusion processes $\mathrm{d}\lambda_i(t)$ by a diffusion process blind to eigenvalues beyond a microscopic distance from $\lambda_i(t)$.

Before we give the construction, we briefly motivate it by introducing a short-range approximation for Wigner matrices. To state this result, we refer to [16] and [17] and introduce the DBM equations for Wigner matrices:

$$(14.1) \qquad \mathrm{d}z_i(t) \ = \ \sqrt{\frac{2}{N}}\mathrm{d}B_{i,\mathrm{W}}(t) \ + \ \frac{1}{N}\sum_{j\neq i}\frac{1}{z_i(t) - z_j(t)} \ \mathrm{d}t.$$

Here, the Brownian motions $B_{i,\mathrm{W}}(t)$ are jointly independent with suitable initial data $z_i(0)$ as will be made precise later. We now introduce the short-range approximation result quite informally. This result is Lemma 3.7 in [16].

Proposition 14.1. *Suppose $z_i(t)$ solves* (14.1) *with initial data either deterministic with a (g, G)-regular potential, or distributed as the eigenvalues of a GOE matrix. Then if $\hat{z}_i(t)$ denotes the solution to the cutoff equation that will be described below for linearized covariance matrices, then the following estimate holds with (ξ, ν)-high probability for constants $\omega_0, \omega_1, \omega_A, \omega_\ell$ and time t_0 to be determined:*

$$(14.2) \qquad \sup_{0\leqslant t\leqslant t_1}\sup_i |\hat{z}_i(t_0 + t) - z_i(t_0 + t)| \ \leqslant \ N^\varepsilon t_1 \left(\frac{N^{\omega_A}}{N^{\omega_0}} + \frac{1}{N^{\omega_\ell}} + \frac{1}{\sqrt{NG}}\right).$$

Before we proceed, we first note the estimate in Proposition 14.1 includes a time-shift by t_0 as to allow for rigidity to hold. We now begin the short-range approximation scheme by introducing the following parameters:

$$(14.3) \qquad 0 \ < \ \omega_1 \ < \ \omega_\ell \ < \ \omega_A \ < \ \frac{\omega_0}{2}.$$

14.1. Time-Shift and Regularization.
We now renormalize the the DBM equations in time. To this end we define two time scales: the natural time scale of the DBM flow and the scale for which we allow the DBM to evolve afterwards.

$$(14.4) \qquad t_0 \ := \ N^{-1+\omega_0}, \quad t_1 \ := \ N^{-1+\omega_1}.$$

This motivates the following time-scale re-shift:

$$(14.5) \qquad \mathrm{d}z_i(t) \ := \ \mathrm{d}\lambda_i(t_0 + t).$$

As alluded to previously, we may now assume the particles $z_i(t)$ exhibit a rigidity phenomenon.

14.2. Short-Range Cutoff.
We now introduce the cutoff in the drift term. To this end, we first define the following which will help regularize the DBM. For a fixed energy E contained in the interior of the support of the free convolution, we define the following classical location minimizing the distance from E:

$$(14.6) \qquad \gamma_E(t) \ := \ \operatorname*{arginf}_{\gamma_{i,t}} |E - \gamma_{i,t}|.$$

We recall here that the infimum is over classical locations of the free convolution law. For notational convenience, we let $k(E)$ denote the index of the above minimizer, so that we have

$$(14.7) \qquad \gamma_E(t) \ = \ \gamma_{k(E),t}.$$

With this, we define the following index set collecting eigenvalues for which we approximate by only short-range interactions:

(14.8) $$\mathscr{I}_{E,\omega_A} := \{j : |j - k(E)| < N^{\omega_A}\}.$$

We also approximate the interval qG on which the initial data is regular by classical locations of the free convolution law:

(14.9) $$\mathscr{C}_q := \{j : \gamma_j \in \mathscr{I}_{E,qG}\}.$$

We now define the short-range cutoff for an individual eigenvalue $z_i(t)$ by collecting nearby eigenvalues via the set

(14.10) $$\mathscr{A}_q := \{(i,j) : |i - j| \leqslant N^{\omega_\ell}\} \bigcup \{(i,j) : ij > 0, \ i, j \notin \mathscr{C}_q\}.$$

We briefly remark that the second set defining $\mathscr{A}_q$ serves to regularize the dynamics for eigenvalues outside the range of regularity for the potential V defined by $\mathscr{I}_{E,qG}$. This will be made precise in the derivation of the short-range approximation.

Before we introduce the short-range cutoffs, we establish the following notation to state the short-range cutoff more conveniently.

Notation 14.2. *For a fixed index i, define the following summation operators:*

(14.11) $$\sum_{j}^{\mathscr{A}_q(i)} := \sum_{j:(i,j)\in\mathscr{A}_q}, \qquad \sum_{j}^{\mathscr{A}_q(i)^C} := \sum_{j:\ (i,j)\in\mathscr{A}_q^C}.$$

Here, the superscript C denotes a set-theoretic complement.

