Finite Rank Perturbations of Random Matrices and Their Continuum Limits

by

Alexander Bloemendal

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Mathematics
University of Toronto

Abstract

Finite Rank Perturbations of Random Matrices
and Their Continuum Limits

Alexander Bloemendal
Doctor of Philosophy
Graduate Department of Mathematics
University of Toronto
2011

We study Gaussian sample covariance matrices with population covariance a bounded-rank perturbation of the identity, as well as Wigner matrices with bounded-rank additive perturbations. The top eigenvalues are known to exhibit a phase transition in the large size limit: with weak perturbations they follow Tracy-Widom statistics as in the unperturbed case, while above a threshold there are outliers with independent Gaussian fluctuations. Baik, Ben Arous and Péché (2005) described the transition in the complex case and conjectured a similar picture in the real case, the latter of most relevance to high-dimensional data analysis.

Resolving the conjecture, we prove that in all cases the top eigenvalues have a limit near the phase transition. Our starting point is the work of Rámirez, Rider and Virág (2006) on the general beta random matrix soft edge. For rank one perturbations, a modified tridiagonal form converges to the known random Schrödinger operator on the half-line but with a boundary condition that depends on the perturbation. For general finite-rank perturbations we develop a new band form; it converges to a limiting operator with matrix-valued potential. The low-lying eigenvalues describe the limit, jointly as the perturbation varies in a fixed subspace. Their laws are also characterized in terms of a diffusion related to Dyson's Brownian motion and in terms of a linear parabolic PDE.

We offer a related heuristic for the supercritical behaviour and rigorously treat the

supercritical asymptotics of the ground state of the limiting operator.

In a further development, we use the PDE to make the first explicit connection between a general beta characterization and the celebrated Painlevé representations of Tracy and Widom (1993, 1996). In particular, for $\beta = 2, 4$ we give novel proofs of the latter.

Finally, we report briefly on evidence suggesting that the PDE provides a stable, even efficient method for numerical evaluation of the Tracy-Widom distributions, their general beta analogues and the deformations discussed and introduced here.

This thesis is based in part on work to be published jointly with Bálint Virág.

Dedication

To my teachers, starting with my parents

Acknowledgements

Without my wonderful thesis advisor Bálint Virág, I would never have done any of this work. Learning from and working with Bálint has been a better experience than anyone could hope for and I will say more about it presently.

I was fortunate to have my graduate work supported by an NSERC postgraduate scholarship as well as departmental research assistantships. I also gained a great deal from attending summer schools at PIMS and MSRI and workshops at CRM, MSRI and AIM. I am further grateful to Alexei Borodin and the department at Caltech as well as Bálint Tóth and the institute at the Technical University of Budapest for their hospitality in the Spring and Fall of 2010.

Special thanks are due to Percy Deift, both for his careful and helpful comments on this thesis and for making the trip to Toronto for my defense! I am also very appreciative of Peter Forrester for his continued interest and encouragement, Alexander Its for his patient and thorough explanations and Benedek Valko for discussions of drafts. I have benefitted from interesting conversations with many others including Mark Adler, Jinho Baik, Alexei Borodin, Iain Johnstone, Arno Kuijlaars, Pierre van Moerbeke, Eric Rains, Brian Rider, Brian Sutton, Craig Tracy, Dong Wang, Harold Widom and Ofer Zeitouni. Credit for the initial conception of certain ideas is shared with José Ramírez.

My professors at U of T have generously shared not only mathematics but guidance over the years. I am especially grateful to Ed Bierstone, Jim Colliander, Michael Goldstein, Kostya Khanin, Robert McCann, Mary Pugh, Joe Repka and Jeremy Quastel.

I am very happy to thank the departmental staff, but it is not possible to thank Ida Bulat enough. I cannot remember all the times she went out of her way; I doubt I know anyone else as patient, kind, dedicated or honestly as good at what they do.

Nor is it really possible to convey my appreciation for my family and friends. I am very lucky to have these wonderful people in my life. To those who were closest to me through the past few years: your company, support, empathy and humour helped me more than you know. I couldn't have done it without you.

Finally, I am forever indebted to Bálint for things that went far beyond the call of duty: for inspiring me out of my initial hesitation; for showing me so much beautiful mathematics; for teaching me the value of thinking simply and expressing each thought directly; for patiently encouraging me, always putting up with me, and never losing faith even when I did. A true teacher, mentor, role model and friend, words do not express the depth of my admiration and gratitude.

Contents

Chapter 1

Introduction

This introductory chapter briefly reviews some background material, sketches the contributions of the thesis and provides an outline of subsequent chapters. Precise definitions and statements as well as many additional references will be found in the opening sections of the chapters, especially Chapters 2 and 3.

1.1 Background

Random matrices

Random matrix theory is predominantly concerned with the asymptotic spectral properties of large matrices with random entries jointly distributed according to one of several natural models. Most classical are the Gaussian orthogonal, unitary and symplectic ensembles (GO/U/SE), introduced to physics by Wigner and Dyson in the 1950s and 60s. These distributions on real symmetric, complex Hermitian or quaternion self-dual matrices have density proportional to $e^{-\frac{\beta}{4}\operatorname{tr} A^2}$ where $\beta = 1, 2, 4$ respectively; they are unique in having the double virtue of invariance under the associated classical group of symmetries as well as independence of the entries up to the algebraic constraints. They may also be described by filling a matrix X with independent standard real/complex/quaternion Gaussians and additively symmetrizing: $A = \frac{1}{\sqrt{2}}(X + X^\dagger)$. General references include Mehta (2004), Deift (1999), Anderson, Guionnet and Zeitouni (2009), Forrester (2010).

Already in the late 1920s, Wishart considered Gaussian sample covariance matrices. Here one begins with a "data matrix" X of independent Gaussian columns and the symmetrization is multiplicative: $A = XX^\dagger$. Wishart matrices continue to be of central

importance in multivariate statistics; see Muirhead (1982), Bai (1999), Anderson (2003).

There are several asymptotic regimes. At the global scale one may study the empirical spectral distribution and find the famous Wigner semicircle and Marčenko-Pastur laws in the Gaussian and Wishart cases respectively. The semicircle law states that, if $\lambda_1, \dots, \lambda_n$ are the eigenvalues of an $n \times n$ GO/U/SE matrix, then with probability one there is the weak convergence

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i/\sqrt{n}} \;\to\; \mu \qquad \text{where} \quad \frac{d\mu}{dx} = \frac{\sqrt{4-x^2}}{2\pi}\, \mathbf{1}_{[-2,2]}(x).$$

This law holds for much more general self-adjoint matrices with independent entries, known as Wigner matrices.

Wigner and Dyson were originally intersted in eigenvalue spacing in the bulk of the spectrum as a model for the excitation spectra of heavy nuclei. The setting of this thesis is the point process formed by the largest eigenvalues, also referred to as the "soft edge" of the spectrum. The fundamental limit law for the $n \times n$ GO/U/SE is due to Tracy and Widom (1993, 1994, 1996); for the largest eigenvalue λ_1 it states that

$$\mathbf{P}_n\Big(n^{1/6}\big(\lambda_1 - 2\sqrt{n}\big) \le x \Big) \to F_\beta(x),$$

again with $\beta = 1, 2, 4$ respectively, where F_β are the celebrated Tracy-Widom distributions. There are striking explicit representations, for example

$$F_2(x) = \exp\left(-\int_x^\infty (s-x)u^2(s)\, ds \right) \tag{1.1}$$

where u is the so-called Hastings-McLeod solution of the Painlevé II equation, a "nonlinear special function" determined by

$$\begin{aligned} u'' &= 2u^3 + xu, \\ u(x) &\sim \mathrm{Ai}(x) \qquad \text{as } x \to +\infty. \end{aligned} \tag{1.2}$$

Universality is the recurring theme: one broadly expects the same local asymptotics irrespective of the details of the model beyond the symmetry class β. There are now large bodies of rigorous results in this direction for broad classes of random matrices retaining either one of the two salient features of the Gaussian ensembles. We give just two references, namely Deift (2007), Erdős and Yau (2011).

Most remarkable, however, is that the relevance of the local limit laws extends to diverse parts of mathematics and physics. In the case of the Tracy-Widom laws the

seminal result is that of Baik, Deift and Johansson (1999) on the longest increasing subsequence of a random permutation, which was followed by work of Johansson (2000) on a model of interface growth in two dimensions. These discoveries have since been reinforced by a decade of intense mathematical activity (Tracy and Widom 2002, Deift 2007). In an exciting recent development, experimental results on interface fluctuations using liquid crystals confirm Tracy-Widom statistics with astonishing accuracy (Takeuchi and Sano 2010, Takeuchi, Sano, Sasamoto and Spohn 2011).

The spiked model and the BBP transition

Multivariate statistical analysis traditionally assumed a fixed sample dimensionality p and a relatively large sample size n. Modern problems typically feature high dimensionality, where p is on the same order as n or perhaps even larger; traditional techniques of covariance estimation then break down and subtle new phenomena arise. In detail, consider a $p \times n$ "data matrix" X with n independent $N_p(0, \Sigma)$ columns where the $p \times p$ population covariance matrix $\Sigma > 0$; then form the $p \times p$ Wishart matrix $XX^\dagger$. Eigenvalues of the sample covariance matrix $\frac{1}{n}XX^\dagger$ no longer consistently estimate those of Σ; even in the so-called null case $\Sigma = I$, if $p \sim cn$ with $0 < c < \infty$ the sample eigenvalues spread out over an interval as described by the Marčenko-Pastur law.

Johansson (2000) and Johnstone (2001) proved respectively GUE/GOE Tracy-Widom fluctuations for the largest eigenvalues of complex/real null Wishart matrices when n, p are both large. Johnstone pointed out the relevance to statistical analysis and called for an understanding of the much more general non-null case. Motivated by prevalent techniques like principal components analysis as well as observed behaviour in real data sets, he proposed the following "spiked population model": fix a finite "rank" r and let Σ have r non-null population eigenvalues $\ell_1, \ldots, \ell_r$ (the spikes) with the rest set to 1. The relevant question is the large n, p behaviour of the top sample covariance eigenvalues.

Baik, Ben Arous and Péché (2005) (**BBP**) gave a fairly complete analysis of the *complex* spiked Wishart model and discovered a fascinating phase transition. If the ℓ_i all remain some positive distance below $1 + \sqrt{p/n}$ then the largest sample eigenvalues behave exactly as in the null case, developing GUE Tracy-Widom fluctuations about the right endpoint of Marčenko-Pastur. The statistical relevance is clear: in this "subcritical regime" one cannot hope to observe the "signal" amidst the high-dimensional "noise". If several ℓ_i exceed $1 + \sqrt{p/n}$, the same number of sample eigenvalues will separate from

the rest and develop larger Gaussian fluctuations around their outlying limits; this is the "supercritical regime". The transition itself occurs on the scale $\ell_i - (1+\sqrt{p/n}) \sim cn^{-1/3}$, and in this "critical regime" one finds subcritical-like behaviour but with new fluctuation laws: multi-parameter families of distributions that deform F_2.

This phenomenon is now often referred to as the BBP transition and has been widely cited. Applications include population genetics, economics and statistical learning (Johnstone 2007, Paul 2007), but these generally involve real-valued data. BBP conjectured a similar transition for real spiked Wishart matrices: all the same scalings, but now with F_1 subcritical fluctuations (and presumably some deformations in the critical regime). Their techniques, however, do not go over to the real case. They begin with the joint eigenvalue density, making essential use of the Harish-Chandra-Itzykson-Zuber formula to integrate out over the unitary group and arrive at a determinantal form for the largest eigenvalue distribution that can be analyzed.

A partial description in the real case was obtained by various other methods: Baik and Silverstein (2006) found the anticipated behaviour on the level of a.s. limits (generalized by Benaych-Georges and Nadakuditi 2009 in work related to free probability), Paul (2007), Bai and Yao (2008) confirmed supercritical Gaussian fluctuations, and Féral and Péché (2009) proved F_1 limits when the spikes are well-separated from the critical point from below. The two latter works also show that the phase transition is universal in the sense that one finds precisely the same behaviour for a large class of non-Gaussian sample covariance matrices.

Analogous results for finite-rank additive perturbations of the GUE were found by Péché (2006). One finds exactly the same transition in the additive setting, which is strong evidence for universality of the effect of finite-rank perturbations on the random matrix soft edge. Once again, the real (GOE) case has proved elusive.

This story will be continued in Section 1.2.

Beta ensembles and the stochastic operator approach

The joint eigenvalue density of the $n \times n$ GO/U/SE is

$$Z_{n,\beta}^{-1} \prod_i e^{-\beta\lambda_i^2/4} \prod_{i<j} |\lambda_j - \lambda_i|^\beta \,, \tag{1.3}$$

again with $\beta = 1, 2, 4$ respectively. One may view (1.3) as the Boltzmann factor for a so-called log gas at inverse temperature β, whereupon it becomes natural to consider general

$\beta > 0$. (An analogous remark holds in the null Wishart setting; the Vandermonde factor persists and only the weight in the background product measure is different.) Forrester (2010) gives a comprehensive account of this point of view. Though one expects many local asymptotics of random matrix eigenvalues to extend to general "β-ensembles", the standard techniques involving orthogonal polynomials are unavailable as they rely heavily on the symmetry underlying the matrix models.

Recent progress began when Dumitriu and Edelman (2002) introduced a family of symmetric tridiagonal (Jacobi) matrix models with (1.3) as eigenvalue density for arbitrary β:

$$\frac{1}{\sqrt{\beta}} \begin{bmatrix} g & \chi_{(n-1)\beta} & & & \\ \chi_{(n-1)\beta} & g & \chi_{(n-2)\beta} & & \\ & \chi_{(n-2)\beta} & g & \ddots & \\ & & \ddots & \ddots & \chi_\beta \\ & & & \chi_\beta & g \end{bmatrix} \tag{1.4}$$

where the entries are independent random variables up to symmetry, $g \sim \text{Normal}(0, 2)$ and $\chi_k \sim \text{Chi}(k)$. Once again there is a similar story in the null Wishart case. Trotter (1984) already used tridiagonal forms of the Gaussian and null Wishart ensembles to give novel derivations of the Wigner semicircle and Marčenko-Pastur laws; Dimitriu and Edelman observed the similarity of the forms for $\beta = 1, 2, 4$, postulated interpolating matrix ensembles and proved their eigenvalue densities were as expected. An extension to more general weights will appear in Krishnapur, Rider and Virág (2011+).

Edelman and Sutton (2007) recognized that these random tridiagonal matrices offer a new route to asymptotic phenomena (see also Sutton 2005). They viewed the matrices as discrete random Schrödinger operators and conjectured limiting continuum random Schrödinger operators in the various spectral scaling regimes (bulk, soft edge, hard edge). Their promising heuristic arguments were first made rigorous in the soft edge case by Ramírez, Rider and Virág (2011) (**RRV**). The hard edge was treated by Ramírez and Rider (2009), and the bulk by Valkó and Virág (2009), Killip and Stoiciu (2009).

At the soft edge, the limiting object is a random Schrödinger operator on $L^2(\mathbb{R}_+)$ called the stochastic Airy operator. It is

$$-\frac{d^2}{dx^2} + x + \tfrac{2}{\sqrt{\beta}} b'_x \tag{1.5}$$

where b'_x is standard Gaussian white noise, and comes with Dirichlet boundary condition at the origin. RRV made sense of this operator and showed it is almost surely bounded

below with purely discrete spectrum. More significantly, they proved that the largest eigenvalues of (1.4) converge, jointly in distribution, to the low-lying eigenvalues of this operator. (The corresponding eigenvectors also converge when suitably embedded as functions.) In particular, the ground state defines Tracy-Widom distribution F_β for arbitrary $\beta > 0$. The authors further characterized F_β in terms of the explosion probability of a certain diffusion.

Broadly speaking, the general β perspective allows for an important distinction: Which facts truly depend on the classical symmetries, and which persist when only the physical character of the eigenvalue repulsion is retained? Many important phenomena seem to fall under the second class; evidence includes the recent general β bulk universality results of Bourgade, Erdős and Yau (2011).

1.2 Contributions and outline

Limits of spiked random matrices

In Chapters 2 and 3 we generalize the results and methods of RRV to develop a comprehensive picture of the BBP transition up to and including the critical regime. Real, complex and quaternion spiked Wishart matices and additively perturbed Gaussian matrices are treated simultaneously in a unified framework. While any description of a limit in the critical regime is new at $\beta = 1$, our results are in some ways more complete even at $\beta = 2$. For one thing, we allow considerably more general scaling assumptions on the parameters; in particular, the Wishart dimensions n, p are allowed to tend to infinity together arbitrarily. Furthermore, we view the perturbation as a parameter and describe a joint limit as the same data are spiked differently.

Chapter 2 treats rank one perturbations. The starting point is the observation that the perturbation commutes with the tridiagonalization in both the Gaussian and the Wishart cases. The resulting "spiked" tridiagonal models make sense for general β. Generalizing the method of RRV, we prove joint convergence of the top eigenvaues and eigenvectors to the low-lying states of the stochastic Airy operator (1.5) but with a boundary condition that now depends on the perturbation. In the subcritical cases we have the usual Dirichlet condition $f(0) = 0$; in the critical window, however, we have the general homogeneous linear condition $f'(0) = wf(0)$ where $w \in \mathbb{R}$ is a scaling parameter for the spike.

The distribution of the ground state forms a one-parameter family of deformations $F_\beta(x, w)$ of Tracy-Widom(β) with the latter recovered at $w = +\infty$. We proceed to characterize this family in terms of the diffusion introduced in RRV, and give a related characterization in terms of a simple parabolic linear PDE:

$$\begin{aligned} &\frac{\partial F}{\partial x} + \frac{2}{\beta}\frac{\partial^2 F}{\partial w^2} + \big(x - w^2\big)\frac{\partial F}{\partial w} = 0 \qquad \text{for } (x, w) \in \mathbb{R}^2, \\ &F(x, w) \to 1 \qquad \text{as } x, w \to \infty \text{ together}, \\ &F(x, w) \to 0 \qquad \text{as } w \to -\infty \text{ with } x \le x_0 < \infty. \end{aligned} \tag{1.6}$$

This boundary value problem is the starting point for Chapters 5 and 6.

In Appendix A, we recast the stochastic Airy eigenvalue problem as a classical Sturm-Liouville problem. This perspective is different from the one taken in RRV and allows for a simpler derivation of the required oscillation theory and spectral theory.

We note that Chapter 2 represents joint work with Bálint Virág that appeared first as Bloemendal and Virág (2010). Since that article was posted, Mo (2011) gave a completely different treatment of the real rank one case. Despite the difficulties mentioned, he succeeds with the standard program of obtaining forms for the joint eigenvalue and largest eigenvalue distributions and doing asymptotic analysis on the latter. Forrester (2011) makes some remarks on the two treatments.

Chapter 3 treats the general case of r spikes, i.e. rank r spiked Wishart matrices and additively perturbed Gaussian ensembles. Here we begin by introducing a new $(2r + 1)$-diagonal form capable of handling a rank r perturbation. The change of basis is in fact uniquely determined by fixing the subspace for the spikes; fortunately, it interacts well with the Gaussian structure of the Gaussian and Wishart models much like in the $r = 1$ case.

Viewing this band form as an $r \times r$ block tridiagonal matrix, we then develop a matrix-valued analogue of the RRV technology. We obtain a limiting Schrödinger operator with matrix-valued potential, in which the Brownian motion in (1.5) is replaced with a standard matrix Brownian motion. We consider spikes that may be subcritical or critical and obtain a general homogeneous linear boundary condition for the vector-valued eigenfunctions. The real, complex and quaternion cases are treated simultaneously; interestingly, however, there is no obvious general β analogue when $r > 1$.

Passage to diffusion and PDE characterizations of the laws involves an additional twist. Matrix Sturm oscillation theory (a topic initiated by Morse 1932) and Dyson's

Brownian (the dynamical version of (1.3) obtained by observing the eigenvalues of a matrix Brownian motion, see Dyson 1962) both play a role.

A note regarding the technical sections of Chapters 2 and 3 that generalize their counterparts in RRV: While certain parts of the RRV argument are recapitulated in Chapter 2, the development is significantly streamlined to use the quadratic form and variational characterization more efficiently. In a few places, we simply refer to corresponding arguments in RRV. Chapter 3 significantly generalizes the entire development, however, and here the proofs are fully self-contained.

Supercritical asymptotics and heuristics

Chapter 4 presents a new heuristic for understanding the supercritical regime of the BBP transition in the operator limit framework. The heuristic is justified by rigorously proving asymptotics for the ground state and ground state energy of the stochastic Airy operator as the parameter in the boundary condition tends to the supercritical limit. It is further illustrated by computing the largest eigenvalue separation for rank one supercritically spiked Wishart matrices and heuristically computing the Gaussian fluctuations. The heuristic also suggests a new description of supercritical eigenvector concentration in the tridiagonal basis.

A connection with Painlevé II

In Chapter 5 we use the PDE characterization (1.6) to give an independent rigorous proof of a formula of Baik (2006) for $F_2(x; w)$ given in terms of the Hastings McLeod solution of Painlevé II (1.2) and the associated Lax pair equations. More precisely, we show directly—at the level of the limiting objects—that Baik's formula satisfies the PDE and the boundary conditions. As a corollary, we establish the Tracy-Widom representation (1.1) in a rigorous and completely novel way. A similar result holds at $\beta = 4$.

This connection between a general β characterization of a random matrix limit law and a known explicit representation at classical β is the first and so far the only one of its kind. In joint work with Alexander Its, the connection is extended to include the Ablowitz-Segur family of solutions of Painlevé II. Finally, we report on a symbolic computation giving very strong evidence that the formulas of Baik (2006) for the higher-rank deformations can be similarly related to the "multispiked" PDE of Chapter 3. On the whole, the connection remains somewhat mysterious.

Numerics

Chapter 6 is a brief and informal report on attempts at solving the boundary value problem (1.6) numerically. The goal is a new, general β method of evaluating the Tracy-Widom(β) laws and the deformations $F_\beta(x, w)$ described above. Evidence is very promising: using an off-the-shelf Mathematica package we can compute an entire table of values for a given β in roughly 10 seconds on a laptop, and for $\beta = 1, 2$ these values were found to agree with published results to 7–8 digits. At the time of writing, development of the method is underway with Brian Sutton.

1.3 Concluding remarks

The contributions of this thesis may be summarized as follows. We develop a comprehensive description of the BBP transition, handling the real, complex and quaternion cases together and extending the existing picture even in the well-studied complex case; to do so, we significantly generalize the methods of RRV; we offer and justify a new heuristic for the supercritical behaviour; we use a PDE characterization of the limit laws to make the first connection between a general β characterization and known Painlevé structure at classical β; and based on the PDE we present a promising new general β method of evaluating the distributions numerically.

The stochastic operator approach stands in sharp contrast to the usual routes to asymptotic results about statistics of a solvable model or matrix ensemble. Typically one first works to obtain useable "finite-n" formulas for distributions of the statistics; one then proceeds with an asymptotic analysis. Although this method has been very successful in random matrix theory, it tends to be highly dependent on symmetry and integrable structure: when the symmetry class changes, the first step must be completely redone (if it can be done at all), even though the model is similar and the scalings are the same. The difficulty is well-illustrated by the spiked model: in spite of the success of BBP at $\beta = 2$, carrying out the program was not at all straightforward at $\beta = 1, 4$ even in the rank one case (Wang 2008, Mo 2011).

The point of view taken here is more probabilistic, essentially what Aldous and Steele (2004) called the "objective method". One begins with a form of the model that has a scaling limit object in some sense; heuristically, one "reads off" the asymptotic behaviour of the desired statistics from this object; the task is then to give a rigorous proof.

Chapter 2

One spike

2.1 Introduction

The study of sample covariance matrices is the oldest random matrix theory, predating Wigner's introduction of the Gaussian ensembles into physics by nearly three decades. Given a sample $X_1, \dots, X_n \in \mathbb{R}^p$ drawn from a large, centred population, form the $p \times n$ data matrix $X = [X_1 \dots X_n]$; the $p \times p$ matrix $S = XX^\dagger$ plays a central role in multivariate statistical analysis (Muirhead 1982, Bai 1999, Anderson 2003). The distribution in the i.i.d. Gaussian case is named after Wishart who computed the density in 1928. The classical story is that of the consistency of the **sample covariance matrix** $\frac{1}{n}S$ as an estimator of the **population covariance matrix** $\Sigma = \mathbf{E}\, X_i X_i^\dagger$ when the dimension p is fixed and the sample size n becomes large. The law of large numbers already gives $\frac{1}{n}S \to \Sigma$. In this fixed dimensional setting, the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p$ of S produce consistent estimators of the eigenvalues $\ell_1 \geq \cdots \geq \ell_p$ of Σ: for example, the **sample eigenvalue** $\frac{1}{n}\lambda_k$ tends almost surely to the **population eigenvalue** ℓ_k as $n \to \infty$, with Gaussian fluctuations on the order $n^{-1/2}$ (Anderson 1963). The same holds in the complex case $X_i \in \mathbb{C}^p$.

Contemporary problems typically involve **high dimensional data**, meaning that p is large as well—perhaps on the same order as n or even larger. In this setting, say with **null covariance** $\Sigma = I$, the sample eigenvalues may no longer concentrate around the population eigenvalue 1 but rather spread out over a certain compact interval. If $p/n \to c$ with $0 < c \leq 1$, Marčenko and Pastur (1967) proved that a.s. the empirical spectral distribution $\frac{1}{p}\sum_k \delta_{\lambda_k/n}$ converges weakly to the continuous distribution with

density

$$\frac{\sqrt{(b-x)(x-a)}}{2\pi cx}\mathbf{1}_{[a,b]}(x)$$

where $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$. (The singular case $c > 1$ is similar by the obvious duality between n and p, except that the $p-n$ zero eigenvalues become an atom at zero of mass $1-c^{-1}$.) This **Marčenko-Pastur law** is the analogue of Wigner's semicircle law in this setting of multiplicative rather than additive symmetrization (see also Silverstein and Bai 1995). The assumption of Gaussian entries may be significantly relaxed.

Often one is primarily interested in the *largest* eigenvalues, as for example in the widely practiced statistical method of principal components analysis. Here the goal is a good low-dimensional projection of a high-dimensional data set, i.e. one that captures most of the variance; the structure of the significant trends and correlations is estimated using the largest sample eigenvalues and their eigenvectors. The challenge is to determine which observed eigenvalues actually represent structure in the population, and understanding the behaviour in the null case is therefore an essential first step.

In the null case the first-order behaviour is simple: $\frac{1}{n}\lambda_k \to b$ a.s. for each fixed k as $n \to \infty$, i.e. none have limits beyond the edge of the support of the limiting spectral distribution (Geman 1980, Yin, Bai and Krishnaiah 1988). More interestingly, the fluctuations are no longer asymptotically Gaussian but are rather those now recognized as universal at a real symmetric or Hermitian **random matrix soft edge**: they are on the order $n^{-2/3}$, asymptotically distributed according to the appropriate **Tracy-Widom law**. The latter were introduced by Tracy and Widom (1994, 1996) as limiting largest eigenvalue distributions for the Gaussian ensembles (see also Forrester 1993) and have since been found to occur in diverse probabilistic models. The limit theorems for sample covariance matrices were proved by Johansson (2000) in the complex case and by Johnstone (2001) in the real case (see Soshnikov 2002 for the first universality results here). Restrictions $c \neq 0, \infty$ on the limiting dimensional ratio were removed by El Karoui (2003) (see also Péché 2009).

Motivated by principal components analysis, it is natural to study the behaviour of the largest sample eigenvalues when the population covariance is not null but rather has a few trends or correlations. Johnstone (2001) proposed the **spiked population model** in which all but a fixed finite number of population eigenvalues (the **spikes**) are taken to be 1 as n, p become large. Baik, Ben Arous and Péché (2005) (**BBP**) analyzed the spiked *complex* Wishart model and discovered a very interesting phenomenon: a phase

transition in the asymptotic behaviour of the largest sample eigenvalue as a function of the spikes. We restrict attention to the case of a single spike in the present chapter, setting $\ell_1 = \ell$, $\ell_2 = \ell_3 = \cdots = 1$.

In this **rank one perturbed case**, BBP describe three distinct regimes. Assume that $p/n = \gamma^2$ is compactly contained in $(0, 1]$. If $\ell_{n,p}$ is in compactly contained in $(0, 1+\gamma)$ then the behaviour of the top eigenvalue is exactly the same as in the null case:

$$\mathbf{P}\left(\frac{\gamma^{-1}}{(1+\gamma^{-1})^{4/3}} n^{2/3}\left(\frac{1}{n}\lambda_1 - (1+\gamma)^2\right) \leq x\right) \to F_2(x),$$

where F_2 is the Tracy-Widom law for the top GUE eigenvalue. This is the **subcritical regime**. If $\ell_{n,p}$ is compactly contained in $(1+\gamma, \infty)$ then the top eigenvalue separates from the bulk and has Gaussian fluctuations on the order $n^{-1/2}$:

$$\mathbf{P}\left(\left(\ell^2 - \gamma^2 \frac{\ell^2}{(\ell-1)^2}\right)^{-1/2} n^{1/2}\left(\frac{1}{n}\lambda_1 - \left(\ell + \gamma^2 \frac{\ell}{(\ell-1)}\right)\right) \leq x\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt.$$

This is the **supercritical regime**. Finally there is a one-parameter family of **critical** scalings in which $\ell_{n,p} - (1+\gamma)$ is on the order $n^{-1/3}$; these double scaling limits are tuned so that the fluctuations—which are on the order $n^{-2/3}$ as in the subcritical case—are asymptotically given by a certain one-parameter family of deformations of F_2. We refer the reader to the original work for details. Subsequent work includes a treatment of the singular case $p > n$ along the same lines (Onatski 2008), deeper investigations into the limiting kernels (Desrosiers and Forrester 2006), and generalizations beyond the spiked model (El Karoui 2007) and away from Gaussianity (Bai and Yao 2008, Féral and Péché 2009). BBP conjectured a similar phase transition for spiked *real* Wishart matrices, in the sense that all scalings should be the same but the limiting distributions would be different.

Now often referred to as the BBP transition, this picture is relevant in various applications. Within mathematics it has been applied to the TASEP model of interacting particles on the line (Ben Arous and Corwin 2011). Spiked complex Wishart matrices occur in problems in wireless communications (Telatar 1999). With these two exceptions, however, most applications involve data that are *real* rather than complex. They include economics and finance—Harding (2008) used the phase transition to explain an old standard example of the failure of PCA—and medical and population genetics—Patterson, Price and Reich (2006) discuss its role in attempting to answer such questions as "Given genotype data, is it from a homogeneous population?" Further applications include speech recognition, statistical learning and the physics of mixtures (see Johnstone 2007,

Paul 2007, Féral and Péché 2009 for references). In general, asymptotic distributions in the non-null cases are relevant when evaluating the power of a statistical test (Johnstone 2007).

Despite these developments, the conjectured BBP picture for spiked real Wishart matrices has proven elusive even in the rank one case. The difficulty is with the joint eigenvalue density: The complex case involves an integral over the unitary group that BBP analyzed via the Harish-Chandra-Itzykson-Zuber integral, a tool originating in representation theory that appears to have no straightforward analogue over the orthogonal group. Much is known, however. At the level of a law of large numbers, the phase transition is described by Baik and Silverstein (2006); a related separation phenomenon was observed already by Bai and Silverstein (1998, 1999). A broad generalization of the results on a.s. limits is developed by Benaych-Georges and Nadakuditi (2009) and dubbed "spiked free probability theory". Paul (2007), Bai and Yao (2008) prove Gaussian central limit theorems in the supercritical regime. Féral and Péché (2009) prove Tracy-Widom fluctuations in the subcritical regime under the scaling assumptions of BBP. Interestingly, Wang (2008) obtained a critical limiting distribution for certain rank one spiked *quaternion* Wishart matrices.

It remains to obtain the asymptotic behaviour in the critically spiked regime around the phase transition in the real case. We do so here, establishing the existence of limiting distributions under the scalings conjectured by BBP and characterizing the laws. Our results apply also to the complex case, and they are more general than the corresponding statements from BBP. We do not restrict the scaling of n, p beyond requiring that they tend to infinity together, nor that of ℓ beyond what is strictly necessary for the existence of a limiting distribution in the subcritical or critical regimes. We therefore allow for certain relevant possibilities that were previously excluded, namely $p \ll n$ and $p \gg n$. The picture of the dependence on the spike is also more complete: we include all intermediate scalings of ℓ with n, p across the subcritical and critical regimes. Separately, we describe a joint convergence in law when the same underlying data is spiked with different ℓ.

Since the work represented in this chapter was first posted (Bloemendal and Virág 2010), Mo (2011) gave a different treatment of the real rank one case. Despite the difficulties mentioned, he succeeds with the standard program of obtaining forms for the joint eigenvalue and largest eigenvalue distributions and doing asymptotic analysis on the latter. His description of the limiting distribution naturally looks very different from ours. See Forrester (2011) for some remarks on the two treatments and an alternative

construction of the "general β" model we now introduce.

We bypass the eigenvalue density altogether; our starting point is rather a *reduction of the matrix to tridiagonal form via Householder's algorithm*, a well-known tool in numerical analysis. Trotter (1984) observed that the algorithm interacts nicely with the Gaussian structure, using the resulting forms to derive the Wigner semicircle and Marčenko-Pastur laws without going through their moments. Observing the similarity of the forms in the $\beta = 1, 2, 4$ cases, Dumitriu and Edelman (2002) introduced interpolating matrix ensembles for all $\beta > 0$ whose eigenvalue density is given by Dyson's **Coulomb** or **log gas model**

$$\frac{1}{Z} \prod_{j<k} |\lambda_j - \lambda_k|^\beta \prod_j v(\lambda_j)^{\beta/2} \tag{2.1}$$

where v is the Hermite or the Laguerre weight and Z is a normalizing factor (see Forrester 2010 for more on such models). Incidentally, Trotter's argument applies to these **general β analogues** and establishes Wigner semicircle and Marčenko-Pastur laws in this setting. An extension to more general weights is part of a forthcoming work of Krishnapur, Rider and Virág (2011+).