We now apply a deterministic shift to the DBM system as follows:

(14.12) $$\widetilde{z}_i(t) := z_i(t) - \gamma_E(t).$$

This deterministic shift is another regularization operator as it will be important in controlling the main error term in the short-range approximation. We note this shift is both deterministic and the same for all indices, so that local eigenvalue statistics should be preserved. By the inverse function theorem and differentiating the quantile representation of $\gamma_E(t)$, we compute the derivative of $\gamma_E(t)$ as follows:

(14.13) $$\partial_t \gamma_E(t) = -\operatorname{Re}\left(m_{\mathrm{fc},t}(\gamma_E(t))\right) - \frac{1}{2}\gamma_E(t).$$

The RHS in the derivative identity above is understood in the principal value sense; this is where we require E lives in the interior of the support or completely separated from the support of the free convolution law. Thus, by the Ito formula we immediately deduce the following perturbed DBM equations for the shifted eigenvalues:

(14.14) $$\mathrm{d}\widetilde{z}_i(t) = \frac{1}{\sqrt{N}}\mathrm{d}B_i(t) + \left(\frac{1}{2N}\sum_{j\notin\zeta_t(X)}^{(\pm i)}\frac{1}{\widetilde{z}_i(t) - \widetilde{z}_j(t)} + \frac{M-N}{N\widetilde{z}_i(t)} + \operatorname{Re}\left(m_{\mathrm{fc},t}(\gamma_E(t))\right) + \frac{1}{2}\gamma_E(t)\right)\mathrm{d}t.$$

14.3. **Short-Range Equations.** We now define the short-range DBM equations as follows. For those indices $i \in \mathscr{I}_{E,\omega_A}$, we define the following short-range interaction equations:

(14.15) $$\mathrm{d}\hat{z}_i(t) := \frac{1}{\sqrt{N}}\mathrm{d}B_i(t) + \frac{1}{2N}\sum_{j}^{\mathscr{A}_q(i)}\frac{1}{\hat{z}_i(t) - \hat{z}_j(t)}\,\mathrm{d}t.$$

For those indices i outside the interval $\mathscr{I}_{E,\omega_A}$, we define the following dynamics instead:

$$\mathrm{d}\hat{z}_i(t) := \frac{1}{\sqrt{N}}\mathrm{d}B_i(t)$$

(14.16) $$+ \left(\frac{1}{2N}\sum_{j}^{\mathscr{A}_q(i)}\frac{1}{\hat{z}_i(t) - \hat{z}_j(t)} + \frac{1}{2N}\sum_{j}^{\mathscr{A}_q(i)^C}\frac{1}{\widetilde{z}_i(t) - \widetilde{z}_j(t)} + \operatorname{Re}\left(m_{\mathrm{fc},t}(\gamma_E(t))\right) + \frac{1}{2}\gamma_E(t)\right)\mathrm{d}t.$$

Although not explicit, the summations in the drift term avoid the index i. To give initial conditions, for all indices we stipulate

(14.17) $$\hat{z}_i(0) = \widetilde{z}_i(0) = \lambda_i(t_0) - \gamma_E(t_0).$$

We now derive and bound the error term in approximating the true DBM $\widetilde{z}_i(t)$ by the short-range dynamics $\hat{z}_i$. The following result shows that the error term is given by $N^{-1-\delta}$ for small $\delta > 0$. This error is smaller than the scaling for eigenvalue gaps and thus suggests the local eigenvalue statistics for the true DBM and short-range DBM coincide in the limit of large N. We return to this point later, however, and proceed with the short-range approximation.

Proposition 14.3. *In the setting of the short-range approximation, for any fixed $\varepsilon > 0$ we have the following estimate with (ξ, ν)-high probability for sufficiently large $N \gg 1$:*

$$(14.18) \qquad \sup_{t\in[0,t_1]} \sup_i |\hat{z}_i(t) - \widetilde{z}_i(t)| \;\leqslant\; N^\varepsilon t_1 \left(\frac{N^{\omega_A}}{N^{\omega_0}} \;+\; \frac{1}{N^{\omega_\ell}} \;+\; \frac{1}{\sqrt{NG}} \right).$$

Here, we recall G is the regularity parameter of the initial data, i.e. the scale on which rigidity holds.

Before we proceed with the proof of Proposition 14.3, we discuss both the result itself and its consequences. First, we deduce the following consequence which follows from Proposition 14.3 combined with Proposition 14.1. The result also gives a rough idea of the proof of gap universality; we return to this idea shortly.

Corollary 14.4. *With (ξ, ν)-high probability for $N \gg 1$ sufficiently large, we have the following estimate:*

$$(14.19) \qquad \sup_{t\in[0,t_1]} \sup_{i\in\mathscr{I}_{E,\omega_A}} |\lambda_i(t_0 + t) \;-\; \lambda_{W,i}(t_0 + t)| \;<\; N^{-1-\delta},$$

where $\delta > 0$ is a small parameter depending on the parameters $\omega_0, \omega_1, \omega_A, \omega_\ell$.

Indeed, Corollary 14.4 follows from the following eigenvalue SDEs for Wigner matrices of dimension $2N$ along short-time matrix-valued Brownian motion flows:

$$(14.20) \qquad \mathrm{d}\lambda_{W,i}(t) \;=\; \sqrt{\frac{2}{2N}}\mathrm{d}B_i(t) \;+\; \frac{1}{2N} \sum_j^{(i)} \frac{1}{\lambda_{W,i}(t) - \lambda_{W,j}(t)} \;\mathrm{d}t.$$

Here, the equation gives the honest dynamics without the short-range cutoff. Thus, with the respective cutoffs, the eigenvalues $\lambda_{W,i}$ and λ_i solve the same SDEs. Thus, we obtain the estimate (14.19) with the bound on the RHS coming from the short-range approximation bounds in Proposition 14.1 and Proposition 14.3, respectively.

Moreover, in the same spirit we deduce the following reduction of the proof of gap universality, which states that it suffices to prove gap universality for the short-range dynamics.

Corollary 14.5. *In the context of Proposition 14.1 and Proposition 14.3, suppose the gap universality estimate holds for the short-range eigenvalues $\hat{\lambda}$. Then the gap universality estimate holds for the full-range bulk eigenvalues λ.*

We now record the following corollary of Proposition 14.3 which provides a weak form of level repulsion. This is a consequence of comparing the short-range DBM of linearized covariance matrices to the short-range DBM of Wigner matrices, i.e. (14.19), as well as level repulsion estimates for Wigner matrices. Although this estimate is too weak for us to use in this thesis, we record it for possible future use.