The second step is to consider the tridiagonal ensemble as a discrete random Schrödinger operator (i.e. discrete Laplacian plus random potential) and then take a scaling limit at the soft edge to obtain a certain continuum random Schrödinger operator on the half-line. This "stochastic operator approach to random matrix theory" was pioneered by Edelman and Sutton (2007), Sutton (2005); in the soft edge case their heuristics were proved by Ramírez, Rider and Virág (2011), who in particular established joint convergence of the largest eigenvalues. Our method is directly based on the latter work and we refer to it throughout by the initials **RRV**. The key point is that both steps can be adapted to the setting of rank one perturbations. As we will see, the limiting operator feels the perturbation in the boundary condition at the origin.

In detail, let X be a $p \times n$ sample matrix whose columns are independent real $N(0, \Sigma)$ with $\Sigma = \mathrm{diag}\big(\ell, 1, \ldots, 1\big)$ for some $\ell > 0$; we shall say $S = XX^\dagger$ has the **ℓ-spiked p-variate real Wishart distribution with n degrees of freedom**. (There is no loss of generality in taking Σ diagonal in the Gaussian case.) We also consider the complex and quaternion cases. The tridiagonalization is carried out in detail in Section 2.3. The result is a symmetric tridiagonal $(n \wedge p) \times (n \wedge p)$ matrix $W^\dagger W$, where W is a certain

bidiagonal matrix with the same nonzero singular values as X. Explicitly, W is given by

$$W_{n,p}^{\beta,\ell} = \frac{1}{\sqrt{\beta}} \begin{bmatrix} \sqrt{\ell}\,\widetilde{\chi}_{\beta n} & & & & & \\ \chi_{\beta(p-1)} & \widetilde{\chi}_{\beta(n-1)} & & & & \\ & \chi_{\beta(p-2)} & \widetilde{\chi}_{\beta(n-2)} & & & \\ & & \ddots & \ddots & & \\ & & & \chi_{\beta(p-(n\wedge p)+1)} & \widetilde{\chi}_{\beta(n-(n\wedge p)+1)} & \\ & & & & \chi_{\beta(p-(n\wedge p))} & \end{bmatrix} \tag{2.2}$$

where $\beta = 1, 2, 4$ in the real, complex and quaternion cases respectively and the $\chi, \widetilde{\chi}$'s are mutually independent chi distributed random variables with parameters given by their indices. In fact (2.2) makes sense for any $\beta > 0$, and the resulting ensemble $W^\dagger W$ is a "spiked version" of the β-Laguerre ensemble of Dumitriu and Edelman (2002); we call it the **ℓ-spiked β-Laguerre ensemble with parameters** n, p. Such a matrix almost surely has exactly $n \wedge p$ distinct nonzero eigenvalues by the theory of Jacobi matrices. In the null case $\ell = 1$, their joint density is (2.1) with the Laguerre weight $v(x) = x^{|n-p|+1-2/\beta} e^{-x} \mathbf{1}_{x>0}$. We note that there is an obvious coupling of (2.2) over all $\ell > 0$; in the spiked Wishart cases it corresponds to the natural coupling obtained by considering X as a matrix of standard Gaussians left multiplied by $\sqrt{\Sigma}$.

In order to state our results, we now recall the **stochastic Airy operator** introduced by Edelman and Sutton (2007). Formally this is the random Schrödinger operator

$$\mathcal{H}_\beta = -\frac{d^2}{dx^2} + x + \tfrac{2}{\sqrt{\beta}} b'_x$$

acting on $L^2(\mathbb{R}_+)$ where b'_x is standard Gaussian white noise. RRV defined this operator rigorously and considered the eigenvalue problem $\mathcal{H}_\beta f = \Lambda f$ with Dirichlet boundary condition $f(0) = 0$. We will consider a general homogeneous boundary condition $f'(0) = wf(0)$, a Neumann or Robin condition for $w \in (-\infty, \infty)$ with the limiting Dirichlet case naturally corresponding to $w = +\infty$. Precise definitions will be given in Section 2.2 in a more general setting; for now, we write $\mathcal{H}_{\beta,w}$ to indicate the stochastic Airy operator together with this boundary condition.

We will see that, almost surely, $\mathcal{H}_{\beta,w}$ is bounded below with purely discrete, simple spectrum $\{\Lambda_0 < \Lambda_1 < \cdots\}$ for all $w \in (-\infty, \infty]$. This fact will be established simultaneously with the standard variational characterization: in Proposition 2.2.9, we show in particular that Λ_k and the corresponding eigenfunction f_k are given recursively by

$$\Lambda_k = \inf_{\substack{f \in L^2,\ \|f\|=1, \\ f \perp f_0, \ldots f_{k-1}}} \int_0^\infty \big(f'(x)^2 + x f^2(x)\big)\,dx + w f(0)^2 + \tfrac{2}{\sqrt{\beta}} \int_0^\infty f^2(x)\,db_x \tag{2.3}$$

in which we consider only candidates f for which the first integral is finite, and the stochastic integral is defined pathwise via integration by parts. Recall from RRV that the distribution $F_{\beta,\infty}$ of $-\Lambda_0$ in the Dirichlet case $w = +\infty$ may be taken as a definition of **Tracy-Widom(β)** for general $\beta > 0$, a one-parameter family of distributions interpolating between those at the standard values $\beta = 1, 2, 4$. Fixing β, the distributions $F_{\beta,w}$ for finite w may be thought of as a family of deformations of Tracy-Widom(β). We note that the pathwise dependence of $\mathcal{H}_{\beta,w}$ on the Brownian motion allows the operators to be coupled over w in a natural way.

Our first result gives a convergence in distribution at the soft edge of the ℓ-spiked β-Laguerre spectrum over the full range of subcritical and critical scalings. Note the absence of extraneous hypotheses on n, p and $\ell_{n,p}$.

Theorem 2.1.1. *Let $\ell_{n,p} > 0$. Let $S = S_{n,p}$ have the real (resp. complex, quaternion) $\ell_{n,p}$-spiked p-variate Wishart distribution with n degrees of freedom and set $\beta = 1$ (resp. 2, 4), or, let $\beta > 0$ and take $S_{n,p}$ from the $\ell_{n,p}$-spiked β-Laguerre ensemble with parameters n, p. Writing $m_{n,p} = \left(n^{-1/2} + p^{-1/2}\right)^{-2/3}$, suppose that*

$$m_{n,p}\left(1 - \sqrt{n/p}\big(\ell_{n,p} - 1\big)\right) \to w \in (-\infty, \infty] \qquad \text{as } n \wedge p \to \infty. \tag{2.4}$$

Let $\lambda_1 > \cdots > \lambda_{n\wedge p}$ be the nonzero eigenvalues of S. Then, jointly for $k = 1, 2, \ldots$ in the sense of finite-dimensional distributions, we have

$$\frac{m_{n,p}^2}{\sqrt{np}}\left(\lambda_k - \left(\sqrt{n} + \sqrt{p}\right)^2\right) \Rightarrow -\Lambda_{k-1} \qquad \text{as } n \wedge p \to \infty$$

where $\Lambda_0 < \Lambda_1 < \cdots$ are the eigenvalues of $\mathcal{H}_{\beta,w}$. Furthermore, the convergence holds jointly with respect to the natural couplings over all $\{\ell_{n,p}\}, w$ satisfying (2.4).

Remark 2.1.2. In the tridiagonal basis, the convergence holds also at the level of the corresponding eigenvectors. If the eigenvector corresponding to λ_k is embedded in $L^2(\mathbb{R}_+)$ as a step-function with step width $m_{n,p}^{-1}$ and support $[0, (n \wedge p)/m_{n,p}]$, then it converges to f_{k-1} in distribution with respect to the L^2 norm; the details are the subject of the next section. In particular, distributional convergence of the rescaled tridiagonal operators to $\mathcal{H}_{\beta,w}$ holds in the norm resolvent sense (see e.g. Weidmann 1997). Defining $\mathcal{H}_{\beta,w}$ as a closed operator on the appropriate (random) dense subspace of L^2 requires some care, however (see e.g. Savchuk and Shkalikov 1999) and we shall not pursue it here.

Remark 2.1.3. The supercritical regime $w = -\infty$ sees a macroscopic separation of the largest eigenvalue from the bulk of the spectrum; the fluctuations of λ_1 are on a larger order and they are asymptotically Gaussian, independent of the rest. See Chapter 4 for a partial treatment in the stochastic Airy framework. Though this regime is understood for both real and complex spiked sample covariance matrices (BBP, Paul 2007, Bai and Yao 2008), existing results do not cover intermediate "vanishingly supercritical" scalings of ℓ with n, p and thus leave a certain gap between the critical and supercritical regimes.

Remark 2.1.4. Work of Féral and Péché (2009) on the universality of real and complex BBP immediately allows extension of the previous theorem in the real and complex spiked Wishart cases to more general spiked sample covariance matrices. More precisely, the i.i.d. multivariate Gaussian columns of the data matrix X may be replaced with i.i.d. columns having zero mean and rank one spiked diagonal covariance, and satisfying some moment conditions. These authors make the same assumptions on the dimension ratio as BBP, but the null case universality result of Péché (2009) suggest these could be removed.

We prove Theorem 2.1.1 by establishing a more general technical result, Theorem 2.2.11 in Section 2.2. The latter theorem gives conditions under which the low-lying eigenvalues and corresponding eigenvectors of a large random symmetric tridiagonal matrix converge in law to those of a random Schrödinger operator on the half-line with a given potential and homogeneous boundary condition at the origin. Verifying the hypotheses for suitably scaled spiked Laguerre matrices will be relatively straightforward; we do it in Section 2.3. The approach follows that of RRV, where the null case of Theorem 2.1.1 is treated.

One advantage of such an approach is that it immediately yields results for other matrix models as well. In particular, finite-rank additive perturbations of **Gaussian orthogonal, unitary and symplectic ensembles (GO/U/SE)** have received considerable attention. The analogue of the BBP result in the perturbed GUE setting was established by Péché (2006), Desrosiers and Forrester (2006). Bassler, Forrester and Frankel (2010) treat an interesting generalization and mention some applications to physics. We consider a simple additive rank one perturbation of the GOE obtained by shifting the mean of every entry by the same constant $\mu/\sqrt{n}$. By orthogonal invariance, this has the same effect on the spectrum as shifting the (1,1) entry by $\sqrt{n}\,\mu$. With this perturbation, the usual tridiagonalization procedure works; the resulting form is the

$\beta = 1$ case of

$$G_n^{\beta,\mu} = \frac{1}{\sqrt{\beta}} \begin{bmatrix} \sqrt{2}\, g_1 + \sqrt{\beta n}\, \mu & \chi_{\beta(n-1)} & & & \\ \chi_{\beta(n-1)} & \sqrt{2}\, g_2 & \chi_{\beta(n-2)} & & \\ & \chi_{\beta(n-2)} & \sqrt{2}\, g_3 & \ddots & \\ & & \ddots & \ddots & \chi_\beta \\ & & & \chi_\beta & \sqrt{2}\, g_n \end{bmatrix}, \tag{2.5}$$

where the g's are independent standard Gaussians and the χ's are independent Chi random variables indexed by their parameter as before. The analogous procedure for a shifted mean GUE (resp. GSE) yields (2.5) with $\beta = 2$ (resp. 4). This matrix ensemble is a perturbed version of the β**-Hermite ensemble** of Dumitriu and Edelman (2002). In the unperturbed case $\mu = 0$, the joint eigenvalue density is (2.1) with the Hermite weight $v(x) = e^{-x^2/2}$. Again, the models are naturally coupled over all $\mu \in \mathbb{R}$.

As in the spiked real Wishart setting, the critical regime for the rank one perturbed GOE has resisted description. We show that the phase transition in the perturbed Hermite ensemble has the same characterization as the one in the Laguerre ensemble.

Theorem 2.1.5. *Let $\mu_n \in \mathbb{R}$. Let $G = G_n$ be a $(\mu_n/\sqrt{n})$-shifted mean $n \times n$ GOE (resp. GUE, GSE) matrix and set $\beta = 1$ (resp. 2, 4), or, let $\beta > 0$ and take $G_n = G_n^{\beta,\mu_n}$ as in* (2.5)*. Suppose that*

$$n^{1/3}\,(1 - \mu_n) \;\to\; w \in (-\infty, \infty] \qquad \text{as } n \to \infty. \tag{2.6}$$

Let $\lambda_1 > \cdots > \lambda_n$ be the eigenvalues of G. Then, jointly for $k = 0, 1, \ldots$ in the sense of finite-dimensional distributions, we have

$$n^{1/6}\left(\lambda_k - 2\sqrt{n}\right) \;\Rightarrow\; -\Lambda_{k-1} \qquad \text{as } n \to \infty$$

where $\Lambda_0 < \Lambda_1 < \cdots$ are the eigenvalues of $\mathcal{H}_{\beta,w}$. Furthermore, the convergence holds jointly with respect to the natural couplings over all $\{\mu_n\}, w$ satisfying (2.6)*.*

Remark 2.1.6. The remarks following the previous theorem apply also to this theorem; the universality issue is discussed in Féral and Péché (2007).

The limit of a rank one perturbed general β soft edge thus seems to be universal, just as at $\beta = 2$. We offer two alternative descriptions.

Theorem 2.1.7. *Fix $\beta > 0$ and let Λ_0 be the ground state energy of $\mathcal{H}_{\beta,w}$ where $w \in (-\infty, \infty]$. The distribution $F_{\beta,w}(x) = \mathbf{P}_{\beta,w}(-\Lambda_0 \le x)$ has the following alternative characterizations.*

(i) (RRV) Consider the stochastic differential equation

$$dp_x = \tfrac{2}{\sqrt{\beta}} db_x + \big(x - p_x^2\big)\, dx \tag{2.7}$$

and let $\mathbf{P}_{(x_0,w)}$ be the Itō diffusion measure on paths $\{p_x\}_{x \geq x_0}$ started from $p_{x_0} = w$. A path almost surely either explodes to $-\infty$ in finite time or grows like $p_x \sim \sqrt{x}$ as $x \to \infty$, and we have

$$F_{\beta,w}(x) = \mathbf{P}_{(x,w)}\big(p \text{ does not explode}\big). \tag{2.8}$$

(ii) The boundary value problem

$$\frac{\partial F}{\partial x} + \frac{2}{\beta}\frac{\partial^2 F}{\partial w^2} + \big(x - w^2\big)\frac{\partial F}{\partial w} = 0 \qquad \text{for } (x,w) \in \mathbb{R}^2, \tag{2.9}$$

$$\begin{aligned} F(x,w) &\to 1 \qquad \text{as } x, w \to \infty \text{ together}, \\ F(x,w) &\to 0 \qquad \text{as } w \to -\infty \text{ with } x \text{ bounded above} \end{aligned} \tag{2.10}$$

has a unique bounded solution, and we have $F_{\beta,w}(x) = F(x,w)$ for $w \in (-\infty, \infty)$. We recover the Tracy-Widom(β) distribution $F_{\beta,\infty}(x) = \lim_{w\to\infty} F(x,w)$.

Remark 2.1.8. These characterizations can be extended to the higher eigenvalues; details appear in Section 2.4.

In RRV the diffusion characterization is derived with classical tools, namely the Riccati transformation and Sturm oscillation theory. We review the relevant facts in Section 2.4 before proceeding to the boundary value problem. For a more classical and fully rigorous approach, see Appendix A.

While the PDE characterization amounts to a fairly straightforward reformulation of the diffusion characterization, the former is appealing in that it involves no stochastic objects. It also turns out to offer a promising way to evaluate the distributions numerically; see Chapter 6. Most interestingly, however, in Chapter 5 we show how it provides a sought-after connection with known integrable structure at classical β.

Separately, we remark that Adler, Delépine and van Moerbeke (2009) derive a completely different, third-order nonlinear PDE for what appears to be the same quantity $F_2(x;w)$ in a different context. It remains to reconcile their PDE with ours.

2.2 Limits of spiked tridiagonal matrices

In this section we strengthen the argument of RRV to apply in the rank one spiked cases. The main convergence result will be applied in the next section to the tridiagonal forms

described in the introduction.

Theorem 2.2.11 below generalizes Theorem 5.1 of RRV in a natural way, giving conditions under which the low-lying eigenvalues and corresponding eigenvectors of a random symmetric tridiagonal matrix converge in law to those of a random Schrödinger operator on the half-line with a given potential *and homogeneous boundary condition at the origin*. We include substantial parts of the original argument both for completeness and to highlight the new material; see Anderson, Guionnet and Zeitouni (2009) for another presentation of the original argument in a special case.

Matrix model and embedding

Underlying the convergence is the embedding of the discrete half-line $\mathbb{Z}_+ = \{0, 1, \ldots\}$ into $\mathbb{R}_+ = [0, \infty)$ via $j \mapsto j/m_n$, where the scale factors $m_n \to \infty$ but with $m_n = o(n)$. Define an associated embedding of function spaces by step functions:

$$\ell^2_n(\mathbb{Z}_+) \hookrightarrow L^2(\mathbb{R}_+), \quad (v_0, v_1, \ldots) \mapsto v(x) = v_{\lfloor m_n x \rfloor},$$

which is isometric with ℓ^2_n-norm $\|v\|^2 = m_n^{-1} \sum_{j=0}^{\infty} v_j^2$. Identify $\mathbb{R}^n$ with the initial coordinate subspace $\{v \in \ell^2_n : v_j = 0, j \geq n\}$. We will generally not refer to the embedding explicitly.

We define some operators on L^2, all of which leave ℓ^2_n invariant. The translation operator $(T_n f)(x) = f(x + m_n^{-1})$ extends the left shift on ℓ^2_n. The difference quotient $D_n = m_n(T_n - 1)$ extends a discrete derivative. Write $E_n = \mathrm{diag}(m_n, 0, 0, \ldots)$ for multiplication by $m_n \mathbf{1}_{[0, m_n^{-1})}$, a "discrete delta function at the origin", and $R_n = \mathrm{diag}(1, \ldots, 1, 0, 0, \ldots)$ for multiplication by $\mathbf{1}_{[0, n/m_n)}$, which extends orthogonal projection $\ell^2_n \to \mathbb{R}^n$.

Let $(y_{n,i;j})_{j=0,\ldots,n}$, $i = 1, 2$ be two discrete-time real-valued random processes with $y_{n,i;0} = 0$, and let w_n be a real-valued random variable. Embed the processes as above. Define a "potential" matrix (or operator)

$$V_n = \mathrm{diag}(D_n y_{n,1}) + \tfrac{1}{2}\big(\mathrm{diag}(D_n y_{n,2}) T_n + T_n^\dagger \, \mathrm{diag}(D_n y_{n,2})\big),$$

and finally set

$$H_n \;=\; R_n\big(D_n^\dagger D_n + V_n + w_n E_n\big). \tag{2.11}$$

This operator leaves the subspace $\mathbb{R}^n$ invariant. The matrix of its restriction with respect

to the coordinate basis is symmetric tridiagonal, with on- and off-diagonal processes

$$m_n^2 + (y_{n,1;1} + w_n)m_n,\ 2m_n^2 + (y_{n,1;2} - y_{n,1;1})m_n,\ \ldots,\ 2m_n^2 + (y_{n,1;n} - y_{n,1;n-1})m_n \tag{2.12}$$

$$-m_n^2 + \tfrac{1}{2}y_{n,2;1}m_n,\ -m_n^2 + \tfrac{1}{2}(y_{n,2;2} - y_{n,2;1})m_n,\ \ldots,\ -m_n^2 + \tfrac{1}{2}(y_{n,2;n-1} - y_{n,2;n-2})m_n \tag{2.13}$$

respectively. We denote this random matrix also as H_n, and call it a **spiked tridiagonal ensemble**. (We could have absorbed w_n into $y_{n,1}$ as an additive constant, but keep it separate for reasons that will soon be apparent.)

As in RRV, convergence rests on a few key assumptions on the random variables just introduced. By choice, no additional scalings will be required.

Assumption 1 (Tightness and convergence). There exists a continuous random process $\{y(x)\}_{x\geq 0}$ with $y(0) = 0$ such that

$$\begin{gathered}\{y_{n,i}(x)\}_{x\geq 0},\ i = 1, 2 \quad \text{are tight in law,} \\ y_{n,1} + y_{n,2} \Rightarrow y \quad \text{in law}\end{gathered} \tag{2.14}$$

with respect to the compact-uniform topology on paths.

Assumption 2 (Growth and oscillation bounds). There is a decomposition

$$y_{n,i;j} = m_n^{-1}\sum_{k=0}^{j-1}\eta_{n,i;k} + \omega_{n,i;j}$$

with $\eta_{n,i;j} \geq 0$ such that for some deterministic unbounded nondecreasing continuous functions $\overline{\eta}(x) > 0$, $\zeta(x) \geq 1$ not depending on n, and random constants $\kappa_n \geq 1$ defined on the same probability spaces, the following hold: The κ_n are tight in distribution, and for each n we have almost surely

$$\overline{\eta}(x)/\kappa_n - \kappa_n \ \leq\ \eta_{n,1}(x) + \eta_{n,2}(x) \ \leq\ \kappa_n\big(1 + \overline{\eta}(x)\big), \tag{2.15}$$

$$\eta_{n,2}(x) \ \leq\ 2m_n^2, \tag{2.16}$$

$$|\omega_{n,1}(\xi) - \omega_{n,1}(x)|^2 + |\omega_{n,2}(\xi) - \omega_{n,2}(x)|^2 \ \leq\ \kappa_n\big(1 + \overline{\eta}(x)/\zeta(x)\big) \tag{2.17}$$

for all $x, \xi \in [0, n/m_n]$ with $|\xi - x| \leq 1$.

Assumption 3 (Critical or subcritical spiking). For some nonrandom $w \in (-\infty, \infty]$, we have

$$w_n \ \to\ w \quad \text{in probability.} \tag{2.18}$$

The necessity of first and third assumptions will be evident when we define a continuum limit and prove convergence. The more technical second assumption ensures tightness of the matrix eigenvalues; its limiting version (derived in the next subsection) will guarantee discreteness of the limiting spectrum. Lastly, we note that for given y_n the models may be coupled over different choices of w_n.

Reduction to deterministic setting

In the next subsection we will define a limiting object in terms of y and w; we want to prove that the discrete models converge to this continuum limit in law. We reduce the problem to a deterministic convergence statement as follows. First, select any subsequence. It will be convenient to extract a further subsequence so that certain additional tight sequences converge jointly in law; Skorokhod's representation theorem (see Ethier and Kurtz 1986) says this convergence can be realized almost surely on a single probability space. We may then proceed pathwise.

In detail, consider (2.14)–(2.18). Note in particular that the upper bound of (2.15) shows that the piecewise linear process $\left\{\int_0^x \eta_{n,i}\right\}_{x\geq 0}$ is tight in distribution under the compact-uniform topology for $i = 1, 2$. Given a subsequence, we pass to a further subsequence so that the following distributional limits exist jointly:

$$\begin{aligned} y_{n,i} &\Rightarrow y_i, \\ \textstyle\int_0 \eta_{n,i} &\Rightarrow \eta_i^\dagger, \\ \kappa_n &\Rightarrow \kappa, \end{aligned} \tag{2.19}$$

for $i = 1, 2$, where convergence in the first two lines is in the compact-uniform topology. We realize (2.19) pathwise a.s. on some probability space and continue in this deterministic setting.

We can take the bounds (2.15),(2.17) to hold with κ_n replaced with a single constant κ. Observe that (2.15) gives a local Lipschitz bound on the $\int \eta_{n,i}$, which is inherited by their limits $\eta_i^\dagger$. Thus $\eta_i = \left(\eta_i^\dagger\right)'$ is defined almost everywhere on $\mathbb{R}_+$, satisfies (2.15), and may be defined to satisfy this inequality everywhere. Furthermore, one easily checks that $m_n^{-1} \sum \eta_{n,i} \to \int \eta_i$ compact-uniformly as well (use continuity of the limit). Therefore $\omega_{n,i} = y_{n,i} - m_n^{-1} \sum \eta_{n,i}$ must have a continuous limit ω_i for $i = 1, 2$; moreover, the bound (2.17) is inherited by the limits. Lastly, put $\eta = \eta_1 + \eta_2$, $\omega = \omega_1 + \omega_2$ and note that $y_i = \int \eta_i + \omega_i$ and $y = \int \eta + \omega$.

Without further reference to the subsequences, we will assume this situation for the remainder of the section.

Limiting operator and variational characterization

Formally, the limit of the spiked tridiagonal ensemble H_n will be the eigenvalue problem

$$\begin{aligned} \mathcal{H}f &= \Lambda f \quad \text{on } \mathbb{R}_+ \\ f'(0) &= wf(0), \qquad\qquad f(+\infty) = 0 \end{aligned} \tag{2.20}$$

where $\mathcal{H} = -d^2/dx^2 + y'(x)$ and $w \in (-\infty, \infty]$ is fixed. If $w = +\infty$, the boundary condition is to be interpreted as $f(0) = 0$; we refer to this as the **Dirichlet case**, and it will require special treatment in what follows. The primary object for us will be a symmetric bilinear form associated with the eigenvalue problem (2.20).

Define a space of test functions C_0^∞ consisting of smooth functions on $\mathbb{R}_+$ with compact support that *may contain the origin, except in the Dirichlet case.* Denote by $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$ the norm and inner product of $L^2[0, \infty)$. Define a weighted Sobolev norm by

$$\|f\|_*^2 = \left\|f'\right\|^2 + \left\|f\sqrt{1+\overline{\eta}}\right\|^2$$

and an associated Hilbert space L^* as the closure of C_0^∞ under this norm. Note that our L^* differs slightly from the one in RRV. We register some basic facts about L^* functions.

Fact 2.2.1. *Any $f \in L^*$ is uniformly Hölder(1/2)-continuous, satisfies $|f(x)| \le \|f\|_*$ for all x, and in the Dirichlet case has $f(0) = 0$.*

Proof. We have $|f(y) - f(x)| = \left|\int_x^y f'\right| \le \|f'\| \, |y - x|^{1/2}$. For $f \in C_0^\infty$ we have

$$f(x)^2 = -\int_x^\infty (f^2)' \le 2 \|f'\| \, \|f\| \le \|f\|_*^2 .$$

An L^*-bounded sequence in C_0^∞ therefore has a compact-uniformly convergent subsequence, so we can extend this bound to $f \in L^*$ and conclude further that $f(0) = 0$ in the Dirichlet case. □

For future reference, we also record some compactness properties of the L^*-norm.

Fact 2.2.2. *Every L^*-bounded sequence has a subsequence converging in the following modes: (i) weakly in L^*, (ii) derivatives weakly in L^2, (iii) uniformly on compacts, and (iv) in L^2.*

Proof. (i) and (ii) are just Banach-Alaoglu; (iii) is the previous fact and Arzelà-Ascoli again; (iii) implies L^2 convergence locally, while the uniform bound on $\int \overline{\eta} f_n^2$ produces the uniform integrability required for (iv). Note that the weak limit in (ii) really is the derivative of the limit function, as one can see by integrating against functions $\mathbf{1}_{[0,x]}$ and using pointwise convergence. □

We introduce a symmetric bilinear form on $C_0^\infty \times C_0^\infty$ by

$$\mathcal{H}_{Y,W}(\varphi,\psi) \;=\; \langle \varphi', \psi' \rangle - \big\langle (\phi\psi)', y \big\rangle + w\,\varphi(0)\psi(0), \tag{2.21}$$

dropping the last term in the Dirichlet case. (We could have absorbed w into y as an additive constant in the finite case, but prefer to keep the boundary term separate.) Formally, $\mathcal{H}_{Y,W}(\varphi, f)$ is just $\langle \varphi, \mathcal{H} f \rangle$; notice how the mixed boundary condition is built "implicitly" into the form, while the Dirichlet boundary condition is built "explicitly" into the space.

Lemma 2.2.3. *There are constants $c, C > 0$ so that the following bounds holds for all $f \in C_0^\infty$:*

$$c\,\|f\|_*^2 - C\,\|f\|^2 \;\le\; \mathcal{H}_{Y,W}(f,f) \;\le\; C\,\|f\|_*^2 . \tag{2.22}$$

In particular, $\mathcal{H}_{Y,W}(\cdot,\cdot)$ extends uniquely to a continuous symmetric bilinear form on $L^* \times L^*$ satisfying the same bounds.

Proof. For the first two terms of (2.21), we use the decomposition $y = \int \eta + \omega$ from the previous subsection. Integrating the $\int \eta$ term by parts, the limiting version of (2.15) easily yields

$$\tfrac{1}{\kappa}\,\|f\|_*^2 - C'\,\|f\|^2 \;\le\; \|f'\|^2 + \big\langle f^2, \eta \big\rangle \;\le\; \kappa\,\|f\|_*^2 .$$

Break up the ω term as follows. The moving average $\overline{\omega}_x = \int_x^{x+1} \omega$ is differentiable with $\overline{\omega}_x' = \omega_{x+1} - \omega_x$; writing $\omega = \overline{\omega} + (\omega - \overline{\omega})$, we have

$$-\big\langle (f^2)', \omega \big\rangle = \big\langle f, \overline{\omega}' f \big\rangle + 2\big\langle f', (\overline{\omega} - \omega) f \big\rangle.$$

The limiting version of (2.17) gives $\max\big(|\omega_\xi - \omega_x|, |\omega_\xi - \omega_x|^2\big) \le C_\varepsilon + \varepsilon\overline{\eta}(x)$ for $|\xi - x| \le 1$, where ε can be made small. In particular, the first term above is bounded absolutely by $\varepsilon\,\|f\|_*^2 + C_\varepsilon\,\|f\|^2$. Averaging, we also get $|\overline{\omega}_x - \omega_x| \le (C_\varepsilon + \varepsilon\overline{\eta}(x))^{1/2}$; Cauchy-Schwarz then bounds the second term above absolutely by $\sqrt{\varepsilon}\int_0^\infty (f')^2 + \frac{1}{\sqrt{\varepsilon}}\int_0^\infty f^2 (C_\varepsilon + \varepsilon\overline{\eta})$ and thus by $\sqrt{\varepsilon}\,\|f\|_*^2 + C_\varepsilon'\,\|f\|^2$. Now combine all the terms and set ε small to obtain the result.

For the boundary term $wf(0)^2$, it suffices to obtain a bound of the form $f(0)^2 \leq \varepsilon \|f\|_*^2 + C''_\varepsilon \|f\|^2$. But $f(0)^2 \leq 2\|f'\|\,\|f\|$ from the proof of Fact 2.2.1 gives such a bound with $C''_\varepsilon = 1/\varepsilon$.

The L^* form bound follows from the fact that the L^*-norm dominates the L^2-norm. We obtain the quadratic form bound $|\mathcal{H}_{Y,W}(f,f)| \leq C\|f\|_*^2$; it is a standard Hilbert space fact that it may be polarized to a bilinear form bound (see e.g. Halmos 1957). □

Definition 2.2.4. Call (Λ, f) an **eigenvalue-eigenfunction pair** if $f \in L^*$, $\|f\| = 1$, and for all $\varphi \in C_0^\infty$ we have

$$\mathcal{H}_{Y,W}(\varphi, f) = \Lambda \langle \varphi, f \rangle . \tag{2.23}$$

Note that (2.23) then automatically holds for all $\varphi \in L^*$, by L^*-continuity of both sides.

Remark 2.2.5. This definition represents a weak or distributional version of the problem (2.20). As further justification, integrate by parts to write the definition

$$\langle \varphi', f' \rangle - \big\langle (\varphi f)', y \big\rangle + w\,\varphi(0) f(0) = \Lambda \langle \varphi, f \rangle$$

in the form

$$\langle \varphi', f' \rangle - \langle \varphi', fy \rangle + \big\langle \varphi', \textstyle\int_0 f' y \big\rangle - wf(0)\,\langle \varphi', \mathbf{1} \rangle = -\Lambda \big\langle \varphi', \textstyle\int_0 f \big\rangle ,$$

which is equivalent to

$$f'(x) = wf(0) + y(x)f(x) - \int_0^x yf' - \Lambda \int_0^x f \qquad \text{a.e. } x. \tag{2.24}$$

In the Dirichlet case the first term on the right is replaced with $f'(0)$. On the one hand (2.24) shows that f' has a continuous version, and the equation may be taken to hold everywhere. In particular, f satisfies the boundary condition of (2.20) at the origin. On the other hand, (2.24) is a straightforward integrated version of the eigenvalue equation in which the potential term has been interpreted via integration by parts. This equation will be useful in Lemma 2.2.7 below and is the starting point for a rigorous derivation of (2.7) in the stochastic Airy case.

Remark 2.2.6. The requirement $f \in L^*$ in Definition 2.2.4 is a technical convenience. Regarding regularity, we need f at least absolutely continuous to make sense of the eigenvalue equation in either an integrated or a distributional sense; we have seen, however, that solutions are in fact C^1. Regarding behaviour at infinity, the diffusion picture developed by RRV shows a dichotomy: almost all solutions of the eigenvalue equation grow super-exponentially at infinity, except for the eigenfunctions which decay sub-exponentially.

We now characterize eigenvalue-eigenfunction pairs variationally. It is easy to see that each eigenspace is finite-dimensional: a sequence of normalized eigenfunctions must have an L^2-convergent subsequence by (2.22) and Fact 2.2.2. By the same argument, eigenvalues can accumulate only at infinity. In fact, more is true:

Lemma 2.2.7. *For each $\Lambda \in \mathbb{R}$, the corresponding eigenspace is at most one-dimensional.*

Proof. By linearity, it suffices to show a solution of (2.24) with $f'(0) = f(0) = 0$ must vanish identically. Integrate by parts to write

$$f'(x) = y(x)\int_0^x f' - \int_0^x yf' - \Lambda x \int_0^x f' + \Lambda \int_0^x tf'(t)dt,$$

which implies that $|f'(x)| \le C(x) \int_0^x |f'|$ with some $C(x) < \infty$ increasing in x. Gronwall's lemma then gives $f'(x) = 0$ for all $x \ge 0$. □

Remark 2.2.8. Compare these simple arguments with Proposition 3.5 of RRV, which requires the diffusion representation.

The eigenfunction corresponding to a given eigenvalue is thus uniquely specified with the additional sign normalization $-\frac{\pi}{2} < \arg\big(f(0), f'(0)\big) \le \frac{\pi}{2}$. We order eigenvalue-eigenfunction pairs by their eigenvalues. As usual, it follows from the symmetry of the form that distinct eigenfunctions are L^2-orthogonal.