Corollary 14.6. *Suppose $\lambda_i(t)$ solves the DBM equations for linearized covariance matrices. Then with (ξ, ν)-high probability, we have the following weak level repulsion for some small $\delta > 0$ and any index i such that the classical location $\gamma_{i,t}$ is bounded away independent of N from the boundary of the support of the free convolution $\varrho_{\mathrm{fc},t}$:*

$$(14.21) \qquad |\lambda_{i+1}(t) - \lambda_i(t)| \;\leqslant\; N^{-1+\varepsilon}.$$

We briefly remark that an optimal level repulsion estimate seems to require analysis of the underlying matrix ensemble and exploiting its Gaussian component, which is the approach taken to obtain optimal level repulsion estimates for Wigner matrices. For a reference concerning level repulsion for Wigner matrices, see [17].

We remark that the upper bound on the RHS of (14.18) is heavily sensitive to the choice of parameters $\omega_0, \omega_1, \omega_A, \omega_\ell$. In particular, Proposition 14.3 is robust in the sense that the parameters are freely adaptable with differences by only small

powers of N. However, the robustness of (14.18) is limited by the quick deterioration of an $N^{-1-\delta}$ estimate into a $\asymp N^{-1+\delta}$ estimate.

Before we provide a proof of Proposition 14.3, we will first make the assumption that $G \asymp 1$ is sufficiently large. In particular, the initial data is very regular, and rigidity holds for all eigenvalues. The proof for weaker regularity on the initial data, although possible, requires a technical and ad hoc argument in applying the strong local law and rigidity estimates. However, focusing on the ensemble of bipartite graphs, the local law in Chapter I of this thesis allows us to assume G is sufficiently large. In other words, with (ξ, ν)-high probability, normalized adjacency matrices of bipartite graphs are (g, G)-regular with $G \asymp 1$ large. For details concerning initial potentials with less regularity, see Lemma 3.3 in [16].

Proof. (of Proposition 14.3).

The key observation is the following dynamics for the difference term $w_i(t) := \hat{z}_i(t) - \tilde{z}_i(t)$:

$$(14.22) \qquad \frac{\mathrm{d}}{\mathrm{d}t} w_i(t) \; = \; \frac{1}{2N} \sum_{j}^{\mathscr{A}_q(i)} \mathscr{B}_{ij}(t) \left(w_j(t) - w_i(t) \right) \; + \; \mathscr{E}_i.$$

Here, the coefficients $\mathscr{B}_{ij}(t)$ and the error terms $\mathscr{E}_i(t)$ are given as follows:

$$(14.23) \qquad \mathscr{B}_{ij}(t) \; = \; \frac{1}{(\hat{z}_i(t) - \hat{z}_j(t))(\tilde{z}_i(t) - \tilde{z}_j(t))},$$

$$(14.24) \qquad \mathscr{E}_i(t) \; = \; \mathbf{1}_{i \in \mathscr{I}_{E, \omega_A}} \left[-\frac{1}{2N} \sum_{j}^{\mathscr{A}_q(i)^C} \frac{1}{\tilde{z}_i(t) - \tilde{z}_j(t)} \; + \; \mathrm{Re} \left(m_{\mathrm{fc}, t}(\gamma_E(t)) \right) \right]$$

Thus, by the Duhamel formula, we have

$$(14.25) \qquad \omega_i(t) \; = \; e^{t\mathscr{B}(t)} \omega_i(0) \; + \; \int_0^t e^{(t-s)\mathscr{B}(t-s)} \mathscr{E}_i(s) \, \mathrm{d}s.$$

Noting the dynamics of the difference terms w_i are given by the dynamics of a jump process, $\mathscr{B}(t)$ denotes the associated semigroup. Because our initial condition $\omega_i(0)$ vanishes by construction, we have

$$(14.26) \qquad \omega_i(t) \; = \; \int_0^t e^{(t-s)\mathscr{B}(t-s)} \mathscr{E}_i(s) \, \mathrm{d}s.$$

Because $\mathscr{B}$ is the generator of a jump process on a discrete state space, it is a contraction on ℓ^∞ of the state space. This implies

$$(14.27) \qquad \|\omega_i(t)\|_{\ell^\infty} \; \leqslant \; \int_0^t \|e^{(t-s)\mathscr{B}(t-s)} \mathscr{E}_i(s)\|_{\ell^\infty} \, \mathrm{d}s \; \leqslant \; t \sup_{s \in [0,t]} \|\mathscr{E}_i(s)\|_{\ell^\infty}.$$

Thus, it remains to obtain an ℓ^∞ estimate on the error terms $\mathscr{E}_i$. To this end, we use the explicit representation of the error terms via the eigenvalue differences. In particular for $i \in \mathscr{I}_{E, \omega_A}$, we first rewrite the error term as follows:

$$(14.28) \qquad \mathscr{E}_i(t) \; = \; \left[-\frac{1}{2N} \sum_{j}^{\mathscr{A}_q(i)^C} \frac{1}{\tilde{z}_i(t) - \tilde{z}_j(t)} \; + \; \int_{\mathscr{I}_{E,t}(i)^C} \frac{\varrho_{\mathrm{fc},t}(x)}{\tilde{z}_i(t) - x} \, \mathrm{d}x \right]$$