Proposition 2.2.9. *There is a well-defined $(k+1)$st lowest eigenvalue-eigenfunction pair (Λ_k, f_k); it is given recursively by the minimum and minimizer in the variational problem*

$$\inf_{\substack{f \in L^*,\ \|f\|=1,\\ f \perp f_0, \dots, f_{k-1}}} \mathcal{H}_{Y,W}(f,f)\,.$$

Remark 2.2.10. Since we must have $\Lambda_k \to \infty$, the min-max principle (Reed and Simon 1978) states that $\{\Lambda_0, \Lambda_1, \dots\}$ exhausts the full spectrum and the operator has compact resolvent. We do not make this precise. Appendix A contains a more classical approach to the spectral theory of the stochastic Airy operator, and includes the statement that the eigenvectors form a complete orthonormal set in L^2.

Proof. First taking $k = 0$, the infimum $\tilde{\Lambda}$ is finite by (2.22). Let f_n be a minimizing sequence; it is L^*-bounded, again by (2.22). Pass to a subsequence converging to $f \in L^*$ in all the modes of Fact 2.2.2. In particular $1 = \|f_n\| \to \|f\|$, so $\mathcal{H}_{Y,W}(f,f) \ge \tilde{\Lambda}$ by

definition. But also

$$\begin{aligned}\mathcal{H}_{Y,W}(f,f) &= \|f'\|^2 + \int f^2\eta + \langle f, \overline{\omega}' f\rangle + 2\langle f', (\overline{\omega}-\omega)f\rangle + wf(0)^2 \\ &\leq \liminf_{n\to\infty} \mathcal{H}_{Y,W}(f_n, f_n)\end{aligned}$$

by a term-by-term comparison. Indeed, the inequality holds for the first term by weak convergence, and for the second term by pointwise convergence and Fatou's lemma; the remaining terms are just equal to the corresponding limits, because the second members of the inner products converge in L^2 by the bounds from the proof of Lemma 2.2.3 together with L^*-boundedness and L^2-convergence. Therefore $\mathcal{H}_{Y,W}(f,f) = \tilde{\Lambda}$.

A standard argument now shows $(\tilde{\Lambda}, f)$ is an eigenvalue-eigenfunction pair: taking $\varphi \in C_0^\infty$ and ε small, put $f^\varepsilon = (f+\varepsilon\varphi)/\|f+\varepsilon\varphi\|$; since f is a minimizer, $\frac{d}{d\varepsilon}\big|_{\varepsilon=0} \mathcal{H}_{Y,W}(f^\varepsilon, f^\varepsilon)$ must vanish; the latter says precisely (2.23) with $\tilde{\Lambda}$. Finally, suppose (Λ, g) is any eigenvalue-eigenfunction pair; then $\mathcal{H}_{Y,W}(g,g) = \Lambda$, and hence $\tilde{\Lambda} \leq \Lambda$. We are thus justified in setting $\Lambda_0 = \tilde{\Lambda}$ and $f_0 = f$.

Proceed inductively, minimizing now over $\{f \in L^* : \|f\| = 1, f \perp f_0, \dots f_{k-1}\}$. Again, L^2-convergence of a minimizing sequence guarantees that the limit remains admissible; as before, the limit is in fact a minimizer; conclude by applying the arguments of the previous paragraph in the ortho-complement. The preceding lemma guarantees that $\Lambda_0 < \Lambda_1 < \cdots$, and that the corresponding eigenfunctions $f_0, f_1, \dots$ are uniquely determined. □

Statement

We are finally ready to state the main result of this section. When we speak of an **eigenvalue-eigenvector pair** (λ, v) of an $n \times n$ matrix, we take $v \in \mathbb{R}^n$ embedded in $L^2(\mathbb{R}_+)$ as usual and normalized by $\|v\| = 1$ and $-\frac{\pi}{2} < \arg(v_0, v_1) \leq \frac{\pi}{2}$.

Theorem 2.2.11. *Suppose that H_n as in (2.11) satisfies Assumptions 1–3 and let $(\lambda_{n,k}, v_{n,k})$ be its $(k+1)$st lowest eigenvalue-eigenvector pair. Define the corresponding form $\mathcal{H}_{y,w}$ as in (2.21) and let (Λ_k, f_k) be its a.s. defined $(k+1)$st lowest eigenvalue-eigenfunction pair. Then, jointly for all $k = 0, 1, \dots$ in the sense of finite dimensional distributions, we have $\lambda_{n,k} \Rightarrow \Lambda_k$ and $v_{n,k} \Rightarrow_{L^2} f_k$ as $n \to \infty$. The convergence holds jointly over different w_n, w for given y_n, y.*

Remark 2.2.12. Essentially, the resolvent matrices (precomposed with the corresponding

finite-rank projections) are converging to the continuum resolvent in L^2-operator norm. We do not define the resolvent operator here.

The proof will be given over the course of the next two subsections. Recall that we proceed in the subsequential almost-sure context of the previous subsection.

Tightness

We will need a discrete analogue of the L^*-norm and a counterpart of Lemma 2.2.3 with constants uniform in n. For $v \in \mathbb{R}^n$, define the L_n^***-norm** by

$$\|v\|_{*n}^2 = \begin{cases} \left\|D_n v\right\|^2 + \left\|v\sqrt{1+\overline{\eta}}\right\|^2 & \text{if } w < \infty, \\ \left\|D_n v\right\|^2 + \left\|v\sqrt{1+\overline{\eta}}\right\|^2 + w_n v_0^2 & \text{if } w = \infty, \end{cases}$$

noting that the additional term in the Dirichlet case is nonnegative for sufficiently large n.

Remark 2.2.13. As in the continuum version, the Dirichlet boundary condition must be put explicitly into the norm (see also Lemma 2.2.16 below). The case considered in RRV has $w_n = m_n$ in our notation; though it is somewhat hidden in the definitions, the L_n^*-norm used there contains a term $m_n v_0^2$.

Lemma 2.2.14. *There are constants $c, C > 0$ so that, for each n and all $v \in \mathbb{R}^n$,*

$$c\,\|v\|_{*n}^2 - C\,\|v\|^2 \le \langle v, H_n v\rangle \le C\,\|v\|_{*n}^2\,. \tag{2.25}$$

Proof. The derivative and potential terms may be handled exactly as in RRV (proof of Lemma 5.6); the proof of Lemma 3.3.13 in the next chapter contains a streamlined version in a more general setting. For the spike term $w_n v_0^2$ we recall Assumption 3. In the $w < \infty$ case the w_n are bounded, so it suffices to obtain a bound of the form $v_0^2 \le \varepsilon\,\|v\|_{*n}^2 + C_\varepsilon\,\|v\|^2$ for each $\varepsilon > 0$ where $\varepsilon, C_\varepsilon$ do not depend on n. Mimicking the continuum version in the proof of Fact 2.2.1, we have

$$v_0^2 = \left\langle -D_n v^2, \mathbf{1}\right\rangle = \langle -(D_n v)(T_n v + v), \mathbf{1}\rangle \le \langle -(D_n v), T_n v + v\rangle \le 2\,\|D_n v\|\,\|v\|\,,$$

which gives the desired bound with $C_\varepsilon = 1/\varepsilon$.

In the Dirichlet case, start with (2.25) but with the spike term left out (both of the form and the norm); it can be easily added back in by simply ensuring that $c \le 1$ and $C \ge 1$. □

Remark 2.2.15. If $w_n \to -\infty$ then the lower bound in Lemma 2.2.14 breaks down: the lowest eigenvalue of H_n really is going to $-\infty$. This is the supercritical regime; see Chapter 4.

Convergence

We begin with a lemma, a discrete-to-continuous version of Fact 2.2.2.

Lemma 2.2.16. *Let $f_n \in \mathbb{R}^n$ with $\|f_n\|_{*n}$ uniformly bounded. Then there exist $f \in L^*$ and a subsequence along which (i) $f_n \to f$ uniformly on compacts, (ii) $f_n \to_{L^2} f$, and (iii) $D_n f_n \to f'$ weakly in L^2.*

Proof. Consider $g_n(x) = f_n(0) + \int_0^x D_n f_n$, a piecewise-linear version of f_n; they coincide at points $x = i/m_n$, $i \in \mathbb{Z}_+$. One easily checks that $\|g_n\|_*^2 \le 2\|f_n\|_{*n}^2$, so some subsequence $g_n \to f \in L^*$ in all the modes of Fact 2.2.2; in the Dirichlet case, the extra term in the L_n^* norm guarantees that $f(0) = 0$. But then also $f_n \to f$ compact-uniformly by a simple argument using the uniform continuity of f, $f_n \to_{L^2} f$ because $\|f_n - g_n\|^2 \le (1/3n^2)\|D_n f_n\|^2$, and $D_n f_n \to f'$ weakly in L^2 because $D_n f_n = g_n'$ a.e. □

Next we establish a kind of weak convergence of the form $\langle \cdot, H_n \cdot \rangle$ to $\mathcal{H}_{Y,W}(\cdot,\cdot)$. Let $\mathcal{P}_n$ be orthogonal projection from L^2 onto $\mathbb{R}^n$. One can check the following: for $f \in L^2$, $\mathcal{P}_n f \to_{L^2} f$ (the Lebesgue differentiation theorem gives pointwise convergence and we have uniform L^2-integrability); for smooth f, $\mathcal{P}_n f \to f$ uniformly on compacts; further, if $f' \in L^2$ then $D_n f \to_{L^2} f'$ ($D_n f$ is a convolution of f' with an approximate delta). Observe that $\mathcal{P}_n$ commutes with R_n and with $D_n R_n$.

Lemma 2.2.17. *Let $f_n \to f$ be as in the hypothesis and conclusion of Lemma 2.2.16. Then for all $\varphi \in C_0^\infty$ we have $\langle \varphi, H_n f_n \rangle \to \mathcal{H}_{Y,W}(\varphi, f)$. In particular, $\mathcal{P}_n\varphi \to \varphi$ in this way and so*

$$\langle \mathcal{P}_n\varphi, H_n \mathcal{P}_n\varphi \rangle = \langle \varphi, H_n \mathcal{P}_n\varphi \rangle \to \mathcal{H}_{Y,W}(\varphi,\varphi)\,. \tag{2.26}$$

Proof. Note that if $f_n \to_{L^2} f$, g_n is L^2-bounded and $g_n \to g$ weakly in L^2, then $\langle f_n, g_n \rangle \to \langle f, g \rangle$. Therefore $\langle \varphi, D_n^\dagger D_n f_n \rangle = \langle D_n\varphi, D_n f_n \rangle \to \langle \varphi', f' \rangle$. The potential term converges as in RRV (proof of Lemma 5.7) or Chapter 3 (proof of Lemma 3.3.16). Moreover, the spike term converges to the boundary term:

$$w_n f_n(0)(\mathcal{P}_n\varphi)(0) \to w\, f(0)\varphi(0),$$

where in the Dirichlet case the left side vanishes for n large because φ is supported away from 0.

For the second statement, the uniform L^*_n bound follows from the following observations: $\left\|(\mathcal{P}_n\varphi)\sqrt{1+\overline{\eta}}\right\| = \left\|\mathcal{P}_n\varphi\sqrt{1+\overline{\eta}}\right\| \leq \left\|\varphi\sqrt{1+\overline{\eta}}\right\|$; for n large enough that $R_n\varphi = \varphi$ we have $\|D_n\mathcal{P}_n\varphi\| = \|\mathcal{P}_nD_n\varphi\| \leq \|D_n\varphi\| \leq \|\varphi'\|$ (Young's inequality); and in the Dirichlet case, the extra term vanishes for n large. The convergence is easy: $\mathcal{P}_n\varphi \to \varphi$ compact-uniformly and in L^2, and for $g \in L^2$ we have $\langle g, D_n\mathcal{P}_n\varphi\rangle = \langle \mathcal{P}_ng, D_n\varphi\rangle \to \langle g, \varphi'\rangle$. □

Finally, we recall the argument of RRV to put all the pieces together.

Proof of Theorem 2.2.11. First we show that for all k we have $\underline{\lambda}_k = \liminf \lambda_{n,k} \geq \Lambda_k$. Assume that $\underline{\lambda}_k < \infty$. The eigenvalues of H_n are uniformly bounded below by Lemma 2.2.14, so there is a subsequence along which $(\lambda_{n,1}, \ldots, \lambda_{n,k}) \to (\xi_1, \ldots, \xi_k = \underline{\lambda}_k)$. By the same lemma the corresponding eigenvector sequences have L^*_n-norm uniformly bounded; pass to a further subsequence so that they all converge as in Lemma 2.2.16. The limit functions are orthonormal, and by Lemma 2.2.17 they are eigenfunctions with eigenvalues ξ_k. There are therefore k distinct eigenvalues at most $\underline{\lambda}_k$, as required.

We proceed by induction, assuming the conclusion of the theorem up to $k-1$. First find $f^\varepsilon_k \in C^\infty_0$ with $\|f^\varepsilon_k - f_k\|_* < \varepsilon$. Consider the vector

$$f_{n,k} = \mathcal{P}_nf^\varepsilon_k - \sum_{j=0}^{k-1} \langle v_{n,j}, \mathcal{P}_nf^\varepsilon_k\rangle\, v_{n,j}.$$

The L^*_n-norm of the sum term is uniformly bounded by $C\varepsilon$: indeed, the $\|v_{n,j}\|_{*n}$ are uniformly bounded by Lemma 2.2.14, while the coefficients satisfy $|\langle v_{n,j}, f^\varepsilon_k\rangle| \leq \|f^\varepsilon_k - f_k\| + \|v_{n,j} - f_j\| < 2\varepsilon$ for large n. By the variational characterization in finite dimensions, and the uniform L^*_n form bound on $\langle\cdot, H_n\cdot\rangle$ (Lemma 2.2.14) together with the uniform bound on $\|\mathcal{P}_nf^\varepsilon_k\|_{*n}$ (Lemma 2.2.17), we then have

$$\limsup \lambda_{n,k} \;\leq\; \limsup \frac{\langle f_{n,k}, H_nf_{n,k}\rangle}{\langle f_{n,k}, f_{n,k}\rangle} \;=\; \limsup \frac{\langle \mathcal{P}_nf^\varepsilon_k, H_n\mathcal{P}_nf^\varepsilon_k\rangle}{\langle \mathcal{P}_nf^\varepsilon_k, \mathcal{P}_nf^\varepsilon_k\rangle} + o_\varepsilon(1), \qquad (2.27)$$

where $o_\varepsilon(1) \to 0$ as $\varepsilon \to 0$. But (2.26) of Lemma 2.2.17 provides $\lim \langle \mathcal{P}_nf^\varepsilon_k, H_n\mathcal{P}_nf^\varepsilon_k\rangle = \mathcal{H}_{Y,W}(f^\varepsilon_k, f^\varepsilon_k)$, so the right hand side of (2.27) is

$$\frac{\mathcal{H}_{Y,W}(f^\varepsilon_k, f^\varepsilon_k)}{\langle f^\varepsilon_k, f^\varepsilon_k\rangle} + o_\varepsilon(1) \;=\; \frac{\mathcal{H}_{Y,W}(f_k, f_k)}{\langle f_k, f_k\rangle} + o_\varepsilon(1) \;=\; \Lambda_k + o_\varepsilon(1).$$

Now letting $\varepsilon \to 0$, we conclude $\limsup \lambda_{n,k} \leq \Lambda_k$.

Thus $\lambda_{n,k} \to \Lambda_k$; Lemmas 2.2.14 and 2.2.17 imply that any subsequence of the $v_{n,k}$ has a further subsequence converging in L^2 to some $g \in L^*$ with (Λ_k, g) an eigenvalue-eigenfunction pair. But then $g = f_k$, and so $v_{n,k} \to_{L^2} f_k$. □

2.3 Application to Wishart and Gaussian models

We now apply Theorem 2.2.11 to prove Theorems 2.1.1 and 2.1.5. The first step is to obtain the tridiagonal forms. Then, after recalling the derivation of the scaling limit at the soft edge, we verify Assumptions 1–3 for certain scalings of the perturbation.

Tridiagonalization

We explain how to tridiagonalize a rank one spiked real Wishart matrix; the algorithm is basically the usual one described by Trotter (1984) with a few careful choices. We restrict for the moment to the case $n \geq p$, but lift this restriction in the Remark 2.3.1 below. For a given $p \times n$ data matrix X we will construct a pair of orthogonal matrices $O \in O(p),\ O' \in O(n)$ so that $W = OXO'$ becomes lower bidiagonal; then X and W have the same singular values and $WW^\dagger$ is a symmetric tridiagonal matrix with the same eigenvalues as $XX^\dagger$. Further, the structure of X and O, O' will be such that the entries of W are independent with explicit known distributions.

We build up O and O' as follows. Let $e_1, \ldots, e_p \in \mathbb{R}^p$ be the standard basis of column vectors and $\tilde{e}_1, \ldots, \tilde{e}_n \in \mathbb{R}^n$ the standard basis of row vectors.

- First, reflect (or rotate) the top row of X into the positive $\tilde{e}_1$ direction via right multiplication by $O_1' \in O(n)$, chosen independently of the other rows. This row becomes $\sqrt{\ell}\,\widetilde{\chi}_n \tilde{e}_1$, where $\widetilde{\chi}_n$ is a $\mathrm{Chi}(n)$ random variable (i.e. distributed as the length of an n-dimensional standard normal vector); the other rows remain independent standard normal vectors, since their distribution is invariant under an independent reflection.
- Next, reflect the first column of XO_1' as follows: leaving $\langle e_1 \rangle$ invariant, reflect the orthogonal $\langle e_2, \ldots, e_p \rangle$ component of the column into the positive e_2 direction via left multiplication by $O_1 \in \{I_1\} \oplus O(p-1)$, chosen independently of the other columns. This component of the column becomes $\chi_{p-1} e_2$ where $\chi_{p-1} \sim \mathrm{Chi}(p-1)$, independent of $\widetilde{\chi}_n$. The same components of the other columns remain independent standard normal vectors, while the first row is untouched.

- Reflect the second row of $O_1 X O_1'$ as follows: leaving $\langle \tilde{e}_1 \rangle$ invariant, reflect the orthogonal component of the row into the positive $\tilde{e}_2$ direction via right multiplication by $O_2' \in \{I_1\} \oplus O(n-1)$, chosen independently of the other rows.
- Reflect the second column of $O_1 X O_1' O_2'$ as follows: leaving $\langle e_1, e_2 \rangle$ invariant, reflect the orthogonal component of the column into the positive e_3 direction via left multiplication by $O_2 \in \{I_2\} \oplus O(p-2)$, chosen independently of the other columns.
- Continue in this way, alternately reflecting rows and columns while leaving the results of previous steps untouched.

The result is that with $O' = O_1' \cdots O_p'$ and $O = O_{p-1} \cdots O_1$ we have

$$W = OXO' = \begin{bmatrix} \sqrt{\ell}\widetilde{\chi}_n & & & & & \\ \chi_{p-1} & \widetilde{\chi}_{n-1} & & & & \\ & \ddots & \ddots & & & \\ & & \chi_2 & \widetilde{\chi}_{n-p+2} & & \\ & & & \chi_1 & \widetilde{\chi}_{n-p+1} \end{bmatrix},$$

where $\{\widetilde{\chi}_{n-j}\}_{j=0}^{p-1}$ and $\{\chi_{p-j}\}_{j=1}^{p-1}$ are independent Chi random variables of parameters given by their indices. We have truncated the $n-p$ rightmost columns of zeros to obtain a $p \times p$ matrix, leaving the product $WW^\dagger$ unchanged. We will actually work with $W^\dagger W$ below, which has the same eigenvalues.

Remark 2.3.1. Attempting the above procedure in the case $n < p$ produces a lower bidiagonal matrix W with $n+1$ nonzero rows. The matrix $W^\dagger W$ is now $n \times n$, has the same *nonzero* eigenvalues as $XX^\dagger$, and looks just like it does in the $n \geq p$ case except for a discrepancy in the bottom-right corner. The two cases may in fact be unified if one agrees that $\chi_0 = 0$; then W is $(n \wedge p + 1) \times (n \wedge p)$ and has the form (2.2) with $\beta = 1$, while $W^\dagger W$ is $(n \wedge p) \times (n \wedge p)$.

The same algorithm will tridiagonalize a rank one spiked complex (resp. quaternionic) Wishart matrix by unitary (resp. symplectic or hyperunitary) conjugations. The lower bidiagonal matrix will be $W_{n,p}^{\beta,\ell}$ from (2.2) with $\beta = 2$ (resp. 4).

The perturbed GOE/GUE/GSE ensembles are even easier to tridiagonalize; as in the Wishart case, the usual procedure of Trotter (1984) works without modification. Starting with an $n \times n$ GOE matrix M with a perturbation in the (1,1) entry, the upshot is that for certain $O_1, \ldots, O_{n-1}$ with $O_j \in \{I_j\} \oplus O(n-j)$ the conjugated matrix $O_{n-1} \cdots O_1 M O_1^\dagger \cdots O_{n-1}^\dagger$ has the form (2.5) with $\beta = 1$. We do not detail it further here.

Scaling limit

Consider the ℓ-spiked β-Laguerre ensemble $S = W^\dagger W$ with $W = W_{n,p} = W_{n,p}^{\beta,\ell_{n,p}}$ as in (2.2), recalling that $S_{n,p}$ is $(n \wedge p) \times (n \wedge p)$. The diagonal and off-diagonal processes of βS are

$$\ell_{n,p}\widetilde{\chi}^2_{\beta n} + \chi^2_{\beta(p-1)}, \quad \widetilde{\chi}^2_{\beta(n-1)} + \chi^2_{\beta(p-2)}, \quad \widetilde{\chi}^2_{\beta(n-2)} + \chi^2_{\beta(p-3)}, \quad \cdots$$
$$\widetilde{\chi}_{\beta(n-1)}\chi_{\beta(p-1)}, \quad \widetilde{\chi}_{\beta(n-2)}\chi_{\beta(p-2)}, \quad \cdots$$

respectively. The usual centering and rescaling for fluctuations at the soft edge—as well as the operator limit itself—can be predicted using the approximations

$$\chi_k \approx \sqrt{k} + \sqrt{1/2}\, g, \qquad \chi^2_k \approx k + \sqrt{2k}\, g,$$

valid for k large, where g is a suitably coupled standard Gaussian. We briefly reproduce the heuristic argument.

To leading order, the top-left corner of S has $n + p$ on the diagonal and $\sqrt{np}$ on the off-diagonal. So the top-left corner of

$$\frac{1}{\sqrt{np}}\Big(S - \big(\sqrt{n} + \sqrt{p}\big)^2 I\Big)$$

is approximately an unscaled discrete Laplacian. If time is scaled by m^{-1}, space has to be scaled by m^2 for this to converge to $\frac{d^2}{dx^2}$. The next order terms for the j'th diagonal and off-diagonal entries of S, where $j \ll n \wedge p$, are respectively

$$\tfrac{1}{\sqrt{\beta}}\big(\sqrt{2n}\,\widetilde{g}_{n-j+1} + \sqrt{2p}\,g_{p-j} - 2j\big),$$
$$\tfrac{1}{\sqrt{\beta}}\big(\sqrt{p/2}\,\widetilde{g}_{n-j} + \sqrt{n/2}\,g_{p-j} - 1/2(\sqrt{p/n} + \sqrt{n/p})j\big).$$

(we have indexed the g's to match the corresponding χ's). The total noise per unit (unscaled) time is like $\frac{2}{\sqrt{\beta}}\big(\sqrt{n}+\sqrt{p}\big)g$; convergence to $\frac{2}{\sqrt{\beta}}$ times standard Gaussian white noise b'_x then requires $\big(\sqrt{n} + \sqrt{p}\big)m_n^2/\sqrt{np} = m^{1/2}$. The averaged part of the potential requires $\big(2+\sqrt{p/n}+\sqrt{n/p}\big)m^2/\sqrt{np} = m^{-1}$ to converge to the function $-x$. Fortunately these two scaling requirements match perfectly; we set

$$m_{n,p} = \left(\frac{\sqrt{np}}{\sqrt{n} + \sqrt{p}}\right)^{2/3}, \quad H_{n,p} = \frac{m_{n,p}^2}{\sqrt{np}}\left(\big(\sqrt{n} + \sqrt{p}\big)^2 I_{n\wedge p} - S_{n,p}\right)$$

and set the integrated limiting potential to

$$y(x) = \tfrac{1}{2}x^2 + \tfrac{2}{\sqrt{\beta}}b_x$$

where b_x is a standard Brownian motion. Note that

$$2^{-2/3}(n \wedge p)^{1/3} \le m \le (n \wedge p)^{1/3},$$

so the conditions $m \to \infty$, $m = o(n \wedge p)$ are met by merely having $n, p \to \infty$ together.

We now carefully decompose $H_{n,p}$ as in (2.11). In (2.12),(2.13) there is a little freedom between $y_{n,1;1}$ and w_n, but only in to an additive constant in $y_{n,1}$ that tends to zero in probability anyway. Thus we may as well set $y_{n,1;1} = 0$ to fix w_n and $y_{n,i}$. Assumptions 1 and 2 (the CLT (2.14) and required tightness (2.15)–(2.17) for the potential terms $y_{n,i}$) are then verified as in the final subsection of RRV; Section 3.4 of the next chapter does it in a more general setting with some clarifications and minor corrections.

It remains to consider Assumption 3. We have

$$w_n = m_{n,p}\left(1 + \sqrt{\frac{n}{p}}\left(1 - \ell_{n,p}\frac{\widetilde{\chi}^2_{\beta n}}{\beta n}\right) + \sqrt{\frac{p}{n}}\left(1 - \frac{\chi^2_{\beta(p-1)}}{\beta p}\right)\right).$$

First order heuristics suggest we take $\ell_{n,p}$ to satisfy

$$\overline{w}_n = m_{n,p}\left(1 + \sqrt{\frac{n}{p}}\,(1 - \ell_{n,p})\right) \to w \in (-\infty, \infty] \qquad \text{as } n \wedge p \to \infty$$

as in (2.4). We want to show that, in this case, $w_n \to w$ in probability; it is certainly enough to show that $w_n - \overline{w}_n \to 0$ in probability.

Second order heuristics say the error terms are on the order $(n \wedge p)^{-1/6}$ or $m^{-1/2}$, and L^2 estimates easily provide the rigour. All we need is that χ^2_k has mean k and variance $2k$. We have

$$w_n - \overline{w}_n = -\frac{m\ell}{\beta\sqrt{np}}\left(\chi^2_{\beta n} - \beta n\right) + \frac{m}{\beta\sqrt{np}}\left(\beta(p-1) - \chi^2_{\beta(p-1)}\right) + \frac{m}{\sqrt{np}}.$$

Using that $\ell \le 1 + 2\sqrt{p/n}$, the mean square of the first term is $O(m^2/p + m^2/n)$, which is $O(m^{-1})$. The mean square of the second term is $O(m^2/n)$, again $O(m^{-1})$. The last term is negligible. This completes the proof of Theorem 2.1.1.

Turning now to the perturbed β-Hermite ensemble, take $G_n = G_n^{\beta,\mu_n}$ as in (2.5). With heuristic motivation similar to that in the previous proof, set

$$m_n = n^{1/3}, \qquad H_n = \frac{m_n^2}{\sqrt{n}}\left(2\sqrt{n}I_n - G_n\right)$$

and $y(x)$ as before. Decompose H_n as in (2.11). Again, the verification of Assumptions 1 and 2 on $y_{n,i}$ proceeds as in RRV (Lemmas 6.2, 6.3) or Section 3.4. Moving on to Assumption 3, we have

$$w_n = m_n\left(1 - \left(\mu_n + \sqrt{2/\beta n}\, g_1\right)\right).$$

Putting

$$\overline{w}_n = m_n\,(1-\mu_n)$$

as in (2.6), the difference is $w_n - \overline{w}_n = -n^{-1/6}\sqrt{2/\beta}\,g_1$. It follows that $w_n - \overline{w}_n \to 0$ in probability, which completes the proof of Theorem 2.1.5.

2.4 Alternative characterizations of the laws

In this section we prove Theorem 2.1.7 and its extension to higher eigenvalues.

Diffusion

The diffusion characterization is developed in RRV; we recall the important facts here with a general initial condition. For a fully rigorous treatment see Appendix A. The starting point is an application of the classical Riccati map $p = f'/f$ to the eigenvalue equation (2.20), or rigorously to (2.24); the result is the first order differential equation

$$p'(x) = x - \lambda + \tfrac{2}{\sqrt{\beta}}b'(x) - p^2(x) \tag{2.28}$$

understood also in the integrated sense. The boundary condition at the origin becomes the initial value

$$p(0) = w,$$

and a zero of f would have p explode to $-\infty$ and immediately restart at $+\infty$.

One can in fact construct the solution for any $\lambda \in \mathbb{R}$. One way to see this is to introduce the variable $q(x) = p(x) + \frac{2}{\sqrt{\beta}}b(x)$; the ODE

$$q' = x - \lambda - \left(q + \tfrac{2}{\sqrt{\beta}}b\right)^2 \tag{2.29}$$

is classical and the Picard existence and uniqueness theorem applies. Although solutions can explode to $-\infty$ in finite time, this is not a problem if we consider the values on the projective line. Behaviour through ∞ can then be understood in the other coordinate $\tilde{q} = 1/q$, which evolves as

$$\tilde{q}' = \left(1 + \tfrac{2}{\sqrt{\beta}}b\tilde{q}\right)^2 - (x-\lambda)\tilde{q}^2;$$

in particular, $\tilde{q}' = 1$ whenever $\tilde{q} = 0$. The solution can thus be continued for all time. Moreover, it depends monotonically and continuously on the the parameter λ, uniformly

on compact time-intervals with respect to the topology of the projective line. Following classical Sturm oscillation theory one can argue that almost surely, for all $\lambda \in \mathbb{R}$, *the number of eigenvalues strictly less than λ equals the number of explosions of p on $\mathbb{R}_+$.*

For a fixed λ, the Riccati equation (2.28) may also be understood in the Itō sense; by translation equivariance the time-shift $x \mapsto x - \lambda$ produces the same path measure as the Itō diffusion (2.7) started at time $x_0 = -\lambda$. Writing $\kappa_{(x_0,w_0)}$ for the distribution of the first explosion time of p_x under $\mathbf{P}_{(x_0,w_0)}$—an improper distribution with some mass on ∞—we have $\mathbf{P}_{\beta,w}(\Lambda_0 < \lambda) = \kappa_{(-\lambda,w)}(\mathbb{R})$ or $F_{\beta,w}(x) = \kappa_{(x,w)}(\{\infty\})$ as in (2.8). More generally, the strong Markov property gives

$$\mathbf{P}_{\beta,w}(-\Lambda_{k-1} > x) = \int_{\mathbb{R}^k} \kappa_{(x,w)}(dx_1)\, \kappa_{(x_1,\infty)}(dx_2) \cdots \kappa_{(x_{k-1},\infty)}(dx_k). \tag{2.30}$$

The stated path properties of (2.7) appear also in RRV (Propositions 3.7 and 3.9).

Boundary value problem

Briefly, the boundary value problem is just the Kolmogorov backward equation for a hitting probability of the diffusion. We assume the diffusion representation $F_{\beta,w}(x) = \kappa_{(x,w)}(\{\infty\})$ for the distribution of $-\Lambda_0$.

Lemma 2.4.1. *For each fixed x, $F_{\beta,w}(x)$ is nondecreasing and continuous in $w \in (\infty, \infty]$ and tends to zero as $w \to -\infty$.*

Remark 2.4.2. There are in fact almost-sure counterparts of these assertions that describe how Λ_0 depends on w for each Brownian path, but we do not need them here.

Proof. The monotonicity is a consequence of uniqueness of the diffusion path from each space-time point: two paths started from (x, w_0) and (x, w_1) with $w_0 < w_1$ never cross, so if the upper path explodes to $-\infty$ then the lower path must do so as well. The continuity is a general property of statistics of diffusions: $\kappa_{(x,p_x)}(\{\infty\})$ is a martingale, so $F_{\beta,w}(x)$ is in fact space-time harmonic. (Again, the behaviour at $w = +\infty$ may be understood by changing coordinates.)

The final assertion is that for fixed x_0 explosion becomes certain as $w \to -\infty$. It may be verified by a domination argument involving the ODE (2.29) (time-shifted as above so that $\lambda = 0$ and the initial time is x_0), whose paths explode simultaneously with those of (2.7). Given $\varepsilon > 0$, let M be such that $\mathbf{P}(\sup_{x\in[x_0,x_0+1]} |b_x| > M) < \varepsilon$. It is easy to check that for r_0 sufficiently negative, the solution of $r' = x - (r + M)^2$ with

initial value $r(x_0) = r_0$ explodes to $-\infty$ before time $x_0 + 1$. Now consider the solution of $q' = x - (q+b)^2$ with $q(x_0) \le r_0 \le -M$. With probability $1-\varepsilon$ we have $q'(x) \le r'(x)$ whenever $q(x) = r(x)$, so the paths never cross and q explodes as well. □

Proof of Theorem 2.1.7 (ii). Writing L for the space-time generator of the SDE (2.7), the PDE (2.9) is simply the equation $LF = 0$. Therefore the hitting probability $F(x,w) = F_{\beta,w}(x)$ satisfies the PDE. The boundary behaviour (2.10) follows from Lemma 2.4.1 and the fact that $F(\cdot, w)$ is a distribution function for each w. Specifically, the lower part of the boundary behaviour follows from the fact that $F(x,w)$ is increasing in x and $F(x,w) \to 0$ as $w \to -\infty$ for each x. The upper part follows from the fact that $F(x,w)$ is increasing in w and $F(x,w) \to 1$ for fixed w as $x \to \infty$.

Toward uniqueness, suppose $\tilde{F}(x,w)$ is another bounded solution of (2.9),(2.10). By the PDE, $\tilde{F}(x, p_x)$ is a local martingale under $\mathbf{P}_{(x_0,w_0)}$ and thus a bounded martingale. Let T be the lifetime of the diffusion; optional stopping gives $\tilde{F}(x,w) = \mathbf{E}_{(x,w)} \tilde{F}(T \wedge t, p_{T\wedge t})$ for all $t \ge x$. Taking $t \to \infty$, we conclude by bounded convergence, the boundary behaviour of $\tilde{F}$ and the stated path properties of the diffusion that $\tilde{F}(x,w)$ is the non-explosion probability. That is, $\tilde{F} = F$. □

As promised, we indicate how the laws of the higher eigenvalues $\Lambda_1, \Lambda_2, \ldots$ may be characterized in terms of the PDE (2.9). The characterization is inductive and follows from (2.30) by reasoning just as in the preceding proof.

Theorem 2.4.3. *Let $F_{(0)}(x,w) = \mathbf{P}_{\beta,w}(-\Lambda_0 < x)$. For each $k = 1, 2, \ldots$, the boundary value problem*

$$\frac{\partial F}{\partial x} + \frac{2}{\beta}\frac{\partial^2 F}{\partial w^2} + \big(x - w^2\big)\frac{\partial F}{\partial w} = 0 \qquad \text{for } (x,w) \in \mathbb{R}^2,$$

$$F(x,w) \to \begin{cases} 1 & \text{as } x, w \to \infty \text{ together,} \\ F_{(k-1)}(x_0, +\infty) & \text{as } w \to -\infty \text{ while } x \to x_0 \in \mathbb{R} \end{cases}$$

has a unique bounded solution $F_{(k)}$, and we have $\mathbf{P}_{\beta,w}(-\Lambda_k < x) = F_{(k)}(x,w)$ for $w \in (-\infty, \infty)$; further, $\mathbf{P}_{\beta,\infty}(-\Lambda_k < x) = \lim_{w\to\infty} F_{(k)}(x,w)$.

Chapter 3

Several spikes

3.1 Introduction

We refer the reader to the introduction of Chapter 2 for background on the spiked Wishart model introduced by Johnstone (2001) in the context of high-dimensional data analysis, especially the phase transition phenomenon that Baik, Ben Arous and Péché (2005) (**BBP**) described in the complex case. We refer the reader to the same place for background on the stochastic Airy operator introduced by Sutton (2005), Edelman and Sutton (2007), and its development byRamírez, Rider and Virág (2011) (**RRV**) that we build on.

In Chapter 2 we considered rank one spiked real/complex/quaternion Wishart matrices and additive rank one perturbations of the Gaussian orthogonal, unitary and symplectic ensembles. We introduced their general β analogues in terms of "spiked" versions of the tridiagonal ensembles of Dumitriu and Edelman (2002) and extended the RRV technology to describe the soft-edge scaling limit in terms of the stochastic Airy operator

$$-\frac{d^2}{dx^2} + \tfrac{2}{\sqrt{\beta}} b'_x + x$$

on $L^2(\mathbb{R}_+)$ with a boundary condition depending on the spike. The boundary condition changes from Dirichlet $f(0) = 0$ to Neumann/Robin $f'(0) = wf(0)$ at the onset of the BBP phase transition, with $w \in \mathbb{R}$ representing a scaling parameter for perturbations in a "critical window". The resulting largest eigenvalue laws form a one-parameter family of deformations of Tracy-Widom(β), naturally generalizing the characterization of RRV in terms of the ground state of this random Schrödinger operator. We went on to characterize the limit laws in terms of the diffusion from RRV and further in terms of an

associated second-order linear parabolic PDE.

Here we deal with r "spikes", or general bounded-rank perturbations of Gaussian and Wishart matrices. To do so we introduce a new "canonical form for perturbations in a fixed subspace", a $(2r+1)$-diagonal band form that has a purely algebraic interpretation. It generalizes the Dimitriu–Edelman forms and is able to handle rank r perturbations. We then develop a generalization of the methods of RRV and Chapter 2 to a matrix-valued setting: block tridiagonal matrices converge to a half-line Schrödinger operator with matrix-valued potential, the spikes once again appearing in the boundary condition. We treat the real, complex and quaternion ($\beta = 1, 2, 4$) cases simultaneously. Once again, even the existence of a near-critical soft-edge limit is new off $\beta = 2$. Unlike in Chapter 2, however, we do not define a general β version of either matrix model, nor of the limiting operator; we will see that in the higher rank cases these objects do not readily admit a β-generalization.

Dyson's Brownian motion makes a surprise appearance, providing nice SDE and PDE characterizations of the limit laws—r parameter deformations of Tracy-Widom(β)—in which β reappears as a simple parameter. The derivation makes use of the matrix versions of classical Sturm oscillation theory and the Riccati transformation.

We highlight two more features of our approach beyond the novelty of bypassing joint densities and handling $\beta = 1, 2, 4$ together. First, we treat the perturbation as a *parameter*. By this we mean that all perturbations in a fixed subspace are considered jointly (on the same probability space); this picture is carried through to the limit, which is therefore a family of point processes parametrized by an $r \times r$ matrix. Second, we allow more general scalings than those considered in BBP. Most importantly, in the Wishart case we do not require the two dimensional parameters n, p to have a positive limiting ratio but rather allow them to tend to infinity together arbitrarily.

To state our results we introduce some objects and notation that will be used throughout the chapter.

Let $\mathbb{F} = \mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$ and $\beta = 1, 2$ or 4, respectively. A **standard $\mathbb{F}$ Gaussian** $Z \sim \mathbb{F}N(0,1)$ is an $\mathbb{F}$-valued random variable described in terms of independent real Gaussians $g_1, \ldots, g_\beta \sim N(0,1)$ as g_1 for $\mathbb{F} = \mathbb{R}$, $(g_1 + g_2 i)/\sqrt{2}$ for $\mathbb{F} = \mathbb{C}$, and $(g_1 + g_2 i + g_3 j + g_4 k)/2$ for $\mathbb{F} = \mathbb{H}$. Note that in each case $\mathbf{E}\,|Z|^2 = 1$ and $uZ \sim \mathbb{F}N(0,1)$ for $u \in \mathbb{F}$ with $|u|^2 = u^* u = 1$.

The space of column vectors $\mathbb{F}^n$ is endowed with the standard inner product $u^\dagger v$ and associated norm $|u|^2 = u^\dagger u$ (we reserve double bars for function spaces). Write $\mathbb{F}N_n(0, I)$

for a vector of independent standard $\mathbb{F}$ Gaussians. With $\Sigma \in M_n(\mathbb{F})$ positive definite, we write $Z \sim \mathbb{F}N_n(0, \Sigma)$ for $Z = \Sigma^{1/2} Z_0$ with $Z_0 \sim \mathbb{F}N_n(0, I)$.

Define the **unitary group** $U_n(\mathbb{F}) = \{U \in \mathbb{F}^{n\times n} : U^\dagger U = I\}$, better known as the orthogonal, unitary or symplectic group for $\mathbb{F} = \mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ respectively. It acts on $\mathbb{F}^n$ by left multiplication, on which the distribution $\mathbb{F}N_n(0, I)$ is invariant. Write $M_n(\mathbb{F}) = \{A \in \mathbb{F}^{n\times n} : A^\dagger = A\}$ for the **self-adjoint matrices**, also known as real symmetric, complex hermitian or quaternion self-dual. $U_n(\mathbb{F})$ acts on $M_n(\mathbb{F})$ by conjugation.

The **Gaussian orthogonal/unitary/symplectic ensemble** (GO/U/SE) is the probability measure on $M_n(\mathbb{F})$ described by $A = (X + X^\dagger)/\sqrt{2}$ where X is an $n \times n$ matrix of independent $\mathbb{F}N(0, 1)$ entries. The distribution is invariant under the unitary action. Furthermore, the algebraically independent entries A_{ij}, $i \geq j$ are statistically independent. (Together, this invariance and independence characterizes the distribution up to a scale factorr.) For an entry-wise description, the diagonal entries are distributed as $N(0, 2/\beta)$ while the off-diagonal entries are $\mathbb{F}N(0, 1)$.

Fixing a positive integer r, we study **rank** r **additive perturbations** $A = A_0 + P$ of a GO/U/SE matrix A_0, where $P = \tilde{P} \oplus 0_{n-r}$ with $\tilde{P} \in M_r(\mathbb{F})$ nonrandom. We will be interested in the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ of A. Of course for a single P their distribution depends only on the eigenvalues of P, but we consider them jointly over all $\tilde{P}$.

We also consider **real/complex/quaternion Wishart matrices**. These are random nonnegative matrices in $M_p(\mathbb{F})$ given by $XX^\dagger$ where the **data matrix** X is $p \times n$ with n independent $\mathbb{F}N_p(0, \Sigma)$ columns. We speak of a p-**variate** Wishart with n **degrees of freedom** and $p \times p$ **covariance** $\Sigma > 0$. Since we are interested in the *nonzero* eigenvalues $\lambda_1 \geq \cdots \geq \lambda_{n\wedge p}$, we can equally well consider $X^\dagger X$. The distribution of $X^\dagger X$ may also be described as $X_0^\dagger \Sigma X_0$ where X_0 is a $p \times n$ matrix of independent $\mathbb{F}N(0, 1)$ entries. The case $\Sigma = I$ is referred to as the **null case**. We study the **rank** r **spiked case** where $\Sigma = \tilde{\Sigma} \oplus I_{p-r}$ with $\tilde{\Sigma} \in M_r(\mathbb{F})$ nonrandom. Once again the eigenvalue distribution depends only on the eigenvalues of Σ, but we consider the spectrum jointly as $\tilde{\Sigma}$ varies.

Our starting point is a new banded or multidiagonal form introduced in Section 3.2, ideally suited to the types of perturbations we consider. It is defined for almost every matrix $A \in M_n(\mathbb{F})$; given vectors $v_1, \ldots, v_r \in \mathbb{F}^n$, the new basis may be obtained by

applying the Gram-Schmidt process to the first n vectors of the sequence

$$v_1, \ldots, v_r, Av_1, \ldots, Av_r, A^2v_1, \ldots, A^2v_r, \ldots.$$

The result is a $(2r+1)$-diagonal matrix with positive outer diagonals. For Gaussian and null Wishart ensembles, the change of basis interacts well with the Gaussian structure; this observation goes back to Trotter (1984) in the $r = 1$ case. In the GO/U/SE case $v_1, \ldots, v_r$ are simply the initial coordinate basis vectors, while in the Wishart case they represent the initial rows of the data matrix X. As in Chapter 2, the key observation is that the perturbations commute with the change of basis.

For the (unperturbed) Gaussian ensembles, the band form looks like

$$\begin{bmatrix}
\tilde{g} & g^* & \cdots & g^* & \chi & & & \\
g & \tilde{g} & g^* & \cdots & g^* & \chi & & \\
\vdots & g & \tilde{g} & g^* & \cdots & g^* & \chi & \\
g & \vdots & g & \ddots & \ddots & & \ddots & \ddots \\
\chi & g & \vdots & \ddots & & & & \\
 & \chi & g & & & & & \\
 & & \chi & \ddots & & & & \\
 & & & \ddots & & & &
\end{bmatrix}$$

where the entries are independent random variables up to the †-symmetry with $\tilde{g} \sim N(0, 2/\beta)$, $g \sim \mathbb{F}N(0,1)$, and $\chi \sim \frac{1}{\sqrt{\beta}}\mathrm{Chi}((n-r-k)\beta)$, $k = 0, 1, 2, \ldots$ going down the matrix. (Recall that if $Z \sim \mathbb{R}N_m(0, I)$ then $|Z| \sim \mathrm{Chi}(m)$.) For the null Wishart ensemble, the form is best described as follows. One first obtains a lower $(r+1)$-diagonal form for the data matix X whose nonzero *singular values* are the same as those of X. It looks like

$$\begin{bmatrix}
\tilde{\chi} & & & \\
g & \tilde{\chi} & & \\
\vdots & g & \tilde{\chi} & \\
g & \vdots & g & \ddots \\
\chi & g & \vdots & \ddots \\
 & \chi & g & \\
 & & \chi & \ddots \\
 & & & \ddots
\end{bmatrix}$$

where the entries are independent random variables with $g \sim \mathbb{F}N(0,1)$, $\tilde{\chi} \sim \frac{1}{\sqrt{\beta}}\mathrm{Chi}((n-k)\beta)$ and $\chi \sim \frac{1}{\sqrt{\beta}}\mathrm{Chi}((n-r-k)\beta)$, $k = 0, 1, 2, \ldots$ going down the matrix. One then forms its multiplicative symmetrization, a $(2r+1)$-diagonal matrix with the same nonzero eigenvalues as X. In both cases the perturbations appear in the upper-left $r \times r$ block. Section 3.2 provides derivations. The obstacle to β-generalization at this level is the presence of $\mathbb{F}$ Gaussians in the intermediate diagonals.

Proceeding with an analogue of the RRV convergence result hinges on reinterpreting these forms as *block tridiagonal* with $r \times r$ blocks. In Section 3.3 we develop an $M_r(\mathbb{F})$-valued analogue of the RRV technology, providing general conditions under which the principal eigenvalues and corresponding eigenvectors of such a random block tridiagonal matrix converge to a those of a continuum half-line random Schrödinger operator with matrix-valued potential. As in Chapter 2 we allow for a general boundary condition at the origin.

In Section 3.4 we apply this result to the band forms just described, proving a process central limit theorem for the potential and verifying the required tightness assumptions. The limiting operator turns out to be a multidimensional version of the stochastic Airy operator, which we now describe.

First, a **standard $\mathbb{F}$ Brownian motion** $\{b_t\}_{t\geq 0}$ is a continuous $\mathbb{F}$-valued random process with $b_0 = 0$ and independent increments $b_t - b_s \sim \mathbb{F}N(0, t-s)$. (It can be described in terms of $\beta = 1$, 2 or 4 independent standard real Brownian motions.) A **standard matrix Brownian motion** $\{B_t\}_{t\geq 0}$ has continuous $M_n(\mathbb{F})$-valued paths with $B_0 = 0$ and independent increments $B_t - B_s$ distributed as $\sqrt{t-s}$ times a GO/U/SE. The diagonal processes are thus $\sqrt{2/\beta}$ times standard real Brownian motions while the off-diagonal processes are standard $\mathbb{F}$ Brownian motions, mutually independent up to symmetry.

Finally, we define the **multivariate stochastic Airy operator**. Operating on the vector-valued function space $L^2(\mathbb{R}_+, \mathbb{F}^r)$ with inner product $\langle f, g\rangle = \int_0^\infty f^\dagger g$ and associated norm $\|f\|^2 = \int_0^\infty |f|^2$, it is the random Schrödinger operator

$$\mathcal{H}_\beta = -\frac{d^2}{dx^2} + \sqrt{2}B_x' + rx \tag{3.1}$$

where B_x' is "standard matrix white noise", the derivative of a standard matrix Brownian motion, and rx is scalar. (Here again β is restricted to the classical values, as the noise term lacks a straightforward β-generalization.) The potential is thus the derivative of a continuous matrix-valued function; rigorous definitions will appear in Section 3.3 in a

more general setting.

For now it is enough to know that, together with a general self-adjoint boundary condition

$$f'(0) = Wf(0), \tag{3.2}$$

the multivariate stochastic Airy operator is bounded below with purely discrete spectrum given by a variational principle. Here $W \in M_r(\mathbb{F})$; actually, writing the spectral decomposition $W = \sum_{i=1}^r w_i u_i u_i^\dagger$, we formally allow $w_i \in (-\infty, \infty]$. Writing $f_i = u_i^\dagger f$, (3.2) is then to be interpreted as

$$\begin{aligned} f_i'(0) &= w_i f_i(0) \qquad && \text{for } w_i \in \mathbb{R} \\ f_i(0) &= 0 && \text{for } w_i = +\infty. \end{aligned}$$

We write $W \in M_r^*(\mathbb{F})$ for this extended set and $\mathcal{H}_{\beta,W}$ for (3.1) together with (3.2).

For concreteness, we record that the eigenvalues $\Lambda_0 \le \Lambda_1 \le \dots$ and corresponding eigenfunctions $f_0, f_1, \dots$ of $\mathcal{H}_{\beta,W}$ are given respectively by the minimum and any minimizer in the recursive variational problem

$$\inf_{\substack{f \in L^2(\mathbb{R}_+) \\ \|f\|=1,\ f \perp f_0,\dots,f_{k-1}}} \int_0^\infty \big(|f'|^2 + rx\,|f|^2\big)dx + f(0)^\dagger W f(0) + \tfrac{2}{\sqrt{\beta}} \int_0^\infty f^\dagger dB_x f.$$

Here canidates f are only considered if the first integral and boundary term are finite; the stochastic integral can then be defined pathwise via integration by parts. The eigenvalues and eigenfunctions are thus jointly defined random processes indexed over W.

Remark 3.1.1. We note one important property of the eigenvalue processes, namely the *pathwise monotonicity* of Λ_k in W with respect to the usual matrix partial order. This is immediate from the variational characterization and the fact that the objective functional is monotone in W. (For the higher eigenvalues it is most apparent from the standard min-max formulation of the variational problem.)

We can now state the main convergence results. As outlined, Sections 3.2–3.4 furnish the proofs. One last shorthand: when we write that a sequence $W_n \in M_r(\mathbb{F})$ tends to $W \in M_r^*(\mathbb{F})$, we mean the following. Writing $W = \sum_{i=1}^r w_i u_i u_i^\dagger$ with $w_i \in (-\infty, \infty]$, one has $W_n = \sum_{i=1}^r w_{n,i} u_i u_i^\dagger$ with $w_{n,i} \in \mathbb{R}$ satisfying $w_{n,i} \to w_i$ for each i. In other words, the matrices are simultaneously diagonal and the eigenvalues tend to the corresponding limits.

Theorem 3.1.2. *Let $A = A_0 + \sqrt{n}P_n$ where A_0 is an $n \times n$ GO/U/SE matrix and $P_n = \tilde{P}_n \oplus 0_{n-r}$ with $\tilde{P}_n \in M_r(\mathbb{F})$, and let $\lambda_1 \geq \cdots \geq \lambda_n$ be its eigenvalues. If*

$$n^{1/3}\big(1 - \tilde{P}_n\big) \;\to\; W \in M_r^*(\mathbb{F}) \qquad \text{as } n \to \infty$$

then, jointly for $k = 1, 2, \ldots$ in the sense of finite-dimensional distributions,

$$n^{1/6}\big(\lambda_k - 2\sqrt{n}\big) \;\Rightarrow\; -\Lambda_{k-1} \qquad \text{as } n \to \infty$$

where $\Lambda_0 \leq \Lambda_1 \leq \ldots$ are the eigenvalues of $\mathcal{H}_{\beta,W}$. Convergence holds jointly over $\{P_n\}, W$ satisfying the condition.

Theorem 3.1.3. *Consider a p-variate real/complex/quaternion Wishart matrix with n degrees of freedom and spiked covariance $\Sigma_{n,p} = \tilde{\Sigma}_{n,p} \oplus I_{p-r} > 0$ with $\tilde{\Sigma}_{n,p} \in M_r(\mathbb{F})$, and let $\lambda_1 \geq \cdots \geq \lambda_{n\wedge p}$ be its nonzero eigenvalues. Writing $m_{n,p} = (n^{-1/2} + p^{-1/2})^{-2/3}$, if*

$$m_{n,p}\left(1 - \sqrt{n/p}\big(\tilde{\Sigma}_{n,p} - 1\big)\right) \;\to\; W \in M_r^*(\mathbb{F}) \qquad \text{as } n \to \infty$$

then, jointly for $k = 1, 2, \ldots$ in the sense of finite-dimensional distributions,

$$\frac{m_{n,p}^2}{\sqrt{np}}\Big(\lambda_k - (\sqrt{n} + \sqrt{p})^2\Big) \;\Rightarrow\; -\Lambda_{k-1} \qquad \text{as } n \to \infty$$

where $\Lambda_0 \leq \Lambda_1 \leq \ldots$ are the eigenvalues of $\mathcal{H}_{\beta,W}$. Convergence holds jointly over $\{\Sigma_{n,p}\}, W$ satisfying the condition.

Remark 3.1.4. In the band basis described above, we also have joint convergence of the corresponding eigenvectors to the eigenfunctions of $\mathcal{H}_{\beta,W}$. In detail, the eigenvectors should be embedded in $L^2(\mathbb{R}_+)$ as step functions with step width $n^{-1/3}$ in the Gaussian case and $m_{n,p}^{-1}$ in the Wishart case, and convergence is in law with respect to the L^2 norm topology. To be precise, one should use either subsequences or spectral projections; one could also formulate the joint eigenvalue-eigenvector convergence in terms of the norm resolvent topology. See Theorem 3.3.9 and the remark that follows.

We now give the two promised alternative characterizations of the limiting eigenvalue laws. Fix $\beta = 1, 2, 4$ and $W \in M_r^*(\mathbb{F})$ with eigenvalues $-\infty < w_1 \leq \ldots \leq w_r \leq \infty$. Writing $\mathbf{P}$ for the probability measure associated with $\mathcal{H}_{\beta,W}$ and its spectrum $\{\Lambda_0 \leq \Lambda_1 \leq \ldots\}$, let

$$F_\beta^k(x; w_1, \ldots, w_r) = \mathbf{P}(-\Lambda_k \leq x)$$

for $k = 0, 1, \ldots$. Write simply $F_\beta = F_\beta^0$ for the ground state distribution (limiting largest eigenvalue law). Once again, the generalization from Chapter 2 is not straightforward. The proofs are contained in Section 3.5.

Theorem 3.1.5. *Let $\mathbf{P}_{x_0,(w_1,\dots,w_r)}$ be the measure on paths $(p_1,\dots,p_r) : [x_0,\infty) \to (-\infty,\infty]^r$ determined by the coupled diffusions*

$$dp_i = \tfrac{2}{\sqrt{\beta}} db_i + \left(rx - p_i^2 + \sum_{j\neq i} \frac{2}{p_i - p_j} \right) dx \tag{3.3}$$

with initial conditions $p_i(x_0) = w_i$ and entering into $\{p_1 < \ldots < p_r\}$, where $b_1,\dots,b_r$ are independent standard Brownian motions; particles p_i may explode to $-\infty$ in finite time whereupon they are restarted at $+\infty$. Then

$$F_\beta(x; w_1,\dots,w_r) = \mathbf{P}_{x/r,(w_1,\dots,w_r)}(\text{no explosions}). \tag{3.4}$$

More generally,

$$F_\beta^k(x; w_1,\dots,w_r) = \mathbf{P}_{x/r,(w_1,\dots,w_r)}(\text{at most } k \text{ explosions}). \tag{3.5}$$

We describe the diffusion more carefully in Section 3.5, asserting that it determines a law on paths valued in an appropriate space. Probabilistic arguments lead to the following reformulation in terms of its generator.

Theorem 3.1.6. *$F_\beta(x; w_1,\dots,w_r)$ is the unique bounded function $F : \mathbb{R} \times \mathbb{R}^r \to \mathbb{R}$ symmetric with respect to permutation of $w_1,\dots,w_r$ that satisfies the PDE*

$$r\frac{\partial F}{\partial x} + \sum_{i=1}^r \left(\frac{2}{\beta}\frac{\partial^2 F}{\partial w_i^2} + (x - w_i^2)\frac{\partial F}{\partial w_i} \right) + \sum_{i<j} \frac{2}{w_i - w_j}\left(\frac{\partial F}{\partial w_i} - \frac{\partial F}{\partial w_j} \right) = 0 \tag{3.6}$$

and the boundary conditions

$$F \to 1 \qquad \text{as } x \to \infty \text{ with } w_1,\dots,w_r \text{ bounded below;} \tag{3.7}$$

$$F \to 0 \qquad \text{as any } w_i \to -\infty \text{ with } x \text{ bounded above.} \tag{3.8}$$

Furthermore, F_β is "continuous to the boundary" as one or several $w_i \to +\infty$. For subsequent eigenvalue laws $F_\beta^k(x; w_1,\dots,w_r)$, (3.8) is replaced with the recursive boundary condition

$$\begin{gathered} F^k(x; w_1,\dots,w_r) \to F^{k-1}(x^*; w_1^*,\dots,w_{r-1}^*, +\infty) \\ \text{as } x \to x^* \in \mathbb{R},\ w_i \to w_i^* \in \mathbb{R} \text{ for } i = 1,\dots,r-1, \text{ and } w_r \to -\infty. \end{gathered} \tag{3.9}$$

At $\beta = 2$ these distributions were obtained in BBP in the form of Fredholm determinants of finite-rank perturbations of the Airy kernel. Baik (2006) derived Painlevé II

formulas, and by a symbolic computation with the computer algebra system Maple we were able to verify that the latter satisfy the PDE (3.6) for $r = 2, 3, 4, 5$. We give the details in Section 5.3 of Chapter 5. Of course a pencil-and-paper proof for all r would be more illuminating and certainly more satisfying.

We make two final remarks. From the finite n matrix models it is clear that the "rank r deformed" limiting distributions $F_{\beta,r}(x; w_1, \ldots, w_r)$ reduce to those for a lower rank $r_0 < r$ in the following way:

$$F_{\beta,r}(x; w_1, \ldots, w_{r_0}, +\infty, \ldots, +\infty) = F_{\beta,r_0}(x; w_1, \ldots, w_{r_0}).$$

Unfortunately this reduction relation is not readily apparent from any of our characterizations (operator, SDE or PDE).

Lastly, the SDE and PDE characterizations seem to make sense for all $\beta > 0$ (although one has to be careful for $\beta < 1$). It would be interesting to find natural "general β multispiked models" at finite n, interpolating between those studied here at $\beta = 1, 2, 4$ and generalizing those introduced in Chapter 2 for $r = 1$. At $\beta = 2$, perhaps one could discover a relationship with formulas of Baik and Wang (2011).

3.2 A canonical form for perturbation in a fixed subspace

In Chapter 2 we observed that the tridiagonal models of Gaussian and Wishart matrices were amenable to rank one perturbation. In this section we introduce a banded (also block tridiagonal) generalization amenable to higher-rank perturbation. We first describe it as a natural object of pure linear algebra; we then show how it interacts with the structure of Gaussian and Wishart random matrices to produce the band forms displayed in the introduction.

The basic facts of "linear algebra over $\mathbb{F}$", where $\mathbb{F}$ may be $\mathbb{R}$, $\mathbb{C}$ or the skew field of quaternions $\mathbb{H}$, are summarized in Appendix E of Anderson et al. (2009). Everything we need (inner product geometry, self-adjointness, eigenvalues, and the spectral theorem) simply works over $\mathbb{H}$ as expected, keeping in mind only that non-real scalars may not commute.

The band Jacobi form as an algebraic object

We present a natural "canonical form" for studying perturbations in a fixed subspace of dimension r. It is a $(2r+1)$-diagonal band matrix generalizing the symmetric tridiagonal Jacobi form, which is the $r = 1$ case. The outermost diagonals continue to be positive; however, intermediate diagonals between the main and outermost ones are not in general real. Once again, the presence of $\mathbb{F}$ Gaussians is the obstacle to writing down a general β analogue.

We begin with a geometric, coordinate-free formulation.

Theorem 3.2.1. *Let T be a self-adjoint linear transformation on a finite-dimensional inner product space V of dimension n over $\mathbb{F}$. An orthonormal sequence $\{v_1, \ldots, v_r\} \subset V$ with $1 \leq r \leq n$ can be extended to an ordered orthonormal basis $\{v_1, \ldots, v_n\}$ for V such that $\langle v_i, Tv_j \rangle \geq 0$ for $|i-j| = r$ and $\langle v_i, Tv_j \rangle = 0$ for $|i-j| > r$. Furthermore, if $\langle v_i, Tv_j \rangle > 0$ for $|i-j| = r$ then the extension is unique.*

The point is that the same extension works for $T' = T + P$ provided $P \in M_n(\mathbb{F})$ satisfies $P\big|_{\{v_1,\ldots,v_r\}^\perp} = 0$. In this case $\operatorname{span}\{v_1, \ldots, v_r\}$ is also an invariant subspace of P and we speak of **perturbing in this subspace**.

Proof. We give an explicit inductive construction. Along the way we will see that the uniqueness condition holds precisely when the choice is forced at each step.

It is convenient to restate the properties of the orthonormal basis in the theorem in the following equivalent way: For $r+1 \leq i \leq n$ we have $\langle v_i, Tv_{i-r} \rangle \geq 0$ and $Tv_{i-r} \in \operatorname{span}\{v_1, \ldots, v_i\}$. Suppose inductively that $v_1, \ldots, v_{k-1}$ have been obtained for some $r+1 \leq k \leq n$, satisfying the preceding conditions for $r+1 \leq i \leq k-1$. Let $w = Tv_{k-r}$; we must choose v_k so that $\langle v_k, w \rangle \geq 0$ and $w \in \operatorname{span}\{v_1, \ldots, v_k\}$. There are two cases to consider. If $w \notin \operatorname{span}\{v_1, \ldots, v_{k-1}\}$ then v_k must be a multiple of $w' = w - \sum_{i=1}^{k-1} \langle v_i, w \rangle v_i$; the positivity condition further forces $v_k = w'/|w'|$, which gives $\langle v_k, w \rangle = |w'| > 0$. If $w \in \operatorname{span}\{v_1, \ldots, v_{k-1}\}$ then any $v_k \in \{v_1, \ldots, v_{k-1}\}^\perp$ will do, and in this case $\langle v_k, w \rangle = 0$. □

Remark 3.2.2. When uniqueness holds, as is generically the case, the basis may also be obtained by applying the Gram-Schmidt process to the first n vectors of the sequence

$$v_1, \ldots, v_r, Tv_1, \ldots, Tv_r, T^2v_1, \ldots, T^2v_r, \ldots.$$

We now state and prove a concrete matrix formulation in which the first r coordinate vectors play the role of $v_1, \dots v_r$. The point of the second proof is that it emphasizes the resulting band matrix rather than the change of basis; the algorithm will be used in the next subsection.

Theorem 3.2.3. *Let $A \in M_n(\mathbb{F})$ and $1 \leq r \leq n$. There there exists $U \in U_n(\mathbb{F})$ of the form $U = I_r \oplus \tilde{U}$ with $\tilde{U} \in U_{n-r}(\mathbb{F})$ such that $B = UAU^\dagger$ satisfies*

$$B_{ij} \geq 0 \quad \text{for } 1 \leq i, j \leq n \text{ with } |i-j| = r, \tag{3.10}$$

$$B_{ij} = 0 \quad \text{for } 1 \leq i, j \leq n \text{ with } |i-j| > r. \tag{3.11}$$

Furthermore, if strict positivity holds in (3.10) *then U and B as such are unique.*

We refer to B as the **band Jacobi form** of A. The allowed perturbations here have the form $P = \tilde{P} \oplus 0_{n-r}$ for $\tilde{P} \in M_r(\mathbb{F})$; these are invariant under conjugation by U, so $U(A+P)U^\dagger = B + P$.

Proof. We prove existence by giving an explicit algorithm; it generalizes the Lanczos algorithm, which applies in the case $r = 1$.

- For the first step, let $v = [A_{i,1}]_{r+1 \leq i \leq n} \in \mathbb{F}^{n-r}$ and take $\tilde{U} \in U_{n-r}(\mathbb{F})$ such that $\tilde{U}v = |v|\tilde{e}_1$, where $\tilde{e}_1$ is the first standard basis vector of $\mathbb{F}^{n-r}$. A concrete choice is the Householder reflection $\tilde{U} = I_{n-r} - 2ww^\dagger/w^\dagger w$ with $w = v - |v|\tilde{e}_1$. Set $U_1 = I_r \oplus \tilde{U}$ and $B_1 = U_1 A U_1^\dagger$.

- Continue inductively: Having obtained U_k, B_k, let $v = [(B_1)_{i,(k+1)}]_{r+k+1 \leq i \leq n} \in \mathbb{F}^{n-r-k}$ and take $\tilde{U} \in U_{n-r-k}(\mathbb{F})$ such that $\tilde{U}v = |v|\tilde{e}_1$. Set $U_{k+1} = I_{r+k} \oplus \tilde{U}$ and $B_{k+1} = U_{k+1} B_k U_{k+1}^\dagger$.

- Stop when $k = n - r$. Let $U = U_{n-r} \cdots U_1$ and $B = B_{n-r} = UAU^\dagger$.

It is immediate that U and B have the required properties. The point is that the kth column of B_k already "looks right", i.e. $(B_k)_{r+k,k} \geq 0$ and $(B_k)_{r+l,k} = 0$ for $l > k$, and subsequent transformations $U_{k+1}, \dots, U_{n-k} \in \{I_{r+k}\} \oplus U_{n-r-k}(\mathbb{F})$ "don't mess it up."

Towards uniqueness, suppose that $U', B' = U'AU'^\dagger$ also have the required properties and let $W = U'U^{-1}$ so that $B' = WBW^\dagger$. Assume inductively that $W \in \{I_{r+k}\} \oplus U_{n-r-k}(\mathbb{F})$, which is certainly true in the base case $k = 0$. Write $W = I_{r+k} \oplus \tilde{W}$. Let $b = [B_{i,k+1}]_{r+k+1 \leq i \leq n} \in \mathbb{F}^{n-r-k}$ and similarly for b'. Then $b' = \tilde{W}b$. But $b = a\tilde{e}_1$ and

$b' = a'\tilde{e}_1$ with $a, a' > 0$ by assumption. It follows that $a = a'$ and $\tilde{W}\tilde{e}_1 = \tilde{e}_1$. Hence $\tilde{W} \in \{I_1\} \oplus U_{n-r-(k+1)}(\mathbb{F})$ and $W \in \{I_{r+k+1}\} \oplus U_{n-r-(k+1)}(\mathbb{F})$, completing the induction step. We conclude that $W = I_n$. □

Perturbed Gaussian and spiked Wishart models

The change of basis described above interacts very nicely with the Gaussian structure in Gaussian and Wishart random matrices. The $r = 1$ case of this observation is due to Trotter (1984), who described the tridiagonal forms explicitly. His forms fall into the framework of Theorem 3.2.1 by taking the initial vector to be fixed in the Gaussian case, and taking it to be the top row of the data matrix in the Wishart case. As we observed in Chapter 2, the change of basis commutes with rank one additive perturbations for the Gaussian case and with rank one spiking for the Wishart case. We now extend the story to the $r > 1$ setting.