$$(14.29) \qquad + \left[\int_{\mathscr{I}_{E,t}(i)^C} \frac{\varrho_{\mathrm{fc},t}(x)}{\tilde{z}_i(t) - x} \, \mathrm{d}x \; - \; \int_{\mathscr{I}_{E,t}(i)^C} \frac{\varrho_{\mathrm{fc},t}(x)}{\gamma_{i,t} - x} \, \mathrm{d}x \right]$$

$$(14.30) \qquad + \left[\mathrm{Re} \left(m_{\mathrm{fc},t}(\gamma_E(t)) \right) \; - \; \mathrm{Re} \left(m_{\mathrm{fc},t}(\gamma_{i,t}) \right) \right] \; + \; \left[\gamma_{i,t} - \gamma_E(t) \right]$$

$$(14.31) \qquad + \int_{\mathscr{I}_{E,t}(i)} \frac{\varrho_{\mathrm{fc},t}(x)}{\gamma_{i,t} - x} \, \mathrm{d}x$$

$$(14.32) \qquad = \; \mathscr{F}_1(t) \; + \; \mathscr{F}_2(t) \; + \; \mathscr{F}_3(t) \; + \; \mathscr{F}_4(t).$$

Here, the integral in defining the error term $\mathscr{F}_4(t)$ is understood in the sense of principal values, as are the Stieltjes transform terms defining the error term $\mathscr{F}_3(t)$. It is now our goal to bound each of the error terms above.

For the first error term, we use the strong local law which holds with (ξ, ν)-high probability for all times:

$$(14.33) \qquad \mathscr{F}_1(t) \;=\; \int_{\mathscr{I}_{E,t}(i)^C} \frac{1}{\widetilde{z}_i(t) - x}\left(\varrho_{\mathrm{fc},t}(x) - \frac{1}{2N}\sum_{\lambda \in \sigma(X(t))} \delta(x - \lambda) \right)\, \mathrm{d}x \;+\; O(N^{-1+\delta})$$

$$(14.34) \qquad \leqslant C \int_{\mathscr{I}_{E,t}(i)^C} \frac{1}{\widetilde{z}_i(t) - x + iN^{\omega_\ell - 1 - \delta}}\left(\varrho_{\mathrm{fc},t}(x) - \frac{1}{2N}\sum_{\lambda \in \sigma(X(t))} \delta(x - \lambda) \right)\, \mathrm{d}x \;+\; O(N^{-1+\delta}).$$

Here, δ is an arbitrarily small but fixed constant. We note the big-Oh term comes from rigidity of eigenvalues whose corresponding classical locations are with $N^{-1+\delta}$ of the boundary of $\mathscr{I}_{E,t}(i)$. The second inequality is taken in the sense of absolute values. This inequality also follows from rigidity, since $\widetilde{z}_i(t)$ is separated from the boundary of $\mathscr{I}_{E,t}(i)$ by a distance bounded below by $N^{\omega_\ell - 1 - \delta}$, upon possibly redefining $\delta > 0$. With this, the strong local law implies, for some $\varepsilon > 0$,

$$(14.35) \qquad \mathscr{F}_1(t) \;\leqslant\; \frac{N^\varepsilon}{N^{\omega_\ell}}.$$

We now estimate the second error term similarly using rigidity:

$$(14.36) \qquad \mathscr{F}_2(t) \;=\; \int_{\mathscr{I}_{E,t}(i)^C} \varrho_{\mathrm{fc},t}(x)\frac{\gamma_{i,t} - \widetilde{z}_i(t)}{(\widetilde{z}_i(t) - x)(\gamma_{i,t} - x)}\, \mathrm{d}x$$

$$(14.37) \qquad \leqslant \frac{N^{1+\varepsilon}}{N^{\omega_\ell}}\int_{\mathscr{I}_{E,t}(i)^C} \varrho_{\mathrm{fc},t}(x)\frac{N^{-1+\varepsilon}}{|\widetilde{z}_i(t) - x| + iN^{\omega_\ell - 1 - \delta}}\, \mathrm{d}x$$

$$(14.38) \qquad \leqslant \frac{N^\varepsilon}{N^{\omega_\ell}}\int_{\mathscr{I}_{E,t}(i)^C} \frac{\varrho_{\mathrm{fc},t}(x)}{|\widetilde{z}_i(t) - x| + iN^{\omega_\ell - 1 - \delta}}\, \mathrm{d}x.$$

To estimate this last term, because $\varrho_{\mathrm{fc},t}$ has compact support, it suffices to estimate the Stieltjes transform of the free convolution for large energy. To this end we use a general result for Stieltjes transforms of compactly supported measures:

$$(14.39) \qquad |m_{\mathrm{fc},t}(E + i\eta)| \;\leqslant\; C\log(N)^C \sup_{\eta' \geqslant \eta} \mathrm{Im}\left(m_{\mathrm{fc},t}(E + i\eta') \right).$$

Here, $C > 0$ is a fixed constant. For a proof of this result, we refer to Lemma 7.1 in [17].

By the fixed-point equation defining the free convolution, the RHS is bounded by $C\log(N)^C$. Thus, we deduce the following bound on the second error term for some possibly adapted constant $\varepsilon > 0$:

$$(14.40) \qquad |\mathscr{F}_2(t)| \;\leqslant\; C\frac{N^\varepsilon}{N^{\omega_\ell}}.$$

We now estimate the third error term with the following bound for an arbitrarily small but fixed $\delta > 0$:

$$(14.41) \qquad |\mathscr{F}_3(t)| \;\leqslant\; C\left(\frac{N^{\omega_A}}{N^{\omega_0}} + \frac{N^{\omega_A}}{N} \right).$$

This follows from bounding the following time-derivative of the Stieltjes transform given in Lemma 3.3 in [16]:

$$(14.42) \qquad \left| \frac{\partial}{\partial z} m_{\mathrm{fc},t_0 + t}(z) \right| \;\leqslant\; \frac{C}{t_0} \;\asymp\; \frac{C}{N^{-1+\omega_0}}.$$

Indeed, by construction, we know $|\gamma_{i,t} - \gamma_E(t)|$ is bounded by $N^{\omega_A - 1}$ up to a constant independent of N. We note this estimate does not require rigidity as there is no randomness in the classical locations of the free convolution. In any case, this gives the desired estimate:

$$(14.43) \qquad |\mathscr{F}_3(t)| \;\leqslant\; C\left(\frac{N^{\omega_A - 1}}{N^{\omega_0 - 1}} + \frac{N^{\omega_A}}{N} \right).$$

It remains to bound the last error term. To this end, we first note the interval $\mathscr{I}_{E,t}(i)$ is almost symmetric by definition. To elaborate, if $\mathscr{I}_{E,t}(i)$ were symmetric about $\gamma_i(t)$, then the error term $\mathscr{F}_4(t)$ would be bounded by the C^1-norm of the free

convolution law up to a constant by the classical regularization procedure of the principal value. However, we instead have the estimate

$$
(14.44) \qquad |\mathscr{F}_4(t)| = \left| \int_{\mathscr{I}_{E,t}(i)} \frac{\varrho_{\mathrm{fc},t}(x)}{\gamma_{i,t} - x} \, \mathrm{d}x \right| \leqslant C \|\nabla \varrho_{\mathrm{fc},t}\|_{C^0} + C \frac{N^{\omega_\ell}}{N t_0}.
$$

Indeed, the first term comes from regularizing the integral on some small symmetric interval centered at $\gamma_{i,t}$ contained in $\mathscr{I}_{E,t}(i)$, and the second term follows from a simple bound using the following estimate on classical locations which in turn follows from the derivative estimate (14.42):

$$
(14.45) \qquad \gamma_{i+k,t_0+t} - \gamma_{i,t_0+t} = (\gamma_{i-k,t_0+t} - \gamma_{i,t_0+t}) \left(1 + O\left(\frac{k}{N t_0} \right) \right).
$$

We may also bound the derivative norm by applying the inverse Stieltjes transform to (14.42). This implies

$$
(14.46) \qquad |\mathscr{F}_4(t)| \leqslant C \frac{N^{\omega_\ell}}{N^{\omega_0}}.
$$

Combining the bounds for all the error terms, we immediately deduce the desired estimate (14.18). $\qquad\square$

Before we proceed, we note our above estimates did not produce a term on the order of $(NG)^{-1/2}$. This is because we assumed high regularity, i.e. $G \asymp 1$, so a direct application of the strong local law and rigidity cannot see this term. We note, however, this missing error term appears naturally in handling the error term $\mathscr{F}_1$ assuming little regularity on the initial data. For details we again refer to Section 3 in [16].

15. Gap Universality: Proof of the Main Theorem

We now use the short-range approximation in Proposition 14.3 to derive gap universality. The method here is taken from the proof of gap universality for the Wigner ensemble given in [17]. We begin by introducing the coupled Dyson Brownian Motions matching the process $\hat{\lambda}_i(t)$ with GOE eigenvalue dynamics. To this end, recall that $\hat{\lambda}_i(t)$ denotes the solutions to the following short-range DBM equations:

$$
(15.1) \qquad \mathrm{d}\hat{\lambda}_i(t_0 + t) = \frac{1}{\sqrt{N}} \mathrm{d}B_i(t_0 + t) + \frac{1}{2N} \sum_{j}^{\mathscr{A}_q(i)} \frac{1}{\hat{\lambda}_i(t_0 + t) - \hat{\lambda}_j(t_0 + t)} \mathrm{d}t.
$$

We emphasize the short-range interactions in the drift term contain at most one contribution from each pair of totally anti-correlated eigenvalues $(\pm \lambda_i)$ in the spectrum of $X(t)$.

15.1. **The Coupled Gaussian Dynamics.** We now fix an energy E in the bulk of the linearized Marchenko-Pastur law and an index $i \in \mathscr{I}_{E,\omega_A}$. We now define the following matrix:

$$
(15.2) \qquad W(t_0) = a_0 W + b_0, \quad a_0 = \frac{\varrho_{\mathrm{sc}}(\mu_i)}{\varrho_{\mathrm{fc},t_0}(\gamma_{i,t_0})}, \quad b_0 = \gamma_{i,t_0} - a_0 \mu_i.
$$

We now define the DBM flow for the eigenvalues of a GOE matrix with initial data given by the eigenvalues of $W(t_0)$. We first define the short-range DBM denoted by $\hat{\nu}(t)$ given by the following system of SDEs for $i \in \mathscr{I}_{E,\omega_A}$:

$$
(15.3) \qquad \mathrm{d}\hat{\nu}_i(t_0 + t) = \frac{1}{\sqrt{N}} \mathrm{d}B_i(t_0 + t) + \frac{1}{2N} \sum_{j}^{\mathscr{A}_q(i)} \frac{1}{\hat{\nu}_i(t_0 + t) - \hat{\nu}_j(t_0 + t)} \mathrm{d}t.
$$

Here, the Brownian motions defining (15.3) are the same Brownian motions driving the short-range process $\hat{\lambda}(t)$.