In the Gaussian case we will be perturbing in a fixed (non-random) subspace; without loss of generality this may be taken as the initial r-dimensional coordinate subspace, and so we take the basis of Theorem 3.2.1 that begins with the first r standard basis vectors. We can therefore obtain the band form by a direct application of the algorithm from the proof of Theorem 3.2.3. The Wishart case is a little more complicated; here we want to perturb in the random subspace spanned by the first r rows of the data matrix. Our new basis will begin with the Gram-Schmidt orthogonalization of these initial rows. As in the $r = 1$ case it is most transparent to construct a lower band form of the data matrix first, afterwards realizing the band Jacobi form as its multiplicative symmetrization. In both the Gaussian and the Wishart cases we will see that the uniqueness condition of Theorem 3.2.1 holds almost surely.

Let A be an $n \times n$ GOE matrix. Applying the algorithm from the proof of Theorem 3.2.3 while keeping track of the distribution of the matrix B_k at each step—the key of course being the unitary invariance of standard Gaussian vectors—yields the following

band Jacobi random matrix $G = UAU^\dagger$:

$$G_{ij} = \begin{cases} \sqrt{\frac{2}{\beta}}\,\widetilde{g}_i & i = j \\ g_{ij} & j < i < j + r \\ \frac{1}{\sqrt{\beta}}\chi_{(n-i+1)\beta} & i = j + r \\ 0 & i > j + r \\ G^*_{ji} & i < j \end{cases} \tag{3.12}$$

for $1 \le i, j \le n$, where the random variables appearing explicitly are independent, $\widetilde{g}_i \sim N(0,1)$, $g_{ij} \sim \mathbb{F}N(0,1)$, and $\chi_k \sim \mathrm{Chi}(k)$. The latter is the distribution of the length of a k-dimensional standard Gaussian vector.

We can introduce a rank r additive perturbation $A = A_0 + \sqrt{n}P$, where $P = \tilde{P} \oplus 0_{n-r}$ with $\tilde{P} \in M_r(\mathbb{F})$; since P commutes with the change of basis $U \in \{I_r\} \oplus U_{n-r}(\mathbb{F})$, we can write

$$G = UAU^\dagger = U(A_0 + \sqrt{n}P)U^\dagger = UA_0U^\dagger + \sqrt{n}P = G_0 + \sqrt{n}P. \tag{3.13}$$

As expected the perturbation shows up undisturbed in the upper-left $r \times r$ corner of G

Turning to the Wishart case, we first consider the null Wishart random matrix $X^\dagger X$, where X is $p \times n$ with independent $\mathbb{F}N(0,1)$ entries. (Remember that $X^\dagger X$ and $XX^\dagger$ have the same nonzero eigenvalues $\lambda_1, \ldots, \lambda_{n \wedge p}$.) The final form can be described abstractly as given in the basis of Theorem 3.2.1 that extends the Gram-Schmidt orthogonalization of the first r rows of X. One cannot readily obtain a description of the resulting random matrix from here, however, so we give another way that generalizes Trotter's original procedure. It is a "singular value analogue" of the algorithm from the proof of Theorem 3.2.3, producing matrices $U \in U_n(\mathbb{F})$ and $V \in U_p(\mathbb{F})$ such that $L = VXU$ has a "lower band form" that is zero off the main and first r subdiagonals and positive on the outermost of these. The key is to work alternately on rows and columns.

- Take $U_1 \in U_n(\mathbb{F})$ so that the first row of XU_1 lies in the (positive) direction of the first coordinate basis vector of $\mathbb{F}^n$.

- Take $V_1 = I_r \oplus U_{p-r}(\mathbb{F})$ so that $[(V_1XU_1)_{i,1}]_{r+1 \le i \le p} \in \mathbb{F}^{p-r}$ lies in the direction of the first coordinate basis vector of the latter subspace.

- Take $U_2 \in I_1 \oplus U_{n-1}(\mathbb{F})$ so that $[(V_1XU_1U_2)_{2,j}]_{2 \le j \le n} \in \mathbb{F}^{n-1}$ lies in the direction of the first coordinate basis vector of the latter subspace.

- Take $V_2 \in I_{r+1} \oplus U_{p-r-1}(\mathbb{F})$ so that $[(V_2V_1XU_1U_2)_{i,2}]_{r+2\le j\le p} \in \mathbb{F}^{p-r-1}$ lies in the direction of the first coordinate basis vector of the latter subspace.

- Continue in this way until the rows and columns both run out (stop alternating if one runs out before the other).

The resulting $L = V_{n\wedge(p-r)}\cdots V_1XU_1\cdots U_{n\wedge p}$ has $n\wedge p$ nonzero columns and $(n+r)\wedge p$ nonzero rows, which can be described as follows:

$$L_{ij} = \begin{cases} \frac{1}{\sqrt{\beta}}\widetilde{\chi}_{(n-i+1)\beta} & i = j \\ g_{ij} & j < i < j+r \\ \frac{1}{\sqrt{\beta}}\chi_{(p-i+1)\beta} & i = j+r \\ 0 & i < j \text{ or } i > j+r \end{cases} \tag{3.14}$$

where the entries are independent, $\widetilde{\chi}_k, \chi_k \sim \mathrm{Chi}(k)$, $g_{ij} \sim \mathbb{F}N(0,1)$. Truncating the remaining zero rows or columns, the matrix $S = L^\dagger L$ is $(n\wedge p)\times(n\wedge p)$ and has the same *nonzero* eigenvalues as $X^\dagger X$. It has the band form

$$S_{ij} = \begin{cases} \frac{1}{\beta}\widetilde{\chi}^2_{(n-i+1)\beta} + \sum_{i<k<i+r}|g_{k,i}|^2 + \frac{1}{\beta}\chi^2_{(p-i-r+1)\beta} & i = j \\ \frac{1}{\sqrt{\beta}}\widetilde{\chi}_{(n-i+1)\beta}g_{ij} + \sum_{i<k<j+r} g^*_{k,i}g_{k,j} + \frac{1}{\sqrt{\beta}}g^*_{j+r,i}\chi_{(p-j-r+1)\beta} & j < i < j+r \\ \frac{1}{\beta}\widetilde{\chi}_{(n-i+1)\beta}\chi_{(p-i+1)\beta} & i = j+r \\ 0 & i > j+r \\ S^*_{ji} & i < j \end{cases} \tag{3.15}$$

where we have ignored the issue of truncation in the final r rows and columns ($g's$ and $\chi's$ with indices beyond the allowed range should simply be zero). The change of basis is thus $U_1\cdots U_{n\wedge p}$; a little thought shows that, as claimed earlier, the new basis begins with the orthogonalization of the first r rows of X. Since the form (3.15) satisfies the uniqueness condition of Theorem 3.2.1 a.s., the basis is indeed the one given by the theorem.

Now we consider the spiked Wishart matrix $X^\dagger X = X_0^\dagger \Sigma X_0$, with $\Sigma = \tilde{\Sigma}\oplus I_{p-r} > 0$. Here X_0 is a null Wishart matrix and $X = \Sigma^{1/2}X_0$. Notice that $X^\dagger X - X_0^\dagger X_0 = X_0^\dagger\big((\tilde{\Sigma} - I_r)\oplus 0\big)X_0$ is indeed an additive perturbation in the subspace spanned by the first r rows of X_0. Since $\Sigma^{1/2} = \tilde{\Sigma}^{1/2}\oplus I$ commutes with the inner transformation $V \in \{I_r\}\oplus U_{p-r}(\mathbb{F})$, we have

$$L^\dagger L = U^\dagger X^\dagger XU = U^\dagger X_0^\dagger \Sigma X_0 U = U^\dagger X_0^\dagger V^\dagger \Sigma V X_0 U = L_0^\dagger \Sigma L_0$$

where $L = VXU$ and $L_0 = VX_0U$. The point is that same change of basis works in the rank r spiked case, and by the lower band structure of L_0, the perturbation shows up in the upper-left $r \times r$ corner:

$$S - S_0 = L^\dagger L - L_0^\dagger L_0 = \tilde{L}_0^\dagger(\tilde{\Sigma} - I_r)\tilde{L}_0 \oplus 0. \tag{3.16}$$

Viewed in terms of the algorithm used to produce L, the point is that the first r rows of X are never "mixed" together or with the lower rows, but only "rotated" within themselves.

3.3 Limits of block tridiagonal matrices

The banded forms of Section 3.2 may also be considered as block tridiagonal matrices with $r \times r$ blocks. In this section we give general conditions under which such random matrices, appropriately scaled, converge at the soft spectral edge to a random Schrödinger operator on the half-line with $r \times r$ matrix-valued potential and general self-adjoint boundary condition at the origin. In Section 3.4 we verify these assumptions for the two specific matrix models we consider.

Proposition 3.3.7 establishes that the limiting operator is a.s. bounded below with purely discrete spectrum via a variational principle. The main result is Theorem 3.3.9, which asserts that the low-lying states of the discrete models converge to those of the operator limit.

The scalar $r = 1$ case of Chapter 2, based in turn on RRV, serves as a prototype. Care is required throughout to adapt the arguments to the matrix-valued setting, and we give a self-contained treatment.

Discrete model and embedding

Underlying the convergence is the embedding of the discrete half-line $\mathbb{Z}_+ = \{0, 1, \ldots\}$ into the continuum $\mathbb{R}_+ = [0, \infty)$ via $j \mapsto j/m_n$, where the scale factors $m_n \to \infty$ but with $m_n = o(n)$. Define an associated embedding of vector-valued function spaces by step functions:

$$\ell_n^2(\mathbb{Z}_+, \mathbb{F}^r) \hookrightarrow L^2(\mathbb{R}_+, \mathbb{F}^r), \quad (v_0, v_1, \ldots) \mapsto v(x) = v_{\lfloor m_n x \rfloor},$$

which is isometric with ℓ_n^2 norm $\|v\|^2 = m_n^{-1} \sum_{j=0}^\infty |v_j|^2$. (Recall that $\mathbb{F}^r$ and L^2 have norms $|v|^2 = v^\dagger v$ and $\|f\|^2 = \int_0^\infty |f|^2$ respectively.) Fix a standard basis for ℓ_n^2 with

lexicographic ordering

$$(e_1, 0, \ldots), (e_2, 0, \ldots), \ldots, (e_r, 0, \ldots), (0, e_1, 0, \ldots), \ldots$$

where $e_1, \ldots, e_r$ is the standard basis for $\mathbb{F}^r$. Identify $\mathbb{F}^n$ with the n-dimensional initial coordinate subspace of ℓ^2_n, consisting of $\mathbb{F}^r$-valued step-functions supported on the interval $\big[0, \lceil n/r \rceil / m_n\big)$ and with the final step value in the subspace spanned by $e_1, \ldots, e_{r-(\lceil n/r \rceil r - n)}$. Our $n \times n$ matrices will act on $\mathbb{F}^n$ with respect to the above basis; we will generally assume the embedding $\mathbb{F}^n \subset \ell^2_n \hookrightarrow L^2$ implicitly.

We define some operators on L^2, all of which leave ℓ^2_n invariant and may also be considered as infinite block matrices with $r \times r$ blocks. The translation operator $(T_n f)(x) = f(x + m_n^{-1})$ extends the left shift on ℓ^2_n. Its adjoint $T_n^\dagger$ is the right shift, where $T_n^\dagger f = 0$ on $[0, m_n^{-1})$. The difference quotient $D_n = m_n(T_n - 1)$ extends a discrete derivative. Write $\operatorname{diag}(A_0, A_1, \ldots)$ for both an $r \times r$ block diagonal matrix and its extension to a pointwise matrix multiplication on L^2. Thus $E_n = \operatorname{diag}(m_n I_r, 0, 0, \ldots)$ is scalar multiplication by $m_n \mathbf{1}_{[0, m_n^{-1})}$, a "discretized delta function at the origin". Orthogonal projection from ℓ^2_n onto $\mathbb{F}^n$ extends to a multiplication $R_n = \operatorname{diag}(I_r, \ldots, I_r, \operatorname{diag}(1, \ldots, 1, 0, \ldots, 0), 0, \ldots)$, in which there are $\lceil n/r \rceil$ non-zero blocks and a total of n 1's.

Let $(Y_{n,i;j})_{j \in \mathbb{Z}_+}$, $i = 1, 2$ be two discrete-time $r \times r$ matrix-valued random processes with $Y_{n,1;j} \in M_r(\mathbb{F})$ for all j. The processes may be embedded into continuous time as above, by setting $Y_{n,i}(x) = Y_{n,i;\lfloor m_n x \rfloor}$. Note also that T_n and $\triangle_n = m_n(1 - T_n^\dagger) = -D_n^\dagger$ may be sensibly applied to such matrix-valued functions. The processes $Y_{n,i}$ are on- and off-diagonal integrated potentials, and we define a "potential operator" by

$$V_n = \operatorname{diag}(\triangle_n Y_{n,1}) + \tfrac{1}{2}\big(\operatorname{diag}(\triangle_n Y_{n,2}) T_n + T_n^\dagger \operatorname{diag}(\triangle_n Y_{n,2}^\dagger)\big). \tag{3.17}$$

Fix $W_n \in M_r(\mathbb{F})$, a nonrandom "boundary term".

Finally, consider

$$H_n = R_n\big(D_n^\dagger D_n + V_n + W_n E_n\big) R_n. \tag{3.18}$$

This operator leaves the initial coordinate subspace $\mathbb{F}^n$ invariant; we shall also use H_n to denote the *matrix of its restriction to* $\mathbb{F}^n$. The matrix $H_n \in M_n(\mathbb{F})$ is self-adjoint and block tridiagonal up to a truncation in the lower-right corner. Its main- and super-diagonal processes are

$$\begin{gathered} m_n^2 + m_n(W_n + Y_{n,1;0}),\ 2m_n^2 + m_n(Y_{n,1;1} - Y_{n,1;0}),\ 2m_n^2 + m_n(Y_{n,1;2} - Y_{n,1;1}),\ \ldots \\ -m_n^2 + \tfrac{1}{2} m_n Y_{n,2;0},\ -m_n^2 + \tfrac{1}{2} m_n (Y_{n,2;1} - Y_{n,2;0}),\ \ldots \end{gathered} \tag{3.19}$$

respectively; the sub-diagonal process is of course the conjugate transpose of the super-diagonal process. (We could have absorbed W_n into $Y_{n,1}$ as an additive constant, but keep it separate for reasons that will soon be clear. Note also that the upper-left block has m_n^2 rather than $2m_n^2$.) We refer to H_n as a **rank r block tridiagonal ensemble**.

As in RRV and Chapter 2, convergence rests on a few key assumptions on the potential and boundary terms just introduced. By choice, no additional scaling will be required. The role of the convergence in the first and third assumption below will be clear as soon as we define the continuum limit. The growth and oscillation bounds of the second assumption (and the lower bound implied by the third) ensure tightness of the low-lying states; in particular, they guarantee that the spectrum remains discrete and bounded below in the limit.

Assumption 1 *(Tightness and convergence).* There exists a continuous $M_r(\mathbb{F})$-valued random process $\{Y(x)\}_{x\geq 0}$ with $Y(0) = 0$ such that

$$\begin{aligned} \{Y_{n,i}(x)\}_{x\geq 0},\ i = 1,2 \quad &\text{are tight in law,} \\ Y_{n,1} + \tfrac{1}{2}(Y_{n,2} + Y_{n,2}^\dagger) \Rightarrow Y \quad &\text{in law} \end{aligned} \tag{3.20}$$

with respect to the compact-uniform topology (defined using any matrix norm).

Assumption 2 *(Growth and oscillation bounds).* There is a decomposition

$$Y_{n,i;j} = m_n^{-1} \sum_{k=0}^{j} \eta_{n,i;k} + \omega_{n,i;j} \tag{3.21}$$

(so $\triangle_n Y_{n,i} = \eta_{n,i} + \triangle_n \omega_{n,i}$) with $\eta_{n,i;j} \geq 0$ (as matrices), such that for some deterministic scalar continuous nondecreasing unbounded functions $\overline{\eta}(x) > 0$, $\zeta(x) \geq 1$ not depending on n, and random constants $\kappa_n \geq 1$ defined on the same probability spaces, the following hold: The κ_n are tight in distribution, and for each n we have almost surely

$$\overline{\eta}(x)/\kappa_n - \kappa_n \ \leq\ \eta_{n,1}(x) + \eta_{n,2}(x) \ \leq\ \kappa_n\big(1 + \overline{\eta}(x)\big), \tag{3.22}$$

$$\eta_{n,2}(x) \ \leq\ 2m_n^2, \tag{3.23}$$

$$\big|\omega_{n,1}(\xi) - \omega_{n,1}(x)\big|^2 + \big|\omega_{n,2}(\xi) - \omega_{n,2}(x)\big|^2 \ \leq\ \kappa_n\big(1 + \overline{\eta}(x)/\zeta(x)\big) \tag{3.24}$$

for all $x, \xi \in [0, \lceil n/r \rceil / m_n)$ with $|\xi - x| \leq 1$. Here matrix inequalities have their usual meaning and single bars denote the spectral (or $\ell^2(\mathbb{F}^r)$ operator) norm.

Assumption 3 *(Critical or subcritical perturbation).* For some orthonormal basis $u_1, \ldots, u_r$ of $\mathbb{F}^r$ and $-\infty < w_1 \leq \ldots \leq w_r \leq \infty$ we have $W_n = \sum_{i=1}^r w_{n,i} u_i u_i^\dagger$, where $w_{n,i} \in \mathbb{R}$ satisfy $\lim_{n\to\infty} w_{n,i} = w_i$ for each i.

We write $r_0 = \#\{i : w_i < \infty\} \in \{0, \ldots, r\}$ for the "critical rank". Formally, $W_n \to W = \sum_{i=1}^r w_i u_i u_i^\dagger \in M_r^*(\mathbb{F})$. It is natural to view W as a parameter: that is, we will consider the joint behaviour of the model (for given $Y_{n,i}$, Y) over all W_n, W satisfying Assumption 3.

Reduction to deterministic setting

In the next subsection we will define a limiting object in terms of $Y(x)$ and W; we want to prove that the discrete models converge to this continuum limit in law. We reduce the problem to a deterministic convergence statement as follows. First, select any subsequence. It will be convenient to extract a further subsequence so that certain additional tight sequences converge jointly in law; Skorokhod's representation theorem (see Ethier and Kurtz 1986) says this convergence can be realized almost surely on a single probability space. We may then proceed pathwise.

In detail, consider (3.20)–(3.24). Note in particular that non-negativity of the $\eta_{n,i}$ and the upper bound of (3.22) give that for $i = 1, 2$ the piecewise linear process $\left\{\int_0^x \eta_{n,i}\right\}_{x\geq 0}$ is tight in distribution, pointwise with respect to the spectral norm and in fact compact-uniformly. Given a subsequence, we pass to a further subsequence so that the following distributional limits exist jointly:

$$\begin{aligned} Y_{n,i} &\Rightarrow Y_i, \\ \textstyle\int_0 \eta_{n,i} &\Rightarrow \widetilde{\eta}_i, \\ \kappa_n &\Rightarrow \kappa, \end{aligned} \tag{3.25}$$

for $i = 1, 2$, where convergence in the first two lines is in the compact-uniform topology. We realize (3.25) pathwise a.s. on some probability space and continue in this deterministic setting.

We can take (3.22)–(3.24) to hold with κ_n replaced with a single κ. Observe that (3.22) gives a local Lipschitz bound on the $\int \eta_{n,i}$, which is inherited by their limits $\widetilde{\eta}_i$ (the spectral norm controls the matrix entries). Thus $\eta_i = \widetilde{\eta}_i'$ is defined almost everywhere on $\mathbb{R}_+$, satisfies (3.22), and may be defined to satisfy this inequality everywhere. Furthermore, one easily checks that $m_n^{-1} \sum \eta_{n,i} \to \int \eta_i$ compact-uniformly as well (use continuity

of the limit). Therefore $\omega_{n,i} = y_{n,i} - m_n^{-1}\sum \eta_{n,i}$ must have a continuous limit ω_i for $i = 1, 2$; moreover, the bound (3.24) is inherited by the limits. Lastly, put $\eta = \eta_1 + \eta_2$, $\omega = \omega_1 + \frac{1}{2}(\omega_2 + \omega_2^\dagger)$ and note that $Y_i = \int \eta_i + \omega_i$ and $Y = \int \eta + \omega$. For convenience, we record the bounds inherited by η, ω:

$$\overline{\eta}(x)/\kappa - \kappa \;\le\; \eta(x) \;\le\; \kappa\big(1 + \overline{\eta}(x)\big) \tag{3.26}$$

$$|\omega(\xi) - \omega(x)|^2 \;\le\; \kappa\big(1 + \overline{\eta}(x)/\zeta(x)\big) \tag{3.27}$$

for $x, \xi \in \mathbb{R}_+$ with $|\xi - x| \le 1$ (and note that $\kappa \ge 1$).

We will assume this **subsequential pathwise coupling** for the remainder of the section.

Limiting object and variational characterization

Formally, the limiting object is the eigenvalue problem

$$\begin{gathered} \mathcal{H} f = \Lambda f \quad \text{on } L^2(\mathbb{R}_+, \mathbb{F}^r) \\ f'(0) = W f(0) \end{gathered} \tag{3.28}$$

where

$$\mathcal{H} = -\frac{d^2}{dx^2} + Y'(x).$$

Writing the spectral decomposition $W = \sum_{i=1}^r w_i u_i u_i^\dagger$, recall (Assumption 3) that we actually allow $w_i \in \mathbb{R}$ for $1 \le i \le r_0$ and, symbolically, $w_i = +\infty$ for $r_0 + 1 \le i \le r$. Writing $f_i = u_i^\dagger f$, the boundary condition is then to be interpreted as

$$\begin{aligned} f_i'(0) &= w_i f_i(0) \qquad && \text{for } i \le r_0, \\ f_i(0) &= 0 && \text{for } i > r_0. \end{aligned} \tag{3.29}$$

We thus have a completely general homogeneous linear self-adjoint boundary condition. We refer to $\operatorname{span}\{u_i : i > r_0\}$ as the **Dirichlet subspace** and the corresponding f_i as **Dirichlet components**; they will require special treatment in what follows.

We will actually work with a symmetric bilinear form (properly, sesquilinear if $\mathbb{F} = \mathbb{C}$ or $\mathbb{H}$) associated with the eigenvalue problem (3.28). Define a space of test functions C_0^∞ consisting of smooth $\mathbb{F}^r$-valued functions φ on $\mathbb{R}_+$ with compact support; we additionally require the Dirichlet components to be supported away from the origin. Introduce a symmetric bilinear form on $C_0^\infty \times C_0^\infty$ by

$$\mathcal{H}(\varphi, \psi) \;=\; \langle \varphi', \psi' \rangle - \langle \varphi', Y\psi \rangle - \langle \varphi, Y\psi' \rangle + \varphi(0)^\dagger W \psi(0), \tag{3.30}$$

where the Dirichlet part of the last term is interpreted as zero. Formally, the form $\mathcal{H}(\cdot,\cdot)$ is just the usual one $\langle\cdot,\mathcal{H}\cdot\rangle$ associated with the operator $\mathcal{H}$; the potential term has been integrated by parts and the boundary condition "built in". See also Remark 3.3.5 below.

The regularity and decay conditions naturally associated with this form are given by the following weighted Sobolev norm:

$$\|f\|_*^2 = \int_0^\infty \left(|f'|^2 + (1+\overline{\eta})\,|f|^2\right) + f(0)^\dagger W^+ f(0) \tag{3.31}$$

where the **positive part of** W is defined as $W^+ = \sum_{i=1}^r w_i^+ u_i u_i^\dagger$ with $w^+ = w \vee 0$. (Define the **negative part** similarly with $w_i^- = -(w \wedge 0)$, so that $W = W^+ - W^-$.) We refer to $\|\cdot\|_*$ as the L^* **norm** and define an associated Hilbert space L^* as the closure of C_0^∞ under this norm. (The formal Dirichlet terms are again interpreted to be zero, but they can also be thought of as imposing the Dirichlet condition.) We record some basic facts about L^*.

Fact 3.3.1. *Any $f \in L^*$ is uniformly Hölder(1/2)-continuous and satisfies $|f(x)|^2 \le 2\,\|f'\|\,\|f\| \le \|f\|_*^2$ for all x; furthermore, $f_i(0) = 0$ for $i > r_0$.*

Proof. We have $|f(y) - f(x)| = \left|\int_x^y f'\right| \le \|f'\|\,|y-x|^{1/2}$. For $f \in C_0^\infty$ we have $|f(x)|^2 = -\int_x^\infty 2\,\mathrm{Re}\, f^\dagger f' \le 2\,\|f\|\,\|f'\| \le \|f\|_*^2$; an L^*-bounded sequence in C_0^∞ therefore has a compact-uniformly convergent subsequence, so we can extend this bound to $f \in L^*$ and also conclude the behaviour in the Dirichlet components. □

Fact 3.3.2. *Every L^*-bounded sequence has a subsequence converging in the following modes: (i) weakly in L^*, (ii) derivatives weakly in L^2, (iii) uniformly on compacts, and (iv) in L^2.*

Proof. (i) and (ii) are just Banach-Alaoglu; (iii) is the previous fact and Arzelà-Ascoli again; (iii) implies L^2 convergence locally, while the uniform bound on $\int \overline{\eta}\,|f_n|^2$ produces the uniform integrability required for (iv). Note that the weak limit in (ii) really is the derivative of the limit function, as one can see by integrating against functions $\mathbf{1}_{[0,x]}$ and using pointwise convergence. □

By the bound in Fact 3.3.1 with $x = 0$, the boundary term in (3.31) could be done away with. It is natural to include the term, however, when considering all W simultaneously and viewing the Dirichlet case as a limiting case. More importantly, it clarifies the role of the boundary terms in the following key bound.

Lemma 3.3.3. *For every $0 < c < 1/\kappa$ there is a $C > 0$ such that, for each $b > 0$, the following holds for all $W \geq -b$ and all $f \in C_0^\infty$:*

$$c\,\|f\|_*^2 - (1+b^2)C\,\|f\|^2 \;\leq\; \mathcal{H}(f,f) \;\leq\; C\,\|f\|_*^2\,. \tag{3.32}$$

In particular, $\mathcal{H}(\cdot,\cdot)$ extends uniquely to a continuous symmetric bilinear form on $L^* \times L^*$.

Proof. For the first three terms of (3.30), we use the decomposition $Y = \int \eta + \omega$ from the previous subsection. Integrating the $\int \eta$ term by parts, (3.26) easily yields

$$\tfrac{1}{\kappa}\,\|f\|_*^2 - \kappa\,\|f\|^2 \;\leq\; \|f'\|^2 + \langle f, \eta f\rangle \;\leq\; \kappa\,\|f\|_*^2\,.$$

Break up the ω term as follows. The moving average $\overline{\omega}_x = \int_x^{x+1} \omega$ is differentiable with $\overline{\omega}'_x = \omega_{x+1} - \omega_x$; writing $\omega = \overline{\omega} + (\omega - \overline{\omega})$, we have

$$-2\,\mathrm{Re}\big\langle f', \omega f\big\rangle = \big\langle f, \overline{\omega}' f\big\rangle + 2\,\mathrm{Re}\big\langle f', (\overline{\omega} - \omega) f\big\rangle.$$

By (3.27), $\max\big(|\omega_\xi - \omega_x|\,, |\omega_\xi - \omega_x|^2\big) \leq C_\varepsilon + \varepsilon\overline{\eta}(x)$ for $|\xi - x| \leq 1$, where ε can be made small. In particular, the first term above is bounded absolutely by $\varepsilon\,\|f\|_*^2 + C_\varepsilon\,\|f\|^2$. Averaging, we also get $|\overline{\omega}_x - \omega_x| \leq (C_\varepsilon + \varepsilon\overline{\eta}(x))^{1/2}$; Cauchy-Schwarz then bounds the second term absolutely by $\sqrt{\varepsilon}\int_0^\infty |f'|^2 + \frac{1}{\sqrt{\varepsilon}}\int_0^\infty (C_\varepsilon + \varepsilon\overline{\eta})\,|f|^2$ and thus by $\sqrt{\varepsilon}\,\|f\|_*^2 + C'_\varepsilon\,\|f\|^2$. Now combine all the terms and set ε small to obtain a version of (3.32) with the boundary terms omitted (from both the form and the norm).

We break the boundary term in (3.30) into its positive and negative parts. For the negative part, Fact 3.3.2 gives $|f(0)|^2 \leq (\varepsilon/b)\,\|f'\|^2 + (b/\varepsilon)\,\|f\|^2$; $W^- \leq b$ then implies that

$$0 \;\leq\; f(0)^\dagger W^- f(0) \;\leq\; \varepsilon\,\|f\|_*^2 + C''_\varepsilon b^2\,\|f\|^2\,,$$

which may be subtracted from the inequality already obtained. For the positive part $f(0)^\dagger W^+ f(0)$, use the fact that $c \leq 1 \leq C$ to simply add it in. We thus arrive at (3.32).

For the L^* bilinear form bound, begin with the quadratic form bound $|\mathcal{H}(f,f)| \leq C_{c,b}\,\|f\|_*^2$; it is a standard Hilbert space fact that it may be polarized to a bilinear form bound (see e.g. §18 of Halmos 1957). □

Definition 3.3.4. We say $f \in L^*$ is an **eigenfunction** with **eigenvalue** Λ if $f \neq 0$ and for all $\varphi \in C_0^\infty$ we have

$$\mathcal{H}(\varphi, f) = \Lambda\,\langle \varphi, f\rangle\,. \tag{3.33}$$

Note that (3.33) then automatically holds for all $\varphi \in L^*$, by L^*-continuity of both sides.

Remark 3.3.5. This definition represents a weak or distributional version of the problem (3.28). As further justification, integrate by parts to write the definition

$$\langle \varphi', f' \rangle - \langle \varphi', Yf \rangle - \langle \varphi, Yf' \rangle + \varphi(0)^\dagger W f(0) = \Lambda \langle \varphi, f \rangle$$

in the form

$$\langle \varphi', f' \rangle - \langle \varphi', Yf \rangle + \left\langle \varphi', \textstyle\int_0 Yf' \right\rangle - \langle \varphi', Wf(0) \rangle = -\Lambda \left\langle \varphi', \textstyle\int_0 f \right\rangle,$$

which is equivalent to

$$f'(x) = Wf(0) + Y(x)f(x) - \int_0^x Yf' - \Lambda \int_0^x f \qquad \text{a.e. } x. \tag{3.34}$$

(For a Dirichlet component f_i the restriction on test functions implies that $\langle \varphi_i', 1 \rangle = 0$, so the first boundary term on the right-hand side is replaced with an arbitrary constant.) Now (3.34) shows that f' has a continuous version, and the equation may be taken to hold everywhere. In particular, f satisfies the boundary condition of (3.28) classically. (For a Dirichlet component we just find that the arbitrary constant is $f_i'(0)$.) One can also view (3.34) as a straightforward integrated version of the eigenvalue equation in which the potential term has been interpreted via integration by parts. This equation will be useful in Lemma 3.3.6 below and is the starting point for the development in Section 3.5.

We now characterize the eigenvalues and eigenfunctions variationally. As usual, it follows from the symmetry of the form that eigenvalues are real (and eigenfunctions with distinct eigenvalues are L^2-orthogonal). The L^2 part of the lower bound in (3.32) says the spectrum is bounded below. The rest of (3.32) implies that there are only finitely many eigenvalues below any given level: a sequence of normalized eigenfunctions with bounded eigenvalues must have an L^2-convergent subsequence by Fact 3.3.2. At a given level, more is true:

Lemma 3.3.6. *For each $\Lambda \in \mathbb{R}$, the corresponding eigenspace is at most r-dimensional.*

Proof. By linearity, it suffices to show a solution of (3.34) with $f'(0) = f(0) = 0$ must vanish identically. Integrate by parts to write

$$f'(x) = Y(x) \int_0^x f' - \int_0^x Yf' - \Lambda x \int_0^x f' + \Lambda \int_0^x t f'(t) dt,$$

which implies that $|f'(x)| \le C(x) \int_0^x |f'|$ with some $C(x) < \infty$ increasing in x. Gronwall's lemma then gives $|f'(x)| = 0$ for all $x \ge 0$. □

Proposition 3.3.7. *There is a well-defined $(k+1)$st lowest eigenvalue Λ_k, counting with multiplicity. The eigenvalues $\Lambda_0 \leq \Lambda_1 \leq \dots$ together with an orthonormal sequence of corresponding eigenvectors $f_0, f_1, \dots$ are given recursively by the variational problem*

$$\Lambda_k = \inf_{\substack{f \in L^*,\ \|f\|=1, \\ f \perp f_0, \dots f_{k-1}}} \mathcal{H}(f,f)$$

in which the minimum is attained and we set f_k to be any minimizer.

Remark 3.3.8. Since we must have $\Lambda_k \to \infty$, $\{\Lambda_0, \Lambda_1, \dots\}$ exhausts the spectrum and the resolvent operator is compact. We do not make this statement precise.

Proof. First taking $k = 0$, the infimum $\tilde{\Lambda}$ is finite by (3.32). Let f_n be a minimizing sequence; it is L^*-bounded, again by (3.32). Pass to a subsequence converging to $f \in L^*$ in all the modes of Fact 3.3.2. In particular $1 = \|f_n\| \to \|f\|$, so $\mathcal{H}(f,f) \geq \tilde{\Lambda}$ by definition. But also

$$\begin{aligned} \mathcal{H}(f,f) &= \|f'\|^2 + \langle f, \eta f \rangle + \big\langle f, \overline{\omega}' f \big\rangle + 2\,\mathrm{Re}\big\langle f', (\overline{\omega} - \omega) f \big\rangle + f(0)^\dagger W f(0) \\ &\leq \liminf_{n \to \infty} \mathcal{H}(f_n, f_n) \end{aligned}$$

by a term-by-term comparison. Indeed, the inequality holds for the first term by weak convergence, and for the second term by pointwise convergence and Fatou's lemma; the remaining terms are just equal to the corresponding limits, because the second members of the inner products converge in L^2 by the bounds from the proof of Lemma 3.3.3 together with L^*-boundedness and L^2-convergence. Therefore $\mathcal{H}(f,f) = \tilde{\Lambda}$.

A standard argument now shows $(\tilde{\Lambda}, f)$ is an eigenvalue-eigenfunction pair: taking $\varphi \in C_0^\infty$ and ε small, put $f^\varepsilon = (f + \varepsilon\varphi)/\|f + \varepsilon\varphi\|$; since f is a minimizer, $\frac{d}{d\varepsilon}\big|_{\varepsilon=0} \mathcal{H}(f^\varepsilon, f^\varepsilon)$ must vanish; the latter says precisely (3.33) with $\tilde{\Lambda}$. Finally, suppose (Λ, g) is any eigenvalue-eigenfunction pair; then $\mathcal{H}(g,g) = \Lambda$, and hence $\tilde{\Lambda} \leq \Lambda$. We are thus justified in setting $\Lambda_0 = \tilde{\Lambda}$ and $f_0 = f$.

Proceed inductively, minimizing now over the orthocomplement $\{f \in L^* : \|f\| = 1, f \perp f_0, \dots f_{k-1}\}$. Again, L^2-convergence of a minimizing sequence guarantees that the limit remains admissible; as before, the limit is in fact a minimizer; conclude by applying the arguments of the previous paragraph with φ, g also restricted to the orthocomplement.

□

Statement

We are finally ready to state the main result of this section. Recall that we consider eigenvectors of a matrix $H_n \in M_n(\mathbb{F})$ in the embedding $\mathbb{F}^n \subset \ell^2_n(\mathbb{Z}_+, \mathbb{F}^r) \hookrightarrow L^2(\mathbb{R}_+, \mathbb{F}^r)$ above.

Theorem 3.3.9. *Let H_n be a rank r block tridiagonal ensemble as in* (3.18) *satisfying Assumptions 1–3, and let $\lambda_{n,k}$ be its $(k+1)$st lowest eigenvalue. Define the associated form $\mathcal{H}$ as in* (3.30) *and let Λ_k be its a.s. defined $(k+1)$st lowest eigenvalue. In the deterministic setting of subsequential pathwise coupling, $\lambda_{n,k} \to \Lambda_k$ for each $k = 0, 1, \ldots$. Furthermore, a sequence of normalized eigenvectors corresponding to $\lambda_{n,k}$ is pre-compact in L^2 norm, and every subsequential limit is an eigenfunction corresponding to Λ_k. Finally, convergence holds uniformly over possible $W_n, W \geq -b > -\infty$. One recovers the corresponding distributional tightness and convergence statements for the full sequence, jointly for $k = 0, 1, \ldots$ in the sense of finite-dimensional distributions and jointly over W_n, W.*

Remark 3.3.10. The eigenvector convergence statement requires subsequences for two reasons: possible multiplicity of the limiting eigenvalues, and the sign or phase ambiguity of the eigenvectors. It is possible to formulate the conclusion of the theorem very simply using spectral projections. (If H has purely discrete spectrum, the spectral projection $\mathbf{1}_A(H)$ is simply orthogonal projection of L^2 onto the span of those eigenvectors of H whose eigenvalues lie in $A \subset \mathbb{R}$.) The joint eigenvalue-eigenvector convergence may be restated in the deterministic setting as follows: *For all $a \in \mathbb{R} \setminus \{\Lambda_0, \Lambda_1, \ldots\}$, the spectral projections $\mathbf{1}_{(-\infty,a)}(H_n) \to \mathbf{1}_{(-\infty,a)}(\mathcal{H})$ in L^2 operator norm.* The corresponding distributional statement holds jointly over all a that are a.s. off the limiting spectrum (or simply all a if the distributions of the Λ_k are non-atomic).

Remark 3.3.11. An operator-theoretic formulation of the theorem (which we do not develop here) would state a norm resolvent convergence: the resolvent matrices, precomposed with the finite-rank projections $L^2 \to \mathbb{F}^n$ associated with the embedding, converge to the continuum resolvent in L^2 operator norm. This mode of convergence is the strongest one can hope for in the unbounded setting (see e.g. VIII.7 of Reed and Simon 1980, Weidmann 1997).

The proof will be given over the course of the next two subsections.

Tightness

We will need a discrete analogue of the L^* norm and a counterpart of Lemma 3.3.3 with constants uniform in n. For $v \in \mathbb{F}^n \hookrightarrow L^2(\mathbb{R}_+, \mathbb{F}^r)$ as above, define the L_n^* **norm** by

$$\begin{aligned} \|v\|_{*n}^2 &= \left\langle v, \left(D_n^\dagger D_n + 1 + \overline{\eta} + E_n W_n^+\right) v \right\rangle \\ &= \int_0^\infty \left(|D_n v|^2 + (1+\overline{\eta})\,|v|^2\right) + v(0)^\dagger W_n^+ v(0) \end{aligned} \tag{3.35}$$

with the non-negative part W_n^+ defined as before.

Remark 3.3.12. When considering just a single W_n, W, the boundary term in (3.35) is really only required when the limit includes Dirichlet terms; it is simpler, however, not to distinguish the two cases here. More importantly, including this term clarifies the role of the boundary term in the following key bound. Note that the original case considered in RRV has $W_n = m_n$ in our notation. (The H_n form and L_n^* norm there contained a term $m_n\,|v_0|^2$, though it is hidden in the fact that, in our notation, they use $\triangle_n$ in place of D_n.)

Lemma 3.3.13. *For every $0 < c < 1/4\kappa$ there is a $C > 0$ such that, for each $b > 0$, the following holds for all n, $W_n \geq -b$ and $v \in \mathbb{F}^n$:*

$$c\,\|v\|_{*n}^2 - (1+b^2)C\,\|v\|^2 \;\leq\; \langle v, H_n v\rangle \;\leq\; C\,\|v\|_{*n}^2\,. \tag{3.36}$$

Proof. We drop the subscript n. The form associated with (3.18) is

$$\langle v, Hv\rangle = \|Dv\|^2 + \langle v, Vv\rangle + v(0)^\dagger W v(0). \tag{3.37}$$

The potential term $\langle v, Vv\rangle = \int_0^\infty v^\dagger V v$, defined in (3.17), is analyzed according to (3.21):

$$\begin{aligned} v^\dagger V v &= v^\dagger(\triangle Y_1)v + \operatorname{Re} v^\dagger(\triangle Y_2)Tv \\ &= \left(v^\dagger \eta_1 v + \operatorname{Re} v^\dagger \eta_2 T v\right) + \left(v^\dagger(\triangle\omega_1)v + \operatorname{Re} v^\dagger(\triangle\omega_2)Tv\right). \end{aligned}$$

Together with $|D_n v|^2$, the η-terms provide the structure of the bound as we now show. Afterwards we will control the ω-terms and lastly deal with the boundary term.

Recall (3.22) and that $\eta_i \geq 0$. For an upper bound, rearrange $(v - Tv)^\dagger \eta_2 (v - Tv) \geq 0$ to

$$\begin{aligned} \operatorname{Re} v^\dagger \eta_2 T v &\leq \tfrac{1}{2} v^\dagger \eta_2 v + \tfrac{1}{2}(Tv)^\dagger \eta_2 T v \\ &\leq \tfrac{1}{2}\kappa(\overline{\eta}+1)\left(|v|^2 + |Tv|^2\right). \end{aligned}$$

Now $\int \overline{\eta}\,|Tv|^2 = \int (T^\dagger\overline{\eta})\,|v|^2 \le \int \overline{\eta}\,|v|^2$ since $\overline{\eta}$ is nondecreasing, and we obtain

$$\|Dv\|^2 + \langle v, \eta_1 v\rangle + \operatorname{Re}\langle v, \eta_2 Tv\rangle \le 2\kappa\,\|v\|_*^2\,. \tag{3.38}$$

Toward a lower bound we use the slightly tricky rearrangement $0 \le (\frac{1}{2}v + Tv)^\dagger\eta_2(\frac{1}{2}v + Tv) = 3\operatorname{Re} v^\dagger\eta_2 Tv + (Tv-v)^\dagger\eta_2(Tv-v) - \frac{3}{4}v^\dagger\eta_2 v$. With (3.23) we get

$$\begin{aligned}\operatorname{Re} v^\dagger\eta_2 Tv &\ge -\tfrac{1}{3}(Tv-v)^\dagger\eta_2(Tv-v) + \tfrac{1}{4}v^\dagger\eta_2 v\\ &\ge -\tfrac{2}{3}\,|Dv|^2 + \tfrac{1}{4}v^\dagger\eta_2 v,\end{aligned}$$

so by (3.22),

$$|Dv|^2 + v^\dagger\eta_1 v + \operatorname{Re} v^\dagger\eta_2 Tv \ge \tfrac{1}{3}\,|Dv|^2 + \tfrac{1}{4}(\overline{\eta}/\kappa - \kappa)\,|v|^2$$

and thus

$$\|Dv\|^2 + \langle v, \eta_1 v\rangle + \operatorname{Re}\langle v, \eta_2 Tv\rangle \ge (1/4\kappa)\,\|v\|_*^2 - (\kappa/4)\,\|v\|^2\,. \tag{3.39}$$

We handle the ω-terms with a discrete analogue of the decomposition used in the continuum proof. Consider the moving average

$$\overline{\omega}_i = \lfloor m\rfloor^{-1}\sum_{j=1}^{\lfloor m\rfloor} T^j\omega_i$$

which has $\triangle\overline{\omega}_i = (m/\lfloor m\rfloor)(T^{\lfloor m\rfloor}-1)\omega_i$; it is convenient to extend $\omega_i(x) = \omega_i(\lceil n/r\rceil/m_n)$ for $x > \lceil n/r\rceil/m_n$. Decompose $\omega_i = \overline{\omega}_i + (\omega_i - \overline{\omega}_i)$. For the ω_1-term,

$$v^\dagger\triangle\omega_1 v = (m/\lfloor m\rfloor)v^\dagger(T^{\lfloor m\rfloor}\omega_1 - \omega_1)v + v^\dagger\triangle(\omega_1 - \overline{\omega}_1)v.$$

By (3.24) and Cauchy-Schwarz, the first term is bounded absolutely by $(C_\varepsilon + \varepsilon\overline{\eta})\,|v|^2$ and its integral by $\varepsilon\,\|v\|_*^2 + C_\varepsilon\,\|v\|^2$. The second term calls for a summation by parts:

$$\begin{aligned}\big\langle v, \triangle(\omega_1-\overline{\omega}_1)v\big\rangle &= m_n\left(\big\langle v, (\omega_1-\overline{\omega}_1)v\big\rangle - \big\langle Tv, (\omega_1-\overline{\omega}_1)Tv\big\rangle\right)\\ &= m_n\operatorname{Re}\big\langle v-Tv, (\omega_1-\overline{\omega}_1)(v+Tv)\big\rangle = \operatorname{Re}\big\langle Dv, (\overline{\omega}_1-\omega_1)(v+Tv)\big\rangle.\end{aligned}$$

The averaged bound $|\overline{\omega}_1 - \omega_1| \le (C_\varepsilon + \varepsilon\overline{\eta})^{1/2}$ and Cauchy-Schwarz bound the integrand

$$\big|(Dv)^\dagger(\overline{\omega}_1-\omega_1)(v+Tv)\big| \le \sqrt{\varepsilon}\,|Dv|^2 + (1/4\sqrt{\varepsilon})(C_\varepsilon+\varepsilon\overline{\eta})(|v|^2 + |Tv|^2),$$

and its integral by $\sqrt{\varepsilon}\,\|v\|_*^2 + C'_\varepsilon\,\|v\|^2$. One thus obtains a similar bound on $\big|\big\langle v, (\triangle\omega_1)v\big\rangle\big|$.

There are corresponding bounds for the ω_2-terms. For the $\overline{\omega}_2$-term, use $2|v||Tv| \le |v|^2 + |Tv|^2$. For the $(\omega_2 - \overline{\omega}_2)$-term, modify the summation by parts:

$$\begin{aligned}\mathrm{Re}\big\langle (v, \triangle(\omega_2 - \overline{\omega}_2)Tv\big\rangle &= m_n \,\mathrm{Re}\left(\big\langle (v - Tv), (\omega_2 - \overline{\omega}_2)Tv\big\rangle + \big\langle Tv, (\omega_2 - \overline{\omega}_2)(Tv - T^2v)\big\rangle\right) \\ &= \mathrm{Re}\big\langle Dv + TDv, (\overline{\omega}_2 - \omega_2)Tv\big\rangle.\end{aligned}$$

Incorporating all the ω-terms into (3.38, 3.39) and setting ε small, we obtain (3.36) but with the boundary terms omitted (from both the form and the norm).

We break the boundary term in (3.37) into its positive and negative parts. A discrete analogue of a bound from Fact 3.3.1 will be useful:

$$|v(0)|^2 = \int_0^\infty -D|v|^2 = \int_0^\infty \mathrm{Re}\, m(v - Tv)^\dagger (v + Tv) \le 2\|Dv\|\,\|v\|\,.$$

It gives $|v(0)|^2 \le (\varepsilon/b)\|Dv\|^2 + (b/\varepsilon)\|v\|^2$, and then $W^- \le b$ implies that

$$0 \;\le\; v(0)^\dagger W^- v(0) \;\le\; \varepsilon\|v\|_*^2 + C''_\varepsilon b^2 \|v\|^2$$

which may be subtracted from the inequality already obtained. The positive part may simply be added in using that $c \le 1 \le C$. We thus arrive at (3.36). □

Remark 3.3.14. If the W_n are not bounded below then the lower bound in (3.36) breaks down: in fact, the bottom eigenvalue of H_n really goes to $-\infty$ like minus the square of the bottom eigenvalue of W_n. This is the supercritical regime.

Convergence

We begin with a simple lemma, a discrete-to-continuous version of Fact 3.3.2.

Lemma 3.3.15. *Let $f_n \in \mathbb{F}^n$ with $\|f_n\|_{*n}$ uniformly bounded. Then there exist $f \in L^*$ and a subsequence along which (i) $f_n \to f$ uniformly on compacts, (ii) $f_n \to_{L^2} f$, and (iii) $D_n f_n \to f'$ weakly in L^2.*

Proof. Consider $g_n(x) = f_n(0) + \int_0^x D_n f_n$, a piecewise-linear version of f_n; they coincide at points $x = i/m_n$, $i \in \mathbb{Z}_+$. One easily checks that $\|g_n\|_*^2 \le 2\|f_n\|_{*n}^2$, so some subsequence $g_n \to f \in L^*$ in all the modes of Fact 3.3.2; for a Dirichlet component, the boundary term in the L_n^* norm guarantees that the limit vanishes at 0. But then also $f_n \to f$ compact-uniformly by a simple argument using the uniform continuity of f, $f_n \to_{L^2} f$ because $\|f_n - g_n\|^2 \le (1/3n^2)\|D_n f_n\|^2$, and $D_n f_n \to f'$ weakly in L^2 because $D_n f_n = g_n'$ a.e. □

Next we establish a kind of weak convergence of the forms $\langle \cdot, H_n \cdot \rangle$ to $\mathcal{H}(\cdot,\cdot)$. Let $\mathcal{P}_n$ be orthogonal projection from L^2 onto $\mathbb{F}^n$ embedded as above. The following facts will be useful and are easy to check. For $f \in L^2$, $\mathcal{P}_n f \to_{L^2} f$ (the Lebesgue differentiation theorem gives pointwise convergence and we have uniform L^2-integrability); further, if $f' \in L^2$ then $D_n f \to_{L^2} f'$ ($D_n f$ is a convolution of f' with an approximate delta); for smooth φ, $\mathcal{P}_n \varphi \to \varphi$ uniformly on compacts. It is also useful to note that $\mathcal{P}_n$ commutes with R_n and with $D_n R_n$. Finally, if $f_n \to_{L^2} f$, g_n is L^2-bounded and $g_n \to g$ weakly in L^2, then $\langle f_n, g_n \rangle \to \langle f, g \rangle$.

Lemma 3.3.16. *Let $f_n \to f$ be as in the hypothesis and conclusion of Lemma 3.3.15. Then for all $\varphi \in C_0^\infty$ we have $\langle \varphi, H_n f_n \rangle \to \mathcal{H}(\varphi, f)$. In particular, $\mathcal{P}_n \varphi \to \varphi$ in this way and so*

$$\langle \mathcal{P}_n \varphi, H_n \mathcal{P}_n \varphi \rangle = \langle \varphi, H_n \mathcal{P}_n \varphi \rangle \to \mathcal{H}(\varphi, \varphi). \tag{3.40}$$

Proof. Since φ is compactly supported we have $R_n \varphi = \varphi$ for n large and the R_ns may be dropped. By assumption $D_n f_n$ is L^2 bounded and $D_n f_n \to f'$ weakly in L^2, so by the preceding observations $D_n \varphi \to_{L^2} \varphi'$ and

$$\left\langle \varphi, D_n^\dagger D_n f_n \right\rangle = \langle D_n \varphi, D_n f_n \rangle \to \langle \varphi', f' \rangle .$$

For the potential term we must verify that

$$\langle \varphi, V_n f_n \rangle = \left\langle \varphi, \left(\triangle_n Y_{n,1} + \tfrac{1}{2}\big((\triangle_n Y_{n,2}) T_n + T_n^\dagger (\triangle_n Y_{n,2}^\dagger) \big) \right) f_n \right\rangle$$

converges to $-\langle \varphi', Y f \rangle - \langle \varphi, Y f' \rangle$. Recall by Assumption 1 (3.20) and (3.25) that $Y_{n,i} \to Y_i$ compact-uniformly ($i = 1, 2$) and $Y = Y_1 + \frac{1}{2}(Y_2 + Y_2^\dagger)$. Writing $Y_n = Y_{n,1} + \frac{1}{2}(Y_{n,2} + Y_{n,2}^\dagger) \to Y$ (and disregarding the notational collision with Y_i), we first approximate V_n by $\triangle Y_n$:

$$\begin{aligned} \langle \varphi, (\triangle_n Y_n) f_n \rangle &= m_n \big(\langle \varphi, Y_n f_n \rangle - \langle T_n \varphi, Y_n T_n f_n \rangle \big) \\ &= m_n \big(\langle \varphi, Y_n f_n \rangle - \langle T_n \varphi, Y_n f_n \rangle + \langle T_n \varphi, Y_n f_n \rangle - \langle T_n \varphi, Y_n T_n f_n \rangle \big) \\ &= - \langle D_n \varphi, Y_n f_n \rangle - \langle T_n \varphi, Y_n D_n f_n \rangle , \end{aligned}$$

which converges to the desired limit by the observations preceding the lemma together with the assumptions on f_n and the fact that $T_n \varphi \to_{L^2} \varphi$ in L^2 since $m_n \|T_n \varphi - \varphi\| = \|D_n \varphi\|$ is bounded. The error in the above approximation comes as a sum of T_n and $T_n^\dagger$ terms. Consider twice the T_n term:

$$\begin{aligned} |\langle \varphi, (\triangle_n Y_{n,2})(T_n - 1) f_n \rangle| &= \left| \left\langle \varphi, (m_n^{-1} \triangle_n Y_{n,2}) D_n f_n \right\rangle \right| \\ &\le \|\varphi\| \sup_I \left| Y_{n,2} - T_n^\dagger Y_{n,2} \right| \|D_n f_n\| \end{aligned}$$

where I is a compact interval supporting φ. (The single bars in the supremum denote the spectral or ℓ_2-operator norm, which is of course equivalent to the max norm on the entries.) Note that $D_n f_n$ is L^2-bounded because it converges weakly in L^2. Now $Y_{n,2}$ and $T_n^\dagger Y_{n,2}$ both converge to Y_2 uniformly on I, in the latter case by the uniform continuity of Y_2 on I; it follows that the supremum, and hence the whole term, vanish in the limit. The $T_n^\dagger$ term is handled similarly, the only difference being that the D_n in the estimate lands on φ instead.

Finally, for the boundary terms Assumption 3 gives

$$(\mathcal{P}_n\varphi)_i^*(0)\, w_{n,i}\, f_{n,i}(0) \to \varphi_i^*(0)\, w_i\, f_i(0),$$

where in the Dirichlet case $i > r_0$ the left side vanishes for n large because φ_i is supported away from 0.

Turning to the second statement, we must verify that $\mathcal{P}_n\varphi \to \varphi$ as in Lemma 3.3.15. The uniform L_n^* bound on $\mathcal{P}_n\varphi$ follows from the following observations: $\left\|(\mathcal{P}_n\varphi)\sqrt{1+\overline{\eta}}\right\| = \left\|\mathcal{P}_n\varphi\sqrt{1+\overline{\eta}}\right\| \le \left\|\varphi\sqrt{1+\overline{\eta}}\right\|$; for n large enough that $R_n\varphi = \varphi$ we have $\|D_n\mathcal{P}_n\varphi\| = \|\mathcal{P}_n D_n\varphi\| \le \|D_n\varphi\| \le \|\varphi'\|$ (Young's inequality); for the boundary term note that $(\mathcal{P}_n\varphi)_i(0)$ is bounded if $i \le r_0$ and in fact vanishes for n large if $i > r_0$. The convergence is easy: $\mathcal{P}_n\varphi \to \varphi$ compact-uniformly and in L^2, and for $g \in L^2$ we have $\langle g, D_n\mathcal{P}_n\varphi\rangle = \langle \mathcal{P}_n g, D_n\varphi\rangle \to \langle g, \varphi'\rangle$. □

We finish by recalling the argument to put all the pieces together. A technical point: unlike in previous treatments we do not assume that the eigenvalues are simple.

Proof of Theorem 3.3.9. We first show that for all k we have $\underline{\lambda}_k = \liminf \lambda_{n,k} \ge \Lambda_k$. Assume that $\underline{\lambda}_k < \infty$. The eigenvalues of H_n are uniformly bounded below by Lemma 3.3.13, so there is a subsequence along which $(\lambda_{n,1}, \dots, \lambda_{n,k}) \to (\xi_1, \dots, \xi_k = \underline{\lambda}_k)$. By the same lemma, corresponding orthonormal eigenvector sequences have L_n^*-norm uniformly bounded. Pass to a further subsequence so that they all converge as in Lemma 3.3.15. The limit functions are orthonormal; by Lemma 3.3.16 they are eigenfunctions with eigenvalues $\xi_j \le \underline{\lambda}_k$ and we are done.

We proceed by induction, assuming the conclusion of the theorem up to $k-1$. For $j = 0, \dots, k-1$ let $v_{n,j}$ be orthonormal eigenvectors corresponding to $\lambda_{n,j}$; for any subsequence we can pass to a further subsequence such that $v_{n,j} \to_{L^2} f_j$, eigenfunctions corresponding to Λ_j. Take an orthogonal eigenfunction f_k corresponding to Λ_k and find

$f_k^\varepsilon \in C_0^\infty$ with $\|f_k^\varepsilon - f_k\|_* < \varepsilon$. Consider the vector

$$f_{n,k} = \mathcal{P}_n f_k^\varepsilon - \sum_{j=0}^{k-1} \langle v_{n,j}, \mathcal{P}_n f_k^\varepsilon \rangle \, v_{n,j}.$$

The L_n^*-norm of the sum term is uniformly bounded by $C\varepsilon$: indeed, the $\|v_{n,j}\|_{*n}$ are uniformly bounded by Lemma 3.3.13, while the coefficients satisfy $|\langle v_{n,j}, f_k^\varepsilon \rangle| \leq \|f_k^\varepsilon - f_k\| + \|v_{n,j} - f_j\| < 2\varepsilon$ for large n. By the variational characterization in finite dimensions and the uniform L_n^* form bound on $\langle \cdot, H_n \cdot \rangle$ (by Lemma 3.3.13) together with the uniform bound on $\|\mathcal{P}_n f_k^\varepsilon\|_{*n}$ (by Lemma 3.3.16), we then have

$$\limsup \lambda_{n,k} \;\leq\; \limsup \frac{\langle f_{n,k}, H_n f_{n,k} \rangle}{\langle f_{n,k}, f_{n,k} \rangle} \;=\; \limsup \frac{\langle \mathcal{P}_n f_k^\varepsilon, H_n \mathcal{P}_n f_k^\varepsilon \rangle}{\langle \mathcal{P}_n f_k^\varepsilon, \mathcal{P}_n f_k^\varepsilon \rangle} + o_\varepsilon(1), \tag{3.41}$$

where $o_\varepsilon(1) \to 0$ as $\varepsilon \to 0$. But (3.40) of Lemma 3.3.16 provides $\lim \langle \mathcal{P}_n f_k^\varepsilon, H_n \mathcal{P}_n f_k^\varepsilon \rangle = \mathcal{H}(f_k^\varepsilon, f_k^\varepsilon)$, so the right hand side of (3.41) is

$$\frac{\mathcal{H}(f_k^\varepsilon, f_k^\varepsilon)}{\langle f_k^\varepsilon, f_k^\varepsilon \rangle} + o_\varepsilon(1) \;=\; \frac{\mathcal{H}(f_k, f_k)}{\langle f_k, f_k \rangle} + o_\varepsilon(1) \;=\; \Lambda_k + o_\varepsilon(1).$$

Now letting $\varepsilon \to 0$, we conclude $\limsup \lambda_{n,k} \leq \Lambda_k$.

Thus $\lambda_{n,k} \to \Lambda_k$; Lemmas 3.3.13 and 3.3.15 imply that any subsequence of the $v_{n,k}$ has a further subsequence converging in L^2 to some $f \in L^*$; Lemma 3.3.16 then implies that f is an eigenfunction corresponding to Λ_k. Finally, convergence is uniform over $W_n, W \geq -b$ since the bound 3.3.13 is. □

3.4 CLT and tightness for Gaussian and Wishart models

We now verify Assumptions 1–3 of Section 3.3 for the band Jacobi forms of Section 3.2 and thus prove Theorems 3.1.2 and 3.1.3 via Theorem 3.3.9.

We must consider the band forms as $(r \times r)$-block tridiagonal matrices. This amounts to reindexing the entries by $(k + rj, l + rj)$, where $j \in \mathbb{Z}_+$ indexes the blocks and $1 \leq k, l \leq r$ give the index within each block. The scalar processes obtained by fixing k, l can then be analyzed jointly; finally, they can be assembled into a matrix-valued process.

The technical tool we use to establish (3.20) is a functional central limit theorem for convergence of discrete time processes with independent increments of given mean and variance (and controlled fourth moments) to Brownian motion plus a nice drift.

Appearing as Corollary 6.1 in RRV, it is just a tailored version of a much more general result given as Theorem 7.4.1 in Ethier and Kurtz (1986). We record it here.

Proposition 3.4.1. *Let $a \in \mathbb{R}$ and $h \in C^1(\mathbb{R}_+)$, and let y_n be a sequence of discrete time real-valued processes with $y_{n,0} = 0$ and independent increments $\delta y_{n,j} = y_{n,j} - y_{n,j-1} = m_n^{-1} \triangle_n y_{n,j}$. Assume that $m_n \to \infty$ and*

$$m_n \, \mathbf{E}\, \delta y_{n,j} = h'(j/m_n) + o(1), \qquad m_n \, \mathbf{E}(\delta y_{n,j})^2 = a^2 + o(1), \qquad m_n \, \mathbf{E}(\delta y_{n,j})^4 = o(1)$$

uniformly for j/m_n on compact sets as $n \to \infty$. Then $y_n(x) = y_{n,\lfloor m_n x \rfloor}$ converges in law, with respect to the compact-uniform topology, to the process $h(x) + ab_x$ where b_x is a standard Brownian motion.

Remark 3.4.2. Since the limit is a.s. continuous, Skorokhod convergence (the topology used in the references) implies uniform convergence on compact intervals (see Theorem 3.10.2 in Ethier and Kurtz 1986) and we may as well speak in terms of the latter.

The Gaussian case

Take $G_n = G_{n;0} + \sqrt{n} P_n$ as in(3.13) with $G_{n;0}$ as in (3.12) and $P_n = \tilde{P}_n \oplus 0_{n-r}$. We denote upper-left $r \times r$ blocks with a tilde throughout. Set

$$m_n = n^{1/3}, \qquad H_n = \frac{m_n^2}{\sqrt{n}} \left(2\sqrt{n} - G_n \right).$$

As usual, this soft-edge scaling can be predicted as follows. Centering G_n by $2\sqrt{n}$ gives, to first order, $\sqrt{n}$ times the discrete Laplacian on blocks of size r. With space scaled down by m_n, the Laplacian must be scaled up by m_n^2 to converge to the second derivative. Finally, the scaling $m_n = n^{1/3}$ is determined by convergence of the next order terms to the noise and drift parts of the limiting potential.

Decompose H_n as in (3.18),(3.19). The upper-left block is

$$\tilde{H}_n = m_n^2 + m_n(W_n + Y_{n,1;0}) = m_n^2(2 - n^{-1/2}\tilde{G}_{n,0} - \tilde{P}_n);$$

we want the boundary term W_n to absorb the "extra" m_n^2 (the 2 in the right hand side "should be" a 1) and the perturbation in order to make $Y_{n,1;0}$ small just like the subsequent increments of $Y_{n,i}$). We therefore set

$$W_n = m_n \big(1 - \tilde{P}_n\big).$$

With this choice Assumption 3 is an immediate consequence of the hypotheses of Theorem 3.1.2. The processes $Y_{n,1}, Y_{n,2}$ are determined and it remains to verify Assumptions 1 and 2.

We begin with Assumption 1, identifying the limiting integrated potential $Y : \mathbb{R}_+ \to M_r(\mathbb{F})$ as that of the multivariate stochastic Airy operator:

$$Y(x) = \sqrt{2}B_x + \tfrac{1}{2}rx^2 \tag{3.42}$$

where B_x is a standard $M_r(\mathbb{F})$ Brownian motion and second term is a scalar matrix.

Proof of (3.20)*, Gaussian case.* Define scalar processes $y_{k,l}$ for $1 \le l \le r$ and $l \le k \le l+r$ by

$$y_{k,l} = \begin{cases} \left(Y_{n,1}\right)_{k,l} & l \le k \le r \\ \left(\frac{1}{2}Y_{n,2}^\dagger\right)_{k-r,l} & r+1 \le k \le l+r. \end{cases} \tag{3.43}$$

(We have dropped the subscript n.) Equivalently, for $1 \le k, l \le r$,

$$\left(Y_{n,1}\right)_{kl} = \begin{cases} y_{l,k}^* & k \le l \\ y_{k,l} & k \ge l, \end{cases} \qquad \left(\tfrac{1}{2}Y_{n,2}^\dagger\right)_{kl} = \begin{cases} y_{k+r,l} & k \le l \\ 0 & k > l. \end{cases} \tag{3.44}$$

Then we have

$$\delta y_{k,l;j} = n^{-1/6} \begin{cases} -\frac{2}{\beta}\,\widetilde{g}_{k+rj} & k = l \\ -g_{k+rj,l+rj} & l < k < l+r \\ \left(\sqrt{n} - \frac{1}{\sqrt{\beta}}\chi_{(n-k-rj+1)\beta}\right) & k = l+r. \end{cases} \tag{3.45}$$

Note that the $y_{k,l}$ are independent increment processes that are mutually independent of one another. With the usual embedding $j = \lfloor n^{1/3}x \rfloor$, Proposition 3.4.1 together with standard moment computations for Gaussian and Gamma random variables—in particular

$$\mathbf{E}\,\chi_\alpha = \sqrt{\alpha} + O(1/\sqrt{\alpha}), \quad \mathbf{E}(\chi_\alpha - \sqrt{\alpha})^2 = 1/2 + O(1/\alpha), \quad \mathbf{E}(\chi_\alpha - \sqrt{\alpha})^4 = O(1),$$

for α large (valid since we consider $j = O(n^{1/3})$ here)—leads to the convergence of processes

$$y_{k,l}(x) \Rightarrow \begin{cases} \sqrt{\frac{2}{\beta}}\,\widetilde{b}_k(x) & k = l \\ b_{k,l}(x), & l < k < l+r \\ \frac{1}{\sqrt{2\beta}}\,b_k(x) + \frac{1}{4}rx^2 & k = l+r \end{cases}$$

where $b_k, \widetilde{b}_k$ are standard real Brownian motions and $b_{k,l}$ are standard $\mathbb{F}$ Brownian motions. By independence, the convergence occurs jointly over k, l and the limiting Brownian motions are all independent. (For the $\mathbb{F}$ Brownian motions apply Proposition 3.4.1 to each of the β real components, which are independent of one another.) Therefore $Y_{n,i}$ are both tight, and using (3.44) we have, jointly for $1 \le k, l \le r$,

$$\big(Y_{n,1} + \tfrac{1}{2}(Y_{n,2}^\dagger + Y_{n,2})\big)_{k,l} = \begin{cases} y_{k,k} + 2y_{k+r,k} \\ y_{k,l} + y_{l+r,k}^* \\ y_{l,k}^* + y_{k+r,l} \end{cases} \Rightarrow \begin{cases} \sqrt{\frac{2}{\beta}}(\widetilde{b}_k + b_k) + \frac{1}{2}rx^2 & k = l \\ b_{k,l} + b_{l+r,k}^* & k > l \\ b_{l,k}^* + b_{k+r,l} & k < l. \end{cases}$$

Noting that the two Brownian motions in each entry are independent and that the entries on and below the diagonal are independent of each other, we conclude that this limiting matrix process is distributed as $Y(x)$ in (3.42). □

We turn to Assumption 2. Here we need bounds over the full range $0 \le j \le \lceil n/r \rceil - 1$. Recall that we can extend the $Y_{n,i}$ processes beyond the end of the matrix arbitrarily (R_n takes care of the truncation), and it is convenient to "continue the pattern" for an extra block or two by setting $\chi_\alpha = 0$ for $\alpha < 0$. For the decomposition (3.21), we simply take $\eta_{n,i}$ to be the expectation of $\triangle Y_{n,i}$ and $\triangle\omega_{n,i}$ to be its centered version; the components of $\eta_{n,i}$ are then easily estimated and those of $\omega_{n,i}$ become independent increment martingales. We further set $\overline{\eta}(x) = rx$.