We now define the full-range DBM with initial data is the full spectrum of $W_{a,b}(t_0)$. We will not study the analysis of the full-range DBM in detail, as it will only be necessary in knowing the global structure of the eigenvalue system. We define this full-range DBM though the following SDEs:

$$
(15.4) \qquad \mathrm{d}\nu_i(t_0 + t) = \frac{1}{\sqrt{N}} \mathrm{d}B_i(t_0 + t) + \frac{1}{2N} \sum_{j}^{(i)} \frac{1}{\nu_i(t_0 + t) - \nu_j(t_0 + t)} \mathrm{d}t.
$$

Because the Gaussian data is invariant under the full-range DBM equations above, we deduce the system $\nu(t)$ gives the spectrum of a sub-linear Gaussian perturbation of $W(t_0)$. This may be realized as processes a_t and b_t with initial data a_0 and b_0 defined above; this is the desired global data mentioned in introducing the full-range DBM.

15.2. **Gap Universality.** We now use the coupled short-range GOE dynamics $\hat{\nu}(t)$ to deduce gap universality for the linearized covariance matrix ensemble. By Proposition 14.3 we have the following raw gap estimates for the short-range dynamics with (ξ, ν)-high probability:

$$(15.5) \qquad \sup_{t \in [0, t_1]} \left| \left(\hat{\lambda}_i(t_0 + t) - \hat{\lambda}_{i+k}(t_0 + t) \right) - (\hat{\nu}_i(t_0 + t) - \hat{\nu}_{i+k}(t_0 + t)) \right| < N^{-1-\varepsilon}.$$

Here, $\varepsilon > 0$ is a small, fixed constant. Because $\varrho_{\mathrm{fc}, t_0} \asymp 1$ in the bulk, the same gap estimate holds multiplying the LHS by $\varrho_{\mathrm{fc}, t_0}(\gamma_{i, t_0})$, upon possibly adjusting the constant $\varepsilon > 0$. This gives the following preliminary estimate upon a possible readjustment of the parameter $\varepsilon > 0$:

$$(15.6) \qquad \sup_{t \in [0, t_1]} \left| \varrho_{\mathrm{fc}, t_0}(\gamma_{i, t_0}) \left(\hat{\lambda}_i(t_0 + t) - \hat{\lambda}_{i+k}(t_0 + t) \right) - \varrho_{\mathrm{fc}, t_0}(\gamma_{i, t_0}) (\hat{\nu}_i(t_0 + t) - \hat{\nu}_{i+k}(t_0 + t)) \right| < N^{-1-\varepsilon}.$$

This, although close, is not the desired bound from which we may deduce gap universality from a first-order Taylor expansion. Through perturbative methods, however, we may deduce the desired bound from (15.6); in particular we aim to perturb the free convolution factors and the classical locations in time, which requires a time- and energy- derivative estimate on the free convolution. We make this more precise shortly.

Lemma 15.1. *In the context of the short-range dynamics $\hat{\lambda}(t)$ and $\hat{\nu}(t)$, we have*

$$(15.7) \qquad \sup_{t \in [0, t_1]} \left| \varrho_{\mathrm{fc}, t}(\gamma_{i, t}) \left(\hat{\lambda}_i(t_0 + t) - \hat{\lambda}_{i+k}(t_0 + t) \right) - \frac{\varrho_{\mathrm{sc}}(\mu_i)}{a_t} (\hat{\nu}_i(t_0 + t) - \hat{\nu}_{i+k}(t_0 + t)) \right| < N^{-1-\varepsilon}.$$

Proof. To replace this free convolution factor at t_0 by the free convolution at time $t_0 + t$, we appeal to the following simple estimate:

$$(15.8) \qquad |\varrho_{\mathrm{fc}, t}(\gamma_{i, t}) - \varrho_{\mathrm{fc}, t_0}(\gamma_{i, t_0})| \leqslant |\varrho_{\mathrm{fc}, t}(\gamma_{i, t}) - \varrho_{\mathrm{fc}, t}(\gamma_{i, t_0})| + |\varrho_{\mathrm{fc}, t}(\gamma_{i, t_0}) - \varrho_{\mathrm{fc}, t_0}(\gamma_{i, t_0})|$$

$$(15.9) \qquad \qquad \leqslant C \frac{N^{\omega_1} \log N}{N^{\omega_0}} + |\varrho_{\mathrm{fc}, t}(\gamma_{i, t_0}) - \varrho_{\mathrm{fc}, t_0}(\gamma_{i, t_0})|.$$

Here, the second inequality follows from an energy-derivative estimate on the free convolution density given by applying the Stieltjes inversion formula to (14.42) and bounding the time derivative of the classical location computed in (14.13) by $\log N$. It remains to bound the second term in the inequality above. We do so by bounding the time-derivative of the free convolution density. Indeed, we may obtain the following time-derivative bound by computing the time-derivative of its Stieltjes transform via its fixed-point equation and applying the Stieltjes inversion formula. This method yields the following estimate; for a reference, see Lemma 7.6 in [17]:

$$(15.10) \qquad |\partial_t \varrho_{\mathrm{fc}, t_0}(E)| \leqslant \frac{C}{t_0} = \frac{C}{N^{-1+\omega_0}}.$$