Proof of (3.22)–(3.24)*, Gaussian case.* From (3.45) we have $\eta_{n,1;j} = 0$ and

$$(\eta_{n,2;j})_{k,l} = \mathbf{E}\, 2m_n \delta y_{k+r,l;j} = 2n^{1/6}\big(\sqrt{n} - \beta^{-1/2}\,\mathbf{E}\,\chi_{(n-k-r(j+1)+1)\beta}\big)\,1_{k=l}.$$

The estimate

$$\sqrt{(\alpha-1)^+} \;\le\; \mathbf{E}\,\chi_\alpha = \sqrt{2}\frac{\Gamma((\alpha+1)/2)}{\Gamma(\alpha/2)} \;\le\; \sqrt{\alpha}. \tag{3.46}$$

is useful. We obtain

$$2n^{1/6}\frac{rj - c}{2\sqrt{n}} \le (\eta_{n,2;j})_{k,k} \le 2n^{1/6}\frac{rj + c}{\sqrt{n}}$$

for some fixed c, which yields the matrix inequalities

$$rx - cn^{-1/3} \le \eta_{n,2}(x) \le 2rx + cn^{-1/3}$$

and verifies (3.22) with $\overline{\eta}(x) = rx$. Separately, we have the upper bound (3.23):

$$\eta_{n,2}(x) \le 2n^{2/3} = 2m_n^2.$$

The bound (3.24) may be done entry by entry, so we consider the process $\{(\omega_{i,n;j})_{k,l}\}_{j\in\mathbb{Z}_+}$ for fixed $i = 1, 2$ and $1 \le k, l \le r$ and further omit these indices; for the $\mathbb{F}$-valued processes we restrict attention further to one of the β real-valued components, and denote the latter simply by $\omega_{n;j}$. Consider (3.45); the key points are that the increments $\delta\omega_{n;j}$ are independent and centered, and that scaled up by $n^{1/6} = m_n^{1/2}$ they have uniformly bounded fourth moments. To prove (3.24) it is enough to consider x at integer points and show that the random variables

$$\sup_{x=0,1,\dots,n/rm_n} x^{\varepsilon-1} \sup_{j=1,\dots,m_n} |\omega_{n;m_nx+j} - \omega_{n;m_nx}|^2$$

are tight over n. Squaring, bounding the outer supremum by the corresponding sum, and then taking expectations gives

$$\sum_{x=0}^{n/rm_n} \frac{\mathbf{E}\sup_{j=1,\dots,m_n} |\omega_{n;m_nx+j} - \omega_{n;m_nx}|^4}{x^{2-2\varepsilon}} \le \sum_{x=0}^{n/rm_n} \frac{16\,\mathbf{E}\left|\omega_{n;m_n(x+1)} - \omega_{n;m_nx}\right|^4}{x^{2-2\varepsilon}},$$

where we have used the L^p maximum inequality for martingales (see e.g. Proposition 2.2.16 of Ethier and Kurtz 1986). To bound the latter expectation, expand the fourth power to obtain $O(m_n^2)$ nonzero terms that are $O(m_n^{-2})$ with constants independent of x and n. It follows that the entire sum is uniformly bounded over n, as required. □

The Wishart case

Take $L_{n,p} = \Sigma_{n,p}^{1/2} L_{n,p,0}$ with $L_{n,p,0}$ as in (3.14) and, denoting the upper-left $r \times r$ block with a tilde, $\Sigma_{n,p} = \tilde{\Sigma}_{n,p} \oplus I_{n\wedge p}$. Recall that $L_{n,p}$ is $((n + r) \wedge p) \times (n \wedge p)$. Put $S_{n,p} = L_{n,p}^\dagger L_{n,p}$ and similarly for $S_{n,p,0}$; these matrices are $(n \wedge p) \times (n \wedge p)$ and the latter is given explicitly in (3.15). We sometimes drop the subscripts n, p. Recall (3.16) that $S - S_0 = \tilde{L}_0^\dagger(\tilde{\Sigma} - 1)\tilde{L}_0 \oplus 0$.

We set

$$m_{n,p} = \left(\frac{\sqrt{np}}{\sqrt{n} + \sqrt{p}}\right)^{2/3}, \quad H_{n,p} = \frac{m_{n,p}^2}{\sqrt{np}} \left(\left(\sqrt{n} + \sqrt{p}\right)^2 - S_{n,p}\right). \tag{3.47}$$

See Chapter 2 for detailed heuristics behind the scaling; written in this way, it allows that $p, n \to \infty$ together arbitrarily, i.e. only $n \wedge p \to \infty$. It is useful to note that

$$2^{-2/3}(n \wedge p)^{1/3} \le m_{n,p} \le (n \wedge p)^{1/3}.$$

Decompose $H_{n,p}$ as in (3.18),(3.19). The upper-left block is

$$\tilde{H} = m^2 + m(W + Y_{1;0}) = 2m^2 - \frac{m^2}{\sqrt{np}}\big(\tilde{S}_0 - n - p + \tilde{L}_0^\dagger(\tilde{\Sigma} - 1)\tilde{L}_0\big).$$

As before we want W to absorb the extra m^2 and the perturbation in order to make $Y_{1;0}$ small. Now the perturbation term is random, but it does not have to be fully absorbed; it is enough that $Y_{1;0} \to 0$ in probability. The reason is that the process Y_1 can absorb an overall additive random constant that tends to zero in probability, as is clear in Assumption 1 while in Assumption 2 the constant may be put into ω_1. Since $\tilde{L}_0 \approx \sqrt{n}$, we set

$$W_{n,p} = m_{n,p}\left(1 - \sqrt{n/p}\big(\tilde{\Sigma}_{n,p} - 1\big)\right). \tag{3.48}$$

Once again, Assumption 3 follows immediately from the hypotheses of Theorem 3.1.3.

We must still deal with the perturbed term in $Y_{1;0}$ and show that

$$\frac{m}{\sqrt{np}}\big(n\tilde{\Sigma} - \tilde{L}_0^\dagger\tilde{\Sigma}\tilde{L}_0\big) \to 0 \tag{3.49}$$

in probability. We defer this to the end of the proof of Assumption 1, to which we now turn. As in the Gaussian case, Y is given by (3.42).

Proof of (3.20)*, Wishart case.* By the preceding paragraph it suffices to treat the null case $\Sigma = I$ and afterwards check (3.49). Define processes $y_{k,l}$ for $1 \le l \le r$ and $l \le k \le l + r$ by (3.43) as in the Gaussian case. From (3.15) with the centering and scaling of (3.47) and (3.19) we obtain

$$\delta y_{k,l;j} = \frac{m}{\sqrt{np}}\begin{cases} n + p - \frac{1}{\beta}\big(\widetilde{\chi}^2_{(n-k-rj+1)\beta} + \chi^2_{(p-k-r(j+1)+1)\beta}\big) + O(1) & k = l \\ -\frac{1}{\sqrt{\beta}}\big(\widetilde{\chi}_{(n-k-rj+1)\beta}g_{k+rj,l+rj} & \\ \qquad +\chi_{(p-l-r(j+1)+1)\beta}g^*_{l+r(j+1),k+rj}\big) + O(1) & l < k < l + r \\ \sqrt{np} - \frac{1}{\beta}\widetilde{\chi}_{(n-k-rj+1)\beta}\chi_{(p-k-rj+1)\beta} & k = l + r \end{cases}$$

where the $O(1)$ terms stand in for the interior Gaussian sums of (3.15), all of whose *moments* are bounded uniformly in n, p. Since $m^{1+k}/(np)^{k/2} \le m^{1-2k} = o(1)$ for $k \ge 1$, these terms are negligible in the scaling of Proposition 3.4.1 in the sense that the associated processes converge to the zero process. Next, use that expressions of type $\chi_n - \sqrt{n}$ are $O(1)$ in the same sense, and that $\sqrt{n} - \sqrt{n-j} = O(j/\sqrt{n}) = O(m/\sqrt{n}) =$

$o(1)$ since we consider j/m bounded here (and similarly for p), to write

$$\delta y_{k,l;j} = \frac{m}{\sqrt{np}} \begin{cases} \frac{2}{\sqrt{\beta}}\big(\sqrt{n}(\sqrt{\beta n} - \widetilde{\chi}_{(n-k-rj+1)\beta}) \\ \qquad +\sqrt{p}(\sqrt{\beta p} - \chi_{(p-k-r(j+1)+1)\beta})\big) + O(1) & k = l \\ -\sqrt{n}\, g_{k+rj,l+rj} - \sqrt{p}\, g^*_{l+r(j+1),k+rj} + O(1) & l < k < l+r \\ \frac{1}{\sqrt{\beta}}\big(\sqrt{p}(\sqrt{\beta n} - \widetilde{\chi}_{(n-k-rj+1)\beta}) \\ \qquad +\sqrt{n}(\sqrt{\beta p} - \chi_{(p-k-rj+1)\beta})\big) + O(1) & k = l+r \end{cases} \tag{3.50}$$

It suffices prove tightness and convergence in distribution along a subsequence of any given subsequence, and we may therefore assume that $p/n \to \gamma^2 \in [0, \infty]$. Each case of (3.50) contains two terms, and each one of these terms forms an independent increment process to which Proposition 3.4.1 may be applied. (Break the $\mathbb{F}$-valued terms up further into their real-valued parts.) Standard moment computations as in the Gaussian case, together with independence, then lead to the joint convergence of processes

$$y_{k,l}(x) \Rightarrow \begin{cases} \sqrt{\frac{2}{\beta}}\big(\frac{1}{1+\gamma}\widetilde{b}_k(x) + \frac{\gamma}{1+\gamma} b_k(x)\big) + \frac{\gamma}{(1+\gamma)^2} r x^2 & k = l \\ \frac{1}{1+\gamma} b_{k,l}(x) + \frac{\gamma}{1+\gamma} b^*_{l+r,k}(x) & l < k < l+r \\ \frac{1}{\sqrt{2\beta}}\big(\frac{\gamma}{1+\gamma}\widetilde{b}_k(x) + \frac{1}{1+\gamma} b_k(x)\big) + \frac{1+\gamma^2}{4(1+\gamma)^2} r x^2 & k = l+r \end{cases}$$

where $b_k, \widetilde{b}_k$ are standard real Brownian motions and $b_{k,l}$ are standard $\mathbb{F}$ Brownian motions, all independent except that $b_{k+r,l+r}$ and $b_{k,l}$ are identified. Therefore $Y_{n,i}$ are both tight. Furthermore, using (3.44) we have

$$\big(Y_{n,1} + \tfrac{1}{2}(Y^\dagger_{n,2} + Y_{n,2})\big)_{k,l} = \begin{cases} y_{k,k} + 2y_{k+r,k} \\ y_{k,l} + y^*_{l+r,k} \\ y^*_{l,k} + y_{k+r,l} \end{cases} \Rightarrow \begin{cases} \sqrt{\frac{2}{\beta}}(\widetilde{b}_k + b_k) + \frac{1}{2} r x^2 & k = l \\ b_{k,l} + b^*_{l+r,k} & k > l \\ b^*_{l,k} + b_{k+r,l} & k < l \end{cases}$$

jointly for $1 \le k, l \le r$. After the dust clears we thus arrive at exactly the same limiting process as in the Gaussian case, namely (3.42).

We now address (3.49). Here we can replace $\tilde{L}_0$ with $\sqrt{n} I_r$ at the cost of an error that has uniformly bounded second and fourth moments. Now (3.48) and the assumed lower bound on $W_{n,p}$ give that $\tilde{\Sigma} \le 1 + 2\sqrt{p/n}$ for n, p large; this matrix inequality holds entrywise in the diagonal basis for $\tilde{\Sigma}$ (which was fixed over n, p). One therefore obtains error terms with mean square $O(m^2/n + m^2/p) = O(m^{-1})$ which is $o(1)$ as required. $\square$

Turning to Assumption 2, we may continue the processes $Y_{n,i}$ past the end of the matrix for convenience just as in the Gaussian case. The Wishart case presents an additional issue at the "end" of the matrix: recall that the final r rows and columns of S in (3.15) may have some apparently non-zero terms set to zero. However, these changes are easily absorbed into the bounds that follow. For (3.21) we once again take $\eta_{n,i}$ to be the expectation of $\triangle Y_{n,i}$ and $\triangle\omega_{n,i}$ to be its centered version. We also set $\overline{\eta}(x) = rx$ as before.

Proof of (3.22)–(3.24)*, Wishart case.* This time we have

$$(\eta_{n,1;j})_{k,l} = \mathbf{E}\, m\, \delta y_{k,l;j} = m^2(np)^{-1/2}\big(2rj - r + 1\big)\, 1_{k=l},$$

$$(\eta_{n,2;j})_{k,l} = \mathbf{E}\, 2m\, \delta y_{k+r,l;j} = 2m^2\big(1 - \beta^{-1}(np)^{-1/2}\, \mathbf{E}\, \widetilde{\chi}_{(n-k-rj+1)\beta}\chi_{(p-k-rj+1)\beta}\big)1_{k=l}.$$

Using (3.46) one finds, for some constant c, that

$$m^{-1}(rj + c) \le (\eta_{n,1;j} + \eta_{n,2;j})_{k,k} \le m^{-1}(2rj + c)$$

which yields (3.22) with $\overline{\eta}(x) = rx$. Separately, we have the upper bound (3.23). The oscillation bound (3.24) may be proved exactly as in the Gaussian case: we have once again that $\{\sqrt{m_n}(\omega_{n,i;j})_{k,l}\}_{j\in\mathbb{Z}_+}$ are martingales with independent increments whose fourth moments are uniformly bounded. □

3.5 Alternative characterizations of the laws

In this section we derive the SDE and PDE characterizations, proving Theorems 3.1.5 and 3.1.6.

First order linear ODE

For each noise path B_x, the eigenvalue equation $\mathcal{H}_{\beta,W} f = \lambda f$ can be rewritten as a first-order linear ODE with continuous coefficients. We begin with the formal second order linear differential equation

$$f''(x) = (x - \lambda + \sqrt{2}B'_x)f(x) \tag{3.51}$$

where $f : \mathbb{R}_+ \to \mathbb{F}^r$, with initial condition

$$f'(0) = Wf(0). \tag{3.52}$$

As usual we allow $W \in M_r^*(\mathbb{F})$ and interpret (3.52) via (3.29) . Rewrite (3.51) in the form

$$\big(f' - \sqrt{2}Bf\big)' = (x-\lambda)f - \sqrt{2}Bf'.$$

Now let $g = f' - \sqrt{2}Bf$. The equation becomes

$$\begin{aligned} g' &= (x-\lambda)f - \sqrt{2}Bf' \\ &= (x-\lambda-2B^2)f - \sqrt{2}Bg. \end{aligned}$$

In other words, the pair $\big(f(x), g(x)\big)$ formally satisfies the first order linear system

$$\begin{bmatrix} f' \\ g' \end{bmatrix} = \begin{bmatrix} \sqrt{2}B & 1 \\ x-\lambda-2B^2 & -\sqrt{2}B \end{bmatrix} \begin{bmatrix} f \\ g \end{bmatrix}. \tag{3.53}$$

Since $B_0 = 0$, g simply replaces f' in the initial condition (3.52). If one prefers, this condition can be written in the standard form

$$-\tilde{W}f(0) + \tilde{I}g(0) = 0 \tag{3.54}$$

where $\tilde{W} = \sum_{i\le r_0} w_i u_i u_i^\dagger + \sum_{i>r_0} u_i u_i^\dagger$ and $\tilde{I} = \sum_{i\le r_0} u_i u_i^\dagger$.

One could allow general measurable coefficients and define a solution to be a pair of absolutely continuous functions (f, g) satisfying (3.53) Lebesgue a.e. This definition, equivalent to writing (3.53) in an integrated form, is easily seen to coincide with (3.34). As in Remark 3.3.5, however, we note the coefficients are continuous; solutions may therefore be taken to satisfy (3.53) everywhere and are in fact continuously differentiable. It is classical that the initial value problem has a unique solution which exists for all $x \in \mathbb{R}_+$ (and further depends continuously on the parameter λ and the initial condition W).

Matrix oscillation theory

The matrix generalization of Sturm oscillation theory goes back to the classic work of Morse Morse (1932) (see also Morse 1973). Textbook treatments of self-adjoint differential systems include that of Reid (1971). Our reference will be the paper of Baur and Kratz (1989), which allows sufficiently general boundary conditions.

We first consider the eigenvalue problem on a finite interval $[0, L]$ with Dirichlet boundary condition $f(L) = 0$ at the right endpoint. In the scalar-valued setting, the number of eigenvalues below λ is found to coincide with the number of zeros of f (the solution of the initial value problem) that lie in $(0, L)$. The correct generalization to the

matrix-valued setting involves tracking a matrix whose columns form a basis of solutions, and counting the so-called "focal points".

We need a little terminology and a few facts from Baur and Kratz (1989), especially Definition 1 on p. 338 there and the points that follow. A **matrix solution** of (3.53) is a pair $F, G : \mathbb{R}_+ \to \mathbb{F}^{r\times r}$ such that each column of $\left[\begin{smallmatrix} F \\ G \end{smallmatrix}\right]$ is a solution. A **conjoined basis** for (3.53) is a matrix solution (F, G) with the additional properties that $F^\dagger G = G^\dagger F$ and $\operatorname{rank}\left[\begin{smallmatrix} F \\ G \end{smallmatrix}\right] = r$. The latter properties hold identically on $\mathbb{R}_+$ as soon as they do at a single point; in particular, we may set $F(0) = \tilde{I}$ and $G(0) = \tilde{W}$ to obtain a conjoined basis for the initial condition (3.54). A point $x \in \mathbb{R}_+$ is called a **focal point** if $F(x)$ is singular, of **multiplicity** nullity $F(x)$. The following proposition summarizes what we need from the more general results of Baur and Kratz (1989).

Proposition 3.5.1. *Consider the differential system*

$$\begin{bmatrix} f' \\ g' \end{bmatrix} = \begin{bmatrix} A & B \\ C - C_0\lambda & -A^\dagger \end{bmatrix} \begin{bmatrix} f \\ g \end{bmatrix}$$

with real parameter λ, where $A(x), B(x), C(x), C_0(x)$ are $n \times n$ matrices depending continuously on $x \in \mathbb{R}$ with B, C, C_0 Hermitian and $B, C_0 > 0$. For each $\lambda \in \mathbb{R}$, let (F, G) be a conjoined basis with some fixed initial condition at 0. Consider also the associated eigenvalue problem on $[0, L]$ with the same boundary condition at 0 and Dirichlet condition $f = 0$ at L. Then, for all $\lambda \in \mathbb{R}$, the number of focal points of (F, G) in $(0, L)$ equals the number of eigenvalues below λ. Furthermore, the spectrum is purely discrete and bounded below with eigenvalues tending to infinity.

Proof. The idea is that focal points are isolated and move continuously to the left as λ increases. For sufficiently negative λ there are no focal points on $(0, L]$; each time λ passes an eigenvalue, a new focal point is introduced at L.

We indicate how the proposition follows from the results of Baur and Kratz (1989). Note that conditions (A1), (A2) on p. 337 are satisfied by our coefficients, and that (A3) on p. 340 is satisfied by our boundary conditions. Theorem 1 on p. 345 thus applies. See (3.5) on p. 341 for the definition of $\Lambda(\lambda)$; the Dirichlet condition at L gives the particularly simple result that the right hand side of (4.1) vanishes, so the quantity $n_2(\lambda)$ is constant. Theorem 2 applies as well, and we obtain $n_1(\lambda) - n_1 = n_3(\lambda)$. Here $n_1(\lambda)$ is the number of focal points in $[0, L)$, $n_1 = \lim_{\lambda\to-\infty} n_1(\lambda)$ and $n_3(\lambda)$ is the number of eigenvalues below λ. To finish we consult Theorem 3 on p. 353; noting that (A4′)

is satisfied by Section 7.2, p. 365, to find that n_1 is simply the multiplicity of the focal point at 0. The oscillation result follows. For the assertion about the spectrum we apply Theorem 4, noting that (A5), p. 358 holds by (i) there, and (A6), p. 359 also holds . $\square$

We conclude the following for our matrix system.

Lemma 3.5.2. *Consider the eigenvalue problem* (3.53) *on* $[0, L]$ *with boundary conditions* (3.54) *and* $f(L) = 0$. *For each* $\lambda \in \mathbb{R}$, *let* (F, G) *be the conjoined basis initialized by* $F(0) = \tilde{I}$ *and* $G(0) = \tilde{W}$; *then the number of focal points in the interval* $(0, L)$, *counting multiplicity, equals the number of eigenvalues below* λ. *Furthermore, the spectrum is purely discrete and bounded below with eigenvalues tending to infinity.*

A soft argument now recovers an oscillation theorem for the original half-line problem.

Theorem 3.5.3. *Consider the eigenvalue problem* (3.53),(3.54) *on* $L^2(\mathbb{R}_+)$. *For each* $\lambda \in \mathbb{R}$, *let* (F, G) *be the conjoined basis as above; then the number of focal points in* $(0, \infty)$ *equals the number of eigenvalues strictly below* λ.

Proof. Let $\Lambda_{L,k}$, Λ_k, $k = 0, 1, \ldots$ denote the lowest eigenvalues of the truncated and half-line operators $\mathcal{H}_L$, $\mathcal{H}$ respectively; it suffices to show that $\lim_{L\to\infty} \Lambda_{L,k} = \Lambda_k$ for each k. Indeed, taking $L \to \infty$ in Lemma 3.5.2 then yields the conclusion for each $\lambda \in \mathbb{R} \setminus \{\Lambda_0, \Lambda_1, \ldots\}$. Letting $\lambda \searrow \Lambda_k$, the right-most focal point must tend to ∞ by monotonicity and continuity, so the claim actually holds for all $\lambda \in \mathbb{R}$.

The variational problem for $\mathcal{H}_L$ simply minimizes over the subset of L^* functions that vanish on $[L, \infty)$; the Dirichlet condition is important here. It follows immediately that $\Lambda_{L,k} \geq \Lambda_k$, using the min-max formulation of the variational characterization. Proceed by induction, assuming that $\Lambda_{L,j} \to \Lambda_L$ for $j = 0, \ldots, k-1$.

Let $f_{L,j}$ be orthonormal eigenvectors corresponding to $\Lambda_{L,j}$. By the induction hypothesis, the variational characterization for $\mathcal{H}$ and the finite-dimensionality of its eigenspaces, every subsequence has a further subsequence such that $f_{L,j} \to_{L^2} f_j$, eigenvectors corresponding to Λ_j. Let f_k be an orthogonal eigenvector corresponding to Λ_k and take f_k^ε compactly supported with $\|f_k^\varepsilon - f_k\|_* < \varepsilon$. Let

$$g_L = f_k^\varepsilon - \sum_{j=0}^{k-1} \langle f_k^\varepsilon, f_{L,j} \rangle \, f_{L,j}.$$

For large L the inner products are at most 2ε, so $\|g_L - f_k\|_* \leq c\varepsilon$. Noting that g_L is eventually supported on $[0, L]$, the variational characterization gives

$$\limsup_{L\to\infty} \Lambda_{L,k} \leq \limsup_{L\to\infty} \frac{\mathcal{H}(g_L, g_L)}{\langle g_L, g_L \rangle}$$

and the right-hand side tends to $\mathcal{H}(f_k, f_k)/\langle f_k, f_k\rangle = \Lambda_k$ as $\varepsilon \to 0$. □

Riccati SDE: stochastic Airy meets Dyson

Let (F, G) be a conjoined basis for (3.53) as defined in the previous subsection. Then, on any interval with no focal points, the matrix $Q = GF^{-1}$ is self-adjoint and satisfies the **matrix Riccati equation**

$$Q' = rx - \lambda - (Q + \sqrt{2}B)^2. \tag{3.55}$$

(see p. 338 of Baur and Kratz 1989).

As x passes through a focal point x_0, an eigenvalue q of Q "explodes to $-\infty$ and restarts at $+\infty$". The precise evolution of Q near x_0 can be seen by choosing $a \in \mathbb{R}$ so that $\tilde{Q} = (Q-a)^{-1} = F(G-aF)^{-1}$ is defined; then $\tilde{Q}$ satisfies

$$\tilde{Q}' = \big(1 + \tilde{Q}(\sqrt{2}B + a)\big)\big(1 + (\sqrt{2}B + a)\tilde{Q}\big) - (x - \lambda)\tilde{Q}^2. \tag{3.56}$$

Writing $\tilde{q} = 1/(q-a)$ and v for the corresponding eigenvector, notice how

$$\tilde{q}'(x_0) = v(x_0)^\dagger \tilde{Q}'(x_0) v(x_0) = 1.$$

Thus $\tilde{q}$ is "pushed up through zero," corresponding to the explosion/restart in $q = 1/\tilde{q} + a$. In this way we may consider $Q(x) \in M_r^*(\mathbb{F})$ to be defined for all x. The initial condition is then simply $Q(0) = W$.

Now let $P = F'F^{-1}$. While $P = Q + \sqrt{2}B$ is not differentiable, by (3.55) it certainly satisfies the integral equation

$$P_{x_2} - P_{x_1} = \sqrt{2}(B_{x_2} - B_{x_1}) + \int_{x_1}^{x_2} (ry - \lambda - P_y^2)\, dy$$

if $[x_1, x_2]$ is free of focal points. In other words, *P is a strong solution of the Itō equation*

$$dP_x = \sqrt{2}\, dB_x + (rx - \lambda - P_x^2)dx \tag{3.57}$$

off the focal points. The evolution of P through a focal point can be described in the coordinate $\tilde{P} = (P-a)^{-1} = F(F'-aF)^{-1}$. Using (3.56) and Itō's lemma one could write down an SDE for $\tilde{P} = \tilde{Q}(1 + \sqrt{2}B\tilde{Q})^{-1}$. The initial condition here is also $P(0) = W$.

Consider the eigenvalues $p_1, \ldots, p_r$ of P. The main point is that the drift term in (3.57) is unitarily equivariant and passes through the usual derivation of Dyson's

Brownian motion (Dyson 1962). The eigenvalues therefore evolve as an autonomous Markov process.

To describe the law on paths we need a space, and there are two issues: it will be necessary to keep the eigenvalues ordered but also allow for explosions/restarts. We therefore define a sequence of **Weyl chambers** $C_k \subset (-\infty, \infty]^r$ by

$$\begin{aligned} C_0 &= \{p_1 < \cdots < p_r\} \\ C_1 &= \{p_2 < \cdots < p_r < p_1\} \\ C_2 &= \{p_3 < \cdots < p_r < p_1 < p_2\} \end{aligned}$$

and so on, permuting cyclically. We glue successive adjacent chambers together at infinity in the natural way to make the disjoint union $\mathcal{C} = C_0 \cup C_1 \cup \ldots$ into a connected smooth manifold. That is, taking $p_1 \to -\infty$ in C_0 puts you at $p_1 = +\infty$ in C_1; the smooth structure is defined by the coordinate $\tilde{p}_1 = 1/p_1$, which vanishes along the seam. Glue C_{k-1} to C_k similarly along $\{p_{k \bmod r} = \infty\}$. We also define $\overline{C}_k$, $\overline{\mathcal{C}}$ in which some coordinates may be equal, and $\partial C_k = \overline{C}_k \setminus C_k$, $\partial\mathcal{C} = \overline{\mathcal{C}} \setminus \mathcal{C}$ in which some coordinates are equal.

Theorem 3.5.4. *Represent the eigenvalues of $W \in M_r^*(\mathbb{F})$ as $\mathbf{w} = (w_1 \ldots, w_r) \in \overline{C}_0$. The eigenvalues $\mathbf{p} = (p_1, \ldots, p_r)$ of P evolve as an autonomous Markov process whose law on paths $\mathbb{R}_+ \to \overline{\mathcal{C}}$ is the unique weak solution of the SDE system*

$$dp_i = \tfrac{2}{\sqrt{\beta}} db_i + \left(rx - \lambda - p_i^2 + \sum_{j \neq i} \frac{2}{p_i - p_j} \right) dx \tag{3.58}$$

with initial condition $\mathbf{p}(0) = \mathbf{w}$, where $b_1, \ldots, b_r$ are independent standard real Brownian motions. An eigenvalue p_i can explode to $-\infty$ and restart at $+\infty$, meaning $\mathbf{p}$ crosses from C_k to C_{k+1}; the evolution through an explosion is described in the coordinate $\tilde{p}_i = 1/p_i$, which satisfies

$$d\tilde{p}_i = -\tfrac{2}{\sqrt{\beta}} \tilde{p}_i^2 db_i + \left(1 + \Big(\lambda - rx + \sum_{j \neq i} \frac{2\tilde{p}_i \tilde{p}_j}{\tilde{p}_i - \tilde{p}_j} \Big) \tilde{p}_i^2 + \tfrac{4}{\beta} \tilde{p}_i^3 \right) dx. \tag{3.59}$$

Proof. Deriving (3.58) from (3.57) is simply a matter of applying Itō's lemma, at least in $\mathcal{C}$ where the eigenvalues are distinct. One needs to differentiate an eigenvalue with respect to a matrix, and this information is given by Hadamard's variation formulas. In detail, let $A \in M_r(\mathbb{F})$ vary smoothly in time and suppose $A(0)$ has distinct spectrum.

Then eigenvalues $\lambda_1, \dots, \lambda_r$ of A and corresponding eigenvectors $v_1, \dots, v_r$ vary smoothly near 0 by the implicit function theorem. Differentiating $Av_i = \lambda_i v_i$ and $v_i^\dagger v_i = 1$ leads to the formulas

$$\dot\lambda_i = v_i^\dagger \dot A v_i, \qquad \ddot\lambda_i = v_i^\dagger \ddot A v_i + 2\sum_{j\neq i} \frac{\left|v_i^\dagger \dot A v_j\right|^2}{\lambda_i - \lambda_j}.$$

Writing $X = \dot A(0)$ and ∇_X for the directional derivative, and taking $v_1(0), \dots, v_r(0)$ to be the standard basis, we find

$$\nabla_X \lambda_i = X_{ii}, \qquad \nabla_X^2 \lambda_i = 2\sum_{j\neq i} \frac{|X_{ij}|^2}{\lambda_i - \lambda_j}.$$

Returning to (3.57), at each fixed time x we can change to the diagonal basis for P_x because the noise term is invariant in distribution and the drift term is equivariant. Itō's lemma amounts to formally writing $dp_i = \nabla_{dP}\, p_i + \frac{1}{2}\nabla^2_{dP}\, p_i$ and using that dB_{ii} are jointly distributed as $\sqrt{2/\beta}\, db_i$ for $i = 1 \dots, r$ while $|dB_{ij}|^2 = dt$ for $j \neq i$. We thus arrive at (3.58).

Recall that the evolution of P through a focal point is still described by an SDE, after changing coordinates. The same is therefore true of $\mathbf{p}$ through an explosion; the form (3.59) is obtained from (3.58) by an application of Itō's lemma.

Just as with the usual Dyson's Brownian motion, the p_i are almost surely distinct at all positive times: $\mathbf{p}(x) \in \mathcal{C}$ for all $x > 0$. One can show this "no collision property" holds for any solution of (3.58),(3.59), even with an initial condition $\mathbf{p}(0) \in \partial C_0$. (Technically, one defines an entrance law from $\partial\mathcal{C}$ by a limiting procedure.) Since the coefficients are regular inside $\mathcal{C}$, this suffices to prove uniqueness of the law. See Anderson et al. (2009), Section 4.3.1 for a detailed proof in the driftless case. □

Proof of Theorem 3.1.5. Explosions of $\mathbf{p}$ as in Theorem 3.5.4 correspond to focal points of F for each λ. By Theorem 3.5.3, the total number of explosions K is equal to the number of eigenvalues strictly below λ. (Notice that $\mathbf{p}$ ends up in C_K.) For a *fixed* λ, translation invariance of the driving Brownian motions b_i allows one to shift time $x \mapsto x - \lambda/r$ and use (3.3) started at $x_0 = -\lambda/r$. Putting $a = -\lambda$ we have $\mathbf{P}(-\Lambda_k \leq a) = \mathbf{P}(\Lambda_k \geq \lambda) = \mathbf{P}_{a/r,\mathbf{w}}(K \leq k)$ as required. □

PDE and boundary value problem

We now prove the PDE characterization, Theorem 3.1.6. We will need two properties of the eigenvalue diffusion.

Lemma 3.5.5. *Let* $\mathbf{p} : [x_0, \infty) \to \overline{\mathcal{C}}$ *have law* $\mathbf{P}_{x_0,\mathbf{w}}$ *as in* (3.3) *and let* K *be the number of explosions. Then the following hold:*

(i) Given x_0, k, $\mathbf{P}_{x_0,\mathbf{w}}(K \leq k)$ *is increasing in* $\mathbf{w}$ *with respect to the partial order* $\mathbf{w} \leq \mathbf{w}'$ *given by* $w_i \leq w_i'$, $i = 1, \ldots, r$.

(ii) $\mathbf{P}_{x_0,\mathbf{w}}$*-almost surely,* $p_1, \ldots, p_r$ *remain bounded below in* C_K *(after the last explosion), or equivalently in* C_0 *on the event* $\{K = 0\}$.

Proof. Part (i) is a consequence Theorem 3.1.5 and Remark 3.1.1, the pathwise monotonicity of the eigenvalues Λ_k as a function of the boundary parameter W with respect to the usual matrix partial order. It can also be seen from the related fact that the matrix partial order is preserved pathwise by the matrix Riccati equation (3.57), which implies that a solution started from W explodes no later than one started from $W' \geq W$. This fact holds for the P evolution if it holds for the Q evolution (3.55), and for the latter it is Theorem IV.4.1 of Reid (1972).