Because $t - t_0 \leqslant t_1 = N^{-1+\omega_1}$, we ultimately deduce

$$(15.11) \qquad |\varrho_{\mathrm{fc}, t}(\gamma_{i, t}) - \varrho_{\mathrm{fc}, t_0}(\gamma_{i, t_0})| \leqslant C \frac{N^{\omega_1}}{N^{\omega_0}} (\log N + 1) \leqslant \frac{1}{N^{\omega_2}},$$

for some fixed constant $\omega_2 > 0$. Combining our estimates thus far, we have the following bound with (ξ, ν)-high probability uniformly over all times $t \in [t_0, t_0 + t_1]$:

$$(15.12) \qquad \left| \varrho_{\mathrm{fc}, t}(\gamma_{i, t}) (\lambda_i - \lambda_{i+k}) - \frac{\varrho_{\mathrm{sc}}(\mu_i)}{a_0} (\nu_i - \nu_{i+k}) \right| \leqslant N^{-1-\varepsilon},$$

upon possibly redefining $\varepsilon > 0$. Here, we keep the $o(N^{-1})$ estimate on the RHS by appealing to the following bound which holds by rigidity:

$$(15.13) \qquad |\varrho_{\mathrm{fc}, t}(\gamma_{i, t}) - \varrho_{\mathrm{fc}, t_0}(\gamma_{i, t_0})| (\lambda_i - \lambda_{i+k}) \leqslant N^{-\omega_2} N^{-1+\delta} \leqslant N^{-1-\varepsilon}.$$

Here, $\delta > 0$ is arbitrarily small and $\varepsilon > 0$ is fixed. However, because eigenvalues differences change at most linearly upon linear perturbations of the underlying matrix, we deduce the simple bound $|a_t - a_0| \leqslant Ct$. This implies the following time-adapted bound on the event on which the upper bound (15.12) holds:

$$(15.14) \qquad \left| \varrho_{\mathrm{fc},t}(\gamma_{i,t}) \left(\lambda_i - \lambda_{i+k} \right) - \frac{\varrho_{\mathrm{sc}}(\mu_i)}{a_t} \left(\nu_i - \nu_{i+k} \right) \right| \leqslant N^{-1-\varepsilon}.$$

This follows from the constraint that we stipulate $t \leqslant t_1$, which concludes the proof. $\qquad \square$

We now make the following observation; by construction the gaps $a_t^{-1}(\nu_i - \nu_{i+k})$ are distributed as the gaps of a standard GOE matrix. Moreover, by applying the proof of Lemma 15.1 iteratively n times for any $n = O(1)$, we also deduce the following multi-gap estimate with (ξ, ν)-high probability:

$$(15.15) \qquad \sup_{k \in [0,n]} \sup_{t \in [0,t_1]} \left| \varrho_{\mathrm{fc},t}(\gamma_{i,t}) \left(\hat{\lambda}_i(t_0 + t) - \hat{\lambda}_{i+k}(t_0 + t) \right) - \frac{\varrho_{\mathrm{sc}}(\mu_i)}{a_t} \left(\hat{\nu}_i(t_0 + t) - \hat{\nu}_{i+k}(t_0 + t) \right) \right| < N^{-1+\varepsilon}.$$

From here, to deduce gap universality we Taylor expand our test function $O \in C_c^\infty(\mathbb{R}^n)$ on the event on which (15.15) holds. In particular, by a first-order Taylor expansion on this event we deduce the following straightforward estimate:

$$(15.16) \qquad \begin{aligned} & O\left(N\varrho_{\mathrm{fc},t}(\gamma_{i,t})(\lambda_i - \lambda_{i+i_1}), \ldots, N\varrho_{\mathrm{fc},t}(\gamma_{i,t})(\lambda_i - \lambda_{i+i_n}) \right) \\ & \quad - O\left(N\varrho_{\mathrm{sc}}(\mu_i)(\lambda_i - \lambda_{i+i_1}), \ldots, N\varrho_{\mathrm{sc}}(\mu_i)(\lambda_i - \lambda_{i+i_n}) \right) \lesssim_{\|O\|_{C^1},n} N^{-\varepsilon}, \end{aligned}$$

where $\varepsilon > 0$ is the same exponent in the estimate (15.14).

On the complement of the event on which (15.14) holds, we may apply the following crude bound which follows from the spectral bound on the initial data V:

$$(15.17) \qquad \begin{aligned} & O\left(N\varrho_{\mathrm{fc},t}(\gamma_{i,t})(\lambda_i - \lambda_{i+i_1}), \ldots, N\varrho_{\mathrm{fc},t}(\gamma_{i,t})(\lambda_i - \lambda_{i+i_n}) \right) \\ & \quad - O\left(N\varrho_{\mathrm{sc}}(\mu_i)(\lambda_i - \lambda_{i+i_1}), \ldots, N\varrho_{\mathrm{sc}}(\mu_i)(\lambda_i - \lambda_{i+i_n}) \right) \lesssim_{\|O\|_{\mathscr{S}},n} N^C \end{aligned}$$

for any $C = O(1)$. However, because (15.14) holds with (ξ, ν)-high probability, by taking expectations we deduce

$$(15.18) \qquad \begin{aligned} & \mathbb{E}^{X(t)} \left[O\left(N\varrho_{\mathrm{fc},t}(\gamma_{i,t})(\lambda_i - \lambda_{i+i_1}), \ldots, N\varrho_{\mathrm{fc},t}(\gamma_{i,t})(\lambda_i - \lambda_{i+i_n}) \right) \right] \\ & \quad - \mathbb{E}^{\mathrm{GOE}} \left[O\left(N\varrho_{\mathrm{sc}}(\mu_i)(\lambda_i - \lambda_{i+i_1}), \ldots, N\varrho_{\mathrm{sc}}(\mu_i)(\lambda_i - \lambda_{i+i_n}) \right) \right] \leqslant N^{-\varepsilon} + N^C e^{-\nu \log^\xi N}. \end{aligned}$$

Because $\xi \gg 1$, appealing to Corollary 14.5, this completes the proof of Theorem 3.1.

Bibliography

References

[1] Adlam, Ben. *The Local Marchenko-Pastur Law for Sparse Covariance Matrices.* Senior Thesis, Harvard University Department of Mathematics (2013).

[2] Anderson, Greg W., Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices. Cambridge.* Cambridge UP, 2010. Print.

[3] Bauerschmidt, Roland and Jiaoyang Huang, Antti Knowles, and H.T. Yau. "Bulk eigenvalue statistics for random regular graphs." arXiv:1505.06700 [math.PR]. Submitted 25 May, 2015.

[4] Bauerschmidt, Roland and Antti Knowles, and H.T. Yau. "Local semicircle law for random regular graphs." arXiv:1503.08702. Submitted 30 Mar, 2015.

[5] Biane, Philippe. "On the free convolution with a semi-circular distribution." *Indiana University Mathematics Journal.* 46(3):705718, 1997.

[6] Bourgade, Paul and H.T. Yau. "The eigenvector moment flow and local quantum unique ergodicity." *Communications in Mathematical Physics*, published online 30 April, 2016, DOI 10.1007/s00220-016-2627-6.

[7] Bru, Marie-France. "Diffusions of perturbed principal component analysis." *Journal of Multivariate Analysis*, 1989. DOI: 10.1016/0047-259X(89)90080-8.

[8] Dumitriu, Ioana and T. Johnson. "The Marchenko-Pastur law for sparse random bipartite biregular graphs." *Random Structures and Algorithms*, published online 12/2014, DOI 10.1002/rsa.20581.

[9] Dyson, Freeman. "A Brownian-Motion Model for the Eigenvalues of a Random Matrix." *Journal of Mathematical Physics*, **3**, 1191 (1962)

[10] Dyson, Freeman. "Correlations between eigenvalues of a random matrix". *Communications in Mathematical Physics*, 19:235250, 1970

[11] Erdos, Laszlo and Benjamin Schlein and H.-T. Yau. "Universality of Random Matrices and Local Relaxation Flow". *Inventiones Mathematicae.* (2011). 185: 75. DOI:10.1007/s00222-010-0302-7

[12] Erdos, Laszlo and Benjamin Schlein, H.T. Yau, and Jun Yin. "The local relaxation flow approach to universality of the local statistics for random matrices". *Annals of the Institute of H. Poincare in Probability and Statistics.* Volume 48, Number 1 (2012), 1-46.

[13] Erdos, Laszlo, and H.-T. Yau and J. Yin. "Bulk Universality for generalized Wigner matrices". *Probability and Related Fields*, 154(1-2):341-407, 2012.

[14] Gaudin, Michel and Madan Lal Mehta. "On the density of eigenvalues of a random matrix." *Nuclear Phys. B*, 18:420427, 1960

[15] Huang, Jiaoyang and Benjamin Landon and H.T. Yau. "Bulk Universality of Sparse Random Matrices". *Journal of Mathematical Physics.* (2015).

[16] Landon, Benjamin and H.T. Yau. "Convergence of local statistics of Dyson Brownian motion". *Communications in Mathematical Physics*, to appear (2015).

[17] Landon, Benjamin, Phillippe Sosoe, and H.T. Yau. "Fixed energy universality of Dyson Brownian motion". arXiv:1609.09011 [math.PR]. September 28, 2016.

[18] Lee, Ji Oon and K. Schnelli. "Local deformed semicircle law and complete delocalization for Wigner matrices with random potential". *Journal of Mathematical Physics.* 54(10):103504, 2013.

[19] Mehta, Madan Lal. "A note on correlations between eigenvalues of a random matrix." *Communications in Mathematical Physics*, 20:245250, 1971

[20] Madan, Mehta Lal. *Random matrices.* Academic Press, Inc., Boston, MA, second edition, 1991.

[21] Pillai, Natesh S, and Jun Yin "Edge universality of correlation matrices." *Annals of Statistics.* 40 (2012), no. 3, 1737–1763. DOI" 10.1214/12-AOS1022.

[22] Pillai, Natesh S., and Jun Yin. "Universality of covariance matrices." *Annals of Applied Probability*, 24 (2014), no. 3, 935–1001. DOI: 10.1214/13-AAP939.

[23] Wigner, Eugene. "Characteristic vectors of bordered matrices with infinite dimensions." *Annals of Mathematics*, 62:548?564, 1955.

[24] Wishart, John. "Generalized product moment distribution in samples." *Biometrika*, 20A (12): 3252. (1928)

Notations

(1) $[[a, b]], \quad a, b \in \mathbb{R}$:

Discretization of the interval $[a, b]$: $[a, b] \cap \mathbb{Z}$.

(2) $[E], \quad E$ a graph:

The vertex set of the graph E.

(3) $vv', \quad v, v' \in [E]$:

The edge determined by vertices v, v'.

(4) $\mathbf{M}_{M \times N}(\mathbb{R})$:

The space of $M \times N$ matrices with real entries.

(5) $\alpha = \gamma^{-1}$:

The limiting dimension ratio $\lim\limits_{N \to \infty} \frac{M}{N}$.

(6) $\sigma(X), \quad X$ a matrix:

The spectrum of the matrix X.

(7) $\mathbb{P}(\Xi), \quad \Xi$: an event of a probability space.

The probability of the event Ξ.

(8) $\mathbb{P}_{\mathscr{F}}(\Xi), \quad \mathscr{F}$ and σ-algebra and Ξ an event:

The conditional probability of Ξ with respect to $\mathscr{F}$.

(9) $\mathbb{E}(X), \quad X$ a random variable:

The expectation of the random variable X.

(10) $\mathbb{E}_{\mathscr{F}}(X), \quad \mathscr{F}$ a σ-algebra and X a random variable:

The conditional expectation of X with respect to $\mathscr{F}$.

(11) $\partial_{ij} F(X), \quad X = (X_{ij})$ a matrix:

The partial derivative of F with respect to the entry X_{ij}.