Part (ii) follows from the stronger assertion that $p_i \sim \sqrt{rx}$ as $x \to \infty$. In the $r = 1$ case this is Proposition 3.7 of RRV. Heuristically, the single particle drift linearizes at the stable equilibrium $\sqrt{rx}$ to $2\sqrt{rx}(\sqrt{rx} - p_i)$; even with the repulsion terms one expects fluctuations of variance only $C/\sqrt{x}$. We omit the proof. □

Proof of Theorem 3.1.6. Assume the diffusion representation of Theorem 3.1.5 for $F_\beta(x; \mathbf{w}) = \mathbf{P}(-\Lambda_0 \leq x)$ on $\mathbb{R} \times \overline{C}_0$. We first show $F = F_\beta$ has the asserted properties and afterwards argue uniqueness. Writing L for the space-time generator of (3.3), the PDE (3.6) is simply the equation $LF = 0$ after replacing x with x/r. In other words, it is the Kolmogorov backward equation for the hitting probability (3.4) (more precisely, the probability of never hitting $\{w_1 = -\infty\}$), which is L-harmonic. This extends to $w_r = +\infty$ by using the local coordinate there; from (3.59) one sees that the coefficients remain regular. Although the diffusivity vanishes at $w_r = +\infty$, the drift does not, and it follows that F is continuous up to $w_r = +\infty$. The PDE holds even at points $\mathbf{w} \in \partial C_0$ with appropriate one-sided derivatives; notice that the apparent singularity in the "Dyson term" of the PDE is in fact removable for F regular and symmetric in the w_i. (For a toy version, consider a function $f : \mathbb{R} \to \mathbb{R}$ that is twice differentiable and even; then f' is odd and $f'(w)/w$ is continuous with value $f''(0)$ at $w = 0$. These functions form the domain of the generator of the Bessel process on the half-line $\{w \geq 0\}$ in the same way that symmetric functions form the domain of the generator of Dyson's Brownian motion

on a Weyl chamber.) Finally, the picture can be copied to $\mathbf{w} \in (-\infty, \infty]^r$ by symmetry, permuting the w_i.

The boundary condition (3.7) follows from the monotonicity property of Lemma 3.5.5 (i). For fixed $\mathbf{w}$, $F(x; \mathbf{w}) \to 1$ as $x \to \infty$ because it is a distribution function in x; by monotonicity in $\mathbf{w}$, the convergence is uniform over a set of $\mathbf{w}$ bounded below. To understand the boundary condition (3.8) (using w_1 in $\overline{C}_0$), change to the coordinate $\tilde{w}_1 = 1/w_1$ and close the domain to include the "bottom boundary" $\{w_1 = -\infty\}$. Then (3.8) becomes an ordinary Dirichlet condition. While the diffusivity vanishes on this boundary, the drift is nonzero into the boundary. The hitting probability is therefore continuous up to the boundary.

For F^k there is the following more general picture. Consider the PDE in $\overline{C}_0 \cup \cdots \cup \overline{C}_k$, defined across the seams by changing coordinates as in (3.59). Put the boundary condition (3.7) on all the chambers and (3.8) on the bottom of $\overline{C}_k$. Then the solution is F^k in $\overline{C}_0$; the reason is the same as for $F = F^0$, but now using (3.5) and the hitting event "at most k explosions". Similarly, the solution is F^{k-1} in $\overline{C}_1$ and so on down to F^0 in $\overline{C}_k$. Continuity holds across the seams and (3.9) follows after permuting coordinates.

Toward uniqueness, suppose $\tilde{F}$ is another bounded solution of the boundary value problem (3.6)–(3.8) on $\mathbb{R} \times \overline{C}_0$. With the notation of Theorem 3.1.5, $\tilde{F}(rx; \mathbf{p}_x)$ is a local martingale under $\mathbf{P}_{x_0,\mathbf{w}}$ by the PDE (3.6). It is therefore a bounded martingale. Let $\zeta \in (x_0, \infty]$ be the time of the first explosion; optional stopping gives $\tilde{F}(rx_0; \mathbf{w}) = \mathbf{E}_{x_0,\mathbf{w}} \tilde{F}\big(r(\zeta \wedge x); \mathbf{p}_{\zeta \wedge x}\big)$ for all $x \geq x_0$. Taking $x \to \infty$, we conclude by bounded convergence, the boundary behavior (3.7),(3.8) of $\tilde{F}$ and Lemma 3.5.5 (ii) that $\tilde{F}(rx_0, \mathbf{w}) = \mathbf{P}_{x_0,\mathbf{w}}(\zeta = \infty)$. By Theorem 3.1.5, this probability is $F_\beta(rx_0, \mathbf{w})$. One argues similarly for the higher eigenvalues. □

Chapter 4

Going supercritical

Chapters 2 and 3 treat the subcritical and critical regimes of the BBP phase transition. The supercritical regime is very different; here the largest eigenvalues no longer tend to the edge of the limiting empirical spectral distribution, but rather separate and tend to outlying points. Furthermore, rather than displaying Tracy-Widom fluctuations on the order $n^{-2/3}$, they display Gaussian fluctuations on the order $n^{-1/2}$. (When k supercritical population eigenvalues coincide, the k largest eigenvalues are actually governed by the joint eigenvalue law of the GOE/GUE of size k.) As mentioned previously, in the complex Wishart case these results were part of BBP. In the real case the limiting eigenvalue location was confirmed by Baik and Silverstein (2006) and the limiting fluctuations by Paul (2007), Bai and Yao (2008) who used perturbation theory arguments. (The former paper just does $k = 1$, while the latter also applies to more general non-Gaussian sample covariance matrices.)

Despite the existing results, it is interesting to try to complete the picture developed in Chapters 2 and 3 and understand the supercritical regime in the operator limit framework. In any case, a natural question is what happens at the "supercritical end of the critical regime" on the level of the limiting eigenvalue process. This question is answered in the rank one case in Section 4.1, where we prove Gaussian asymptotics of the stochastic Airy ground state energy as the boundary condition tends to supercritical.

In Section 4.2 we use the tridiagonal form of a rank one spiked Wishart matrix to prove that, with supercritical spiking, the largest eigenvalue tends to the correct outlying location. The proof offers a simple heuristic for this location. In Section 4.3 we build on these heuristics to predict the precise limiting fluctuations. One could proceed very similarly for the rank one perturbed Gaussian model but we do not do so here.

There is an underlying idea, a version of which is made precise in Section 4.1. Recall the stochastic Airy operator

$$-\frac{d^2}{dx^2} + \tfrac{2}{\sqrt{\beta}} b'_x + x \tag{4.1}$$

acting on $L^2(\mathbb{R}_+)$, where b'_x is standard Gaussian white noise. The three terms correspond to effects in the tridiagonal models that are balanced on the scale of soft edge fluctuations. On larger scales, the effects separate into three distinct orders: Laplacian $\gg$ noise $\gg$ Airy potential. The Laplacian term therefore governs the principal eigenvector; the eigenvector will have a deterministic limit on the correct scale, namely a decaying geometric sequence on $\mathbb{Z}_+$ (or a decaying exponential function on $\mathbb{R}_+$ for "vanishingly supercritical" spiking). This limiting eigenvector determines the limiting eigenvalue and governs its fluctuations, which will be a deterministic linear functional of the noise and therefore Gaussian. The Airy term only helps to control behaviour at infinity.

In some sense, the separation phenomenon is captured in the following "cartoon version": the operator $-d^2/dx^2$ on $L^2(\mathbb{R}_+)$ with boundary condition $f'(0) = af(0)$ has continuous spectrum $[0, \infty)$, but when $a < 0$ there is also an outlying eigenvalue $-a^2$.

We believe it is feasible to give a complete treatment of the supercritical regime using the ideas of this chapter. Beyond the heuristics just discussed, one could proceed by discrete analogues of the arguments of the Section 4.1, especially the coercivity/uniform convexity argument for Lemma 4.1.3. One motivation to pursue this line is that it would offer a new description of eigenvector concentration in the supercritical regime, an important phenomenon studied by several authors including Paul (2007), Benaych-Georges and Nadakuditi (2009).

4.1 The supercritical end of the critical regime

Fix $\beta > 0$ and consider the stochastic Airy operator (4.1) acting on $L^2(\mathbb{R}_+)$ with boundary condition $f'(0) = wf(0)$ for some $w \in \mathbb{R}$. Recall from Chapter 2 that there is almost surely a well defined ground state f_0 (normalized by $\|f\| = 1$ and $f(0) > 0$) at finite energy Λ_0. Theorem 2.2.9 provides the variational characterization

$$\Lambda_0 = \inf_{f \in L^2,\ \|f\|=1} \int_0^\infty \big(f'(x)^2 + xf(x)^2\big)dx + wf(0)^2 + \tfrac{2}{\sqrt{\beta}} \int_0^\infty f(x)^2\, db_x \tag{4.2}$$

in which the minimum is attained at f_0. (Candidate minimizers should have the first integral finite, and for these f the other two terms are a.s. defined and finite; the stochastic integral is defined pathwise via integration by parts.)

We are concerned here with the limit $w \to -\infty$. In this case we know from Lemma 2.4.1 and monotonicity in w that $\Lambda_0 \to -\infty$ a.s. We now describe the asymptotic behaviour of f_0 and Λ_0. As usual, $\|f\|_{H^1}^2 = \|f\|^2 + \|f'\|^2$.

Theorem 4.1.1. *Almost surely* $|w|^{-1/2} f_0\big(|w|^{-1}\, t\big) \to \sqrt{2}\, e^{-t}$ *in* $H^1(\mathbb{R}_+)$*, and we have the distributional convergence*

$$\frac{\Lambda_0 + w^2}{\sqrt{|w|}} \Rightarrow N\big(0, \tfrac{4}{\beta}\big) \qquad \text{as } w \to -\infty.$$

When the left-hand side is considered jointly with b_x*, they are asymptotically jointly Gaussian and uncorrelated.*

Remark 4.1.2. As for the higher eigenvalues, Lemma 2.4.1 and monotonicity imply that

$$\big(\Lambda_1(w), \Lambda_2(w), \ldots\big) \to \big(\Lambda_0(+\infty), \Lambda_1(+\infty), \ldots\big) \quad \text{a.s.} \qquad \text{as } w \to -\infty,$$

where $w = +\infty$ indicates the Dirichlet (unperturbed) spectrum. The last assertion of the theorem implies that the Gaussian limit of Λ_0 is independent of this limit when they are taken jointly.

Proof. By way of motivation, the "first order behaviour" of the ground state should already be described by the Laplacian and boundary terms. As soon as $w < 0$, the minimizer in

$$\inf_{\|f\|=1} \|f'\|^2 + w f(0)^2$$

is $\sqrt{2|w|}e^{wx}$ with corresponding minimum $-w^2$; this suggests centering by w^2 and reparametrizing time as $t = |w|\, x$. Furthermore, the order of the fluctuations should be predicted by the stochastic integral, which for the latter function has variance $|w|$; this suggests scaling by $1/\sqrt{|w|}$.

Writing

$$E_w = \frac{\Lambda_0 + w^2}{\sqrt{|w|}},$$
$$\psi(t) = |w|^{-1/2}\, f(|w|^{-1}\, t), \qquad B_t = |w|^{1/2}\, b_{|w|^{-1}t}$$

(note that $\|\psi\| = \|f\|$ and B is again a standard Brownian motion), we are thus led to consider

$$E_w = \inf_{\|\psi\|=1} I_w(\psi), \tag{4.3}$$

where the objective quadratic functional is

$$I_w(\psi) \;=\; |w|^{3/2}\, J(\psi) + \tfrac{2}{\sqrt{\beta}} \int_0^\infty \psi^2\, dB + |w|^{-3/2} \int_0^\infty \psi^2\, t\, dt \tag{4.4}$$

in which we put

$$J(\psi) \;=\; \|\psi'\|^2 - \psi(0)^2 + \|\psi\|^2 \;=\; \big\|\psi' + \psi\big\|^2.$$

The minimum is attained at

$$\psi_w(t) = |w|^{-1/2}\, f_0(|w|^{-1}\, t).$$

The separation of scales mentioned in the introduction can already be seen in (4.4).

Now on the one hand,

$$E_w = I_w(\psi_w) \geq \tfrac{2}{\sqrt{\beta}} \int \psi_w^2\, dB. \tag{4.5}$$

On the other hand, $J(\psi) = 0$ for ψ a multiple of e^{-t}; writing

$$\psi_*(t) = \sqrt{2}\, e^{-t},$$

certainly

$$E_w \leq I_w(\psi_*) = \tfrac{2}{\sqrt{\beta}} \int \psi_*^2\, dB + \tfrac{1}{2}\, |w|^{-3/2}\,. \tag{4.6}$$

It is convenient to recouple the models over w so that now paths of B are fixed as w varies. The proof is completed by Proposition 4.1.5 below, where we show that $\psi_w \to \psi_*$ as $w \to -\infty$ in a sense sufficient to guarantee pathwise convergence of the stochastic integral in (4.5) to the one in (4.6). Then

$$E_w \to \tfrac{2}{\sqrt{\beta}} \int \psi_*^2\, dB \quad \text{a.s.} \qquad \text{as } w \to -\infty, \tag{4.7}$$

which implies the distributional convergence in the theorem.

As for the joint behaviour with b_x, it is enough to show that for $s \in \mathbb{R}$ and f a compactly supported step function on $\mathbb{R}_+$ we have

$$\mathbf{E}\, e^{i\left(sE_w + \int f\, db\right)} \to e^{-(2/\beta)s^2 - (1/2)\|f\|^2} \qquad \text{as } w \to -\infty.$$

Now

$$\mathbf{E} \left| e^{i\left(sE_w + \int f\, db\right)} - e^{i\left(s(2/\sqrt{\beta}) \int \psi_* dB + \int f\, db\right)} \right| = \mathbf{E} \left| e^{isE_w} - e^{is(2/\sqrt{\beta}) \int \psi_* dB} \right| \to 0$$

by (4.7) and the bounded convergence theorem. But

$$\begin{aligned}\mathbf{E}\, e^{i\left(s(2/\sqrt{\beta})\int \psi_* dB + \int f\, db\right)} &= \mathbf{E}\, e^{i\int_0^\infty \left(s(2/\sqrt{\beta})\psi_*(t) + |w|^{-1/2} f(|w|^{-1}t)\right) dB_t} \\ &= e^{-\frac{1}{2}\int_0^\infty \left(s(2/\sqrt{\beta})\psi_*(t) + |w|^{-1/2} f(|w|^{-1}t)\right)^2 dt}\end{aligned}$$

which has the desired limit because the cross term $\int_0^\infty \psi_*(t)\ |w|^{-1/2} f(|w|^{-1}\, t)\, dt \to 0$. (The basic fact that L^2 functions on distinct scales are asymptotically orthogonal may be proved by breaking the integral into two pieces on an intermediate scale.) □

Lemma 4.1.3. *If $\psi \in H^1(\mathbb{R}_+)$ with $\|\psi\|_{L^2} = 1$ and $\langle \psi, \psi_* \rangle \geq 0$, then*

$$\|\psi - \psi_*\|_{H^1}^2 \ \leq\ 8\, J(\psi).$$

Proof. We first claim that for $\varphi \in H^1$ with $\langle \varphi, e^{-t} \rangle = 0$ we have

$$J(\varphi) \ \geq\ \|\varphi\|^2 + \tfrac{1}{2}\, \|\varphi'\|^2 .$$

Indeed, the orthogonality relation can be integrated by parts to give

$$\varphi(0) = \left\langle \varphi', e^{-t} \right\rangle ;$$

an application of Cauchy-Schwarz then yields

$$\tfrac{1}{2}\, \|\varphi'\|^2 \geq \varphi(0)^2,$$

which is equivalent to the claim.

Now put $\varphi = \psi - \langle \psi, \psi_* \rangle\, \psi_*$. By the previous estimate and the hypotheses on ψ,

$$\|\psi - \psi_*\|^2 \ =\ 2\big(1 - \langle \psi, \psi_* \rangle\big) \ \leq\ 2\big(1 - \langle \psi, \psi_* \rangle^2\big) \ =\ 2\, \|\varphi\|^2 \ \leq\ 2J(\varphi) \ =\ 2J(\psi).$$

In other words,

$$\|\psi - \psi_*\| \ \leq\ \sqrt{2} J(\psi)^{1/2}.$$

But

$$J(\psi)^{1/2} \ =\ J(\psi - \psi_*)^{1/2} \ =\ \|(\psi' - \psi_*') + (\psi - \psi_*)\| \ \geq\ \|\psi' - \psi_*'\| - \|\psi - \psi_*\|\,,$$

so also

$$\|\psi' - \psi_*'\| \ \leq\ \big(1 + \sqrt{2}\big) J(\psi)^{1/2}.$$

Square and add to conclude. □

Note that ψ_w satisfies the hypotheses of this lemma by facts about f_0 from Chapter 2. In particular, $\psi_w > 0$ by the oscillation theory (or by Perron-Frobenius theory), so certainly $\langle \psi_w, \psi_* \rangle \geq 0$.

Lemma 4.1.4. *There is a random, almost surely finite constant C_0 such that*

$$\left| \tfrac{2}{\sqrt{\beta}} \int \psi^2 \, dB \right| \leq C_0 \left(\|\psi'\|^2 + \log\left(\tfrac{1}{\varepsilon}\right) \|\psi\|^2 + \varepsilon \|\psi\sqrt{t}\|^2 \right) \tag{4.8}$$

for all ψ and all $0 < \varepsilon \leq \frac{1}{2}$.

Proof. Decompose B as in RRV (Lemma 2.3 and Proposition 2.4 there); in the last step, use $\log(2+t) \leq \log\left(\frac{1}{\varepsilon}\right) + \varepsilon t$. (The function $\varepsilon t + \log\left(\frac{1}{\varepsilon}\right) - \log(2+t)$ is minimized at $t = \frac{1}{\varepsilon} - 2$ with minimum $1 - 2\varepsilon$, so the inequality holds for $\varepsilon \leq \frac{1}{2}$.) □

It is standard (see e.g. Halmos 1957 §18, Theorem 3) that the above quadratic form bound may be polarized to a bilinear form bound. Denoting the quantity in parentheses on the right hand side of (4.8) by $\|\psi\|^2_{L^*_\varepsilon}$, we have

$$\left| \int \varphi\psi \, dB \right| \leq C_0 \|\varphi\|_{L^*_\varepsilon} \|\psi\|_{L^*_\varepsilon}. \tag{4.9}$$

Proposition 4.1.5. *Almost surely, $\psi_w \to \psi_*$ in $H_1(\mathbb{R}_+)$ and*

$$\int \psi_w^2 \, dB \to \int \psi_*^2 \, dB.$$

Proof. Throughout, C denotes a random, almost surely finite constant that is allowed to change from line to line.

First, using $2\,|\langle \psi', \psi \rangle| \leq (1-\delta)\,\|\psi'\|^2 + \frac{1}{1-\delta}\,\|\psi\|^2$ gives

$$J(\psi) \geq \delta\,\|\psi'\|^2 + (1 - \tfrac{1}{1-\delta})\,\|\psi\|^2 \geq \delta\,\|\psi'\|^2 - 2\delta\,\|\psi\|^2$$

for $0 \leq \delta \leq \frac{1}{2}$. Then by (4.6) together with Lemma 4.1.4 (any ε), one has

$$\begin{aligned} C &\geq I_w(\psi_*) \\ &\geq I_w(\psi_w) \\ &\geq |w|^{3/2} \left(\tfrac{1}{2} J(\psi_w) + (\delta/2)\,\|\psi_w'\|^2 - \delta \right) + \tfrac{2}{\sqrt{\beta}} \int \psi_w^2 \, dB + |w|^{-3/2} \left\|\psi_w \sqrt{t}\right\|^2. \end{aligned}$$

Applying Lemma 4.1.4 now with $\varepsilon = |w|^{-3/2}/2C_0$, and taking $\delta = 2\,|w|^{-3/2}\,(C_0 + 1)$ above, we find

$$|w|^{3/2}\,J(\psi_w) + \|\psi_w'\|^2 + |w|^{-3/2} \left\|\psi_w \sqrt{t}\right\|^2 \leq C \log|w| \tag{4.10}$$

for w sufficiently negative. In particular

$$\left\|\psi_w - \psi_*\right\|_{H^1}^2 \leq 8\, J(\psi_w) \leq C\,|w|^{-3/2} \log|w| \tag{4.11}$$

by Lemma 4.1.3, establishing the first assertion.

We conclude by using (4.9) with $\varepsilon = |w|^{-3}$ as follows:

$$\begin{aligned}
\left|\int (\psi_w^2 - \psi_*^2)\, dB\right| &\leq C_0 \left\|\psi_w + \psi_*\right\|_{L_\varepsilon^*} \left\|\psi_w - \psi_*\right\|_{L_\varepsilon^*} \\
&\leq C\left(\left\|\psi_w'\right\|^2 + \left\|\psi_*'\right\|^2 + \log|w| + |w|^{-3}\left(\left\|\psi_w\sqrt{t}\right\|^2 + \left\|\psi_*\sqrt{t}\right\|^2\right)\right)^{1/2} \\
&\times\left(\left\|\psi_w' - \psi_*'\right\|^2 + \log|w|\left\|\psi_w - \psi_*\right\|^2 + |w|^{-3}\left(\left\|\psi_w\sqrt{t}\right\|^2 + \left\|\psi_*\sqrt{t}\right\|^2\right)\right)^{1/2} \\
&\leq C\,|w|^{-3/4}\log^{3/2}|w|\,,
\end{aligned}$$

where the last inequality is a consequence of (4.10) and (4.11). □

4.2 The limiting location

Here we focus on the rank one spiked Wishart ensemble and its general β analogue as defined in Chapter 2. We find the limiting location of the largest eigenvalue using a simple heuristic together with a little matrix perturbation theory.

Our heuristic is the following observation: A spiked unscaled discrete Laplacian has a certain geometric sequence as its principal eigenvector, which is decaying in the supercritical case. Specifically, for $r > 0$ the matrix

$$\begin{bmatrix}
r-2 & 1 & & & & & \\
1 & -2 & 1 & & & & \\
 & 1 & -2 & \ddots & & & \\
 & & \ddots & \ddots & 1 & & \\
 & & & 1 & -2 & 1 & \\
 & & & & 1 & r^{-1}-2
\end{bmatrix}$$

has an eigenvector $v_i = r^{-i}$ with corresponding eigenvalue $\lambda = (r-1)^2/r$. We know λ is the top eigenvalue because v has no sign-changes. (The observation also has an infinite Jacobi matrix version.) The supercritical case is $r > 1$. In this case we argue that it is good enough for the matrix to look roughly like this in the top-left corner, which gets us the limiting location. In the next section we use first-order perturbation theory to give a heuristic computation of the limiting fluctuations.

Recall the tridiagonal form $S_{n,p} = W_{n,p}^\dagger W_{n,p}$ with $W_{n,p} = W_{n,p}^{\beta,\ell_{n,p}}$ as in (2.2). We recall the following approximations from the heuristics in Section 2.3. The diagonal and off-diagonal processes of $\beta S_{n,p}$ are

$$\ell_{n,p}\widetilde{\chi}^2_{\beta n} + \chi^2_{\beta(p-1)}, \quad \widetilde{\chi}^2_{\beta(n-1)} + \chi^2_{\beta(p-2)}, \quad \widetilde{\chi}^2_{\beta(n-2)} + \chi^2_{\beta(p-3)}, \quad \cdots$$
$$\widetilde{\chi}_{\beta(n-1)}\chi_{\beta(p-1)}, \quad \widetilde{\chi}_{\beta(n-2)}\chi_{\beta(p-2)}, \quad \cdots$$

respectively. The approximations

$$\chi_k \approx \sqrt{k} + \sqrt{1/2}\, g, \qquad \chi^2_k \approx k + \sqrt{2k}\, g,$$

are valid for k large, where g is a suitably coupled standard Gaussian. To leading order, the top-left corner of S has $n+p$ on the diagonal and $\sqrt{np}$ on the off-diagonal (ignoring the spike). So the top-left corner of

$$\frac{1}{\sqrt{np}}\Big(S - \big(\sqrt{n} + \sqrt{p}\big)^2 I\Big)$$

is approximately an unscaled discrete Laplacian.

Theorem 4.2.1. *Assuming the notation of Theorem 2.1.1, suppose that*

$$\sqrt{n/p}\,(\ell_{n,p} - 1) \to r > 1 \qquad \text{as } n \wedge p \to \infty.$$

Then in probability (in fact almost surely),

$$\frac{1}{\sqrt{np}}\Big(\lambda_1 - \big(\sqrt{n} + \sqrt{p}\big)^2\Big) \to \frac{(r-1)^2}{r} \qquad \text{as } n \wedge p \to \infty.$$

Remark 4.2.2. Of course for all other λ_k, $k \geq 2$ the corresponding limit is zero in this scaling, by Theorem 2.1.1 and Weyl interlacing.

To recover a statement in the form familiar from Baik and Silverstein (2006), assume that $p/n \to \gamma^2 \in (0, \infty)$ and let $\ell_{n,p} = \ell = 1 + \gamma r$. Our theorem implies the following: If $r > 1$, then a.s.

$$\frac{1}{n}\lambda_1 \to (1+\gamma)^2 + \gamma\frac{(r-1)^2}{r} = \ell + \gamma^2\frac{\ell}{\ell - 1}.$$

Recall that $(1+\gamma)^2$ is the right endpoint of the support of Marcenko-Pastur; the term $\gamma(r-1)^2/r$ therefore quantifies the separation. Our formulation treats n and p symmetrically (apart from the inherent asymmetry of the spike) and has the advantage of allowing $n, p \to \infty$ together arbitrarily.

Toward a proof, we recall a standard fact.

Lemma 4.2.3. *Let A be a symmetric $n \times n$ matrix and endow $\mathbb{R}^n$ with the standard inner product and norm. Suppose that $\tilde{\lambda} \in \mathbb{R}$ and $\tilde{v} \in \mathbb{R}^n$ satisfy $\|\tilde{v}\| = 1$ and $\|(A - \tilde{\lambda})\tilde{v}\| < \varepsilon$ for some $\varepsilon > 0$. Then there is an eigenvalue λ of A with $|\lambda - \tilde{\lambda}| < \varepsilon$.*

Proof. Let $v_1, \dots, v_n$ be an orthonormal basis of eigenvalues with corresponding eigenvectors $\lambda_1, \dots, \lambda_n$. Write $\tilde{v} = \sum_i a_i v_i$ and note that $\sum_i |a_i|^2 = 1$. Then

$$\varepsilon^2 > \|(A - \tilde{\lambda})\tilde{v}\|^2 = \|\textstyle\sum_i a_i(\lambda_i - \tilde{\lambda})v_i\|^2 = \sum_i |a_i|^2 |\lambda_i - \tilde{\lambda}|^2,$$

which implies that $|\lambda_i - \tilde{\lambda}|^2 < \varepsilon^2$ for some i. □

Proof of Theorem 4.2.1. Put

$$A_{n,p} = \frac{1}{\sqrt{np}} \left(S_{n,p} - (\sqrt{n} + \sqrt{p})^2\right) \qquad \text{and} \qquad r_{n,p} = \sqrt{n/p}\,(\ell_{n,p} - 1)\,.$$

By the remark and the lemma it suffices to show that $\|(A_{n,p} - \tilde{\lambda}_{n,p})\tilde{v}_{n,p}\| \to 0$ a.s., where $\tilde{\lambda}_{n,p} = (r_{n,p} - 1)^2 / r_{n,p}$ and $\tilde{v}_{n,p(i)} = r_{n,p}^{-i}$.

In the following, C denotes a constant that is independent of n, p but may change from line to line. Let $r_0 > 1$ be such that $r_{n,p} \geq r_0$ for $n \wedge p$ sufficiently large. Using standard facts about moments of chi random variables, independence, and that $\ell_{n,p} \leq C\sqrt{p/n}$, we estimate

$$\begin{aligned}
\mathbf{E}\,\|(A - \tilde{\lambda})\tilde{v}\|^2 &\leq \frac{C}{np}\,\mathbf{E} \sum_{i=0}^{n \wedge p} r_0^{-2i} \Bigg(\left(\frac{p}{n} + 1\right) \left(\tfrac{1}{\beta}\widetilde{\chi}^2_{\beta(n-i)} - n\right)^2 \\
&\qquad + \left(\tfrac{1}{\beta}\chi^2_{\beta(p-i)} - p\right)^2 + \left(\tfrac{1}{\beta}\widetilde{\chi}_{\beta(n-i)}\chi_{\beta(p-i)} - \sqrt{np}\right)^2 \Bigg) \\
&\leq \frac{C}{np} \sum_{i=0}^{n \wedge p} r_0^{-2i} \big(n + p + i(n+p) + i^2(n+p)\big) \\
&\leq C \left(\frac{1}{n} + \frac{1}{p}\right) \sum_{i=0}^{\infty} i^2 r_0^{-2i} \\
&\leq \frac{C}{n \wedge p}.
\end{aligned}$$

(We omitted an additional error of $r_0^{-(n \wedge p)}$ from the last term.) While this is only good enough for convergence in probability, similar calculations with higher moments give the fourth moment bound $\mathbf{E}\|(A - \tilde{\lambda})\tilde{v}\|^4 \leq C/(n \wedge p)^2$. □

4.3 Heuristic for the limiting fluctuations

In the previous section we also predicted the limiting eigenvector. One can therefore predict the first-order eigenvalue correction, i.e. the fluctuations. First-order perturbation theory states that if λ, v is an eigenvalue/eigenvector pair of A (with v normalized) and P is small, then to first order $A + P$ has an eigenvalue $\lambda + v^\dagger P v$.

We take up the approximations of the previous section. To first order, we found that $\left(S_{n,p} - (\sqrt{n} + \sqrt{p})^2\right) / \sqrt{np}$ is a "spiked discrete Laplacian". The next order terms are, omitting a $\sqrt{2/\beta}$ prefactor and going down the diagonal and off-diagonal respectively:

$$\begin{gathered}\tfrac{1}{\sqrt{p}}\ell\tilde{g}_n + \tfrac{1}{\sqrt{n}}g_{p-1}, \quad \tfrac{1}{\sqrt{p}}\tilde{g}_{n-1} + \tfrac{1}{\sqrt{n}}g_{p-2}, \quad \tfrac{1}{\sqrt{p}}\tilde{g}_{n-2} + \tfrac{1}{\sqrt{n}}g_{p-3}, \quad \cdots \\ \tfrac{1}{2\sqrt{n}}\tilde{g}_{n-1} + \tfrac{1}{2\sqrt{p}}g_{p-1}, \quad \tfrac{1}{2\sqrt{n}}\tilde{g}_{n-2} + \tfrac{1}{2\sqrt{p}}g_{p-2}, \quad \cdots\end{gathered}$$

where we have labeled the g's to match the corresponding χ's, the important thing being that distinctly labeled g's are independent. Again with $r_{n,p} = \sqrt{n/p}\,(\ell_{n,p} - 1)$, the limiting eigenvector is

$$\sqrt{1 - r^{-2}}\,\begin{bmatrix} 1 & r^{-1} & r^{-2} & \dots \end{bmatrix}^\dagger.$$

First order perturbation theory then predicts that $\left(\lambda_1 - (\sqrt{n} + \sqrt{p})^2\right) / \sqrt{np}$ has Gaussian fluctuations with variance

$$\begin{aligned}&\tfrac{2}{\beta}(1 - r^{-2})^2 \left(\tfrac{1}{p}\ell^2 + (\tfrac{1}{\sqrt{n}} + \tfrac{1}{\sqrt{p}}r^{-1})^2 + (\tfrac{1}{\sqrt{n}}r^{-1} + \tfrac{1}{\sqrt{p}}r^{-2})^2 + \cdots\right) \\ &= \tfrac{2}{\beta}(1 - r^{-2}) \left(\tfrac{1}{p}\ell^2(1 - r^{-2}) + (\tfrac{1}{\sqrt{n}} + \tfrac{1}{\sqrt{p}}r^{-1})^2\right) \\ &= \tfrac{2}{\beta}(1 - r^{-2}) \left(\tfrac{1}{p}\ell^2\big(1 - \tfrac{p}{n(\ell-1)^2}\big) + \tfrac{1}{n}\big(\tfrac{\ell}{\ell-1}\big)^2\right) \\ &= \frac{2}{\beta}\left(\frac{\ell^2}{p} - \frac{\ell^2}{n(\ell-1)^2}\right).\end{aligned}$$

This prediction agrees with results of BBP, Paul (2007) for $\beta = 1, 2$ respectively. Note once again that we state the scaling slightly differently, dividing λ_1 by $\sqrt{np}$ rather than n; our formulation has the advantage of symmetry in n, p and allowing $n, p \to \infty$ arbitrarily.

Chapter 5

Connection with Painlevé II

In this chapter we connect our PDE characterization of the deformed Tracy-Widom(β) laws, Theorems 2.1.7(ii), 2.4.3 and 3.1.6, with known Painlevé II representations for these laws at $\beta = 2$ due to Baik (2006) (derived of course from the Fredholm determinant representations of BBP). In the rank one case we obtain a rigorous alternative proof of these representations. Furthermore, using results of Baik and Rains (2000, 2001) we guess and prove a similar formula at $\beta = 4$. In particular, we recover a novel proof of the Painlevé II representations of the Tracy-Widom laws for $\beta = 2, 4$.

At $\beta = 2$ we indicate how the connection extends to the second largest and subsequent eigenvalue laws. Put another way, the connection between the PDE and Painlevé II includes not only the Hastings-McLeod solution but also the Ablowitz-Segur solutions. For the multi-spiked deformations, we describe a computer-assisted symbolic computation providing strong evidence that the connection holds here as well.

A number of points remain somewhat mysterious. Most obviously, we lack a connection in the $\beta = 1$ case; while the literature previously did not even suggest a guess, it would now be satisfying to reconcile (2.9), (2.10) with the formula obtained by Mo (2011). We point out that in the work of Dieng (2005), the $\beta = 1$ case also involves additional complications; perhaps they are related to ours, though this is idle speculation.

Even at $\beta = 2, 4$ it seems there should be a more direct way to derive or at least understand the connection. From the point of view of the PDE, some kind of extra structure appears to be present at certain special values of the parameter β; what about other values? From the point of view of nonlinear special functions, we show directly—independently of any limit theorems—how certain well-studied solutions of Painlevé II admit a new characterization in terms of a linear parabolic boundary value problem in

the plane.

5.1 Tracy-Widom laws and rank one deformations

We introduce the functions involved in the Painlevé representations. The **Hastings-McLeod solution** $u(x)$ of the **homogeneous Painlevé II equation**

$$u'' = 2u^3 + xu, \tag{5.1}$$

is characterized by

$$u(x) \sim \mathrm{Ai}(x) \quad \text{as } x \to +\infty \tag{5.2}$$

where $\mathrm{Ai}(x)$ is the Airy function (characterized in turn by $\mathrm{Ai}'' = x\,\mathrm{Ai}$ and $\mathrm{Ai}(+\infty) = 0$). It is known that there is a unique such function and that it has no singularities on $\mathbb{R}$ (Hastings and McLeod 1980). Put

$$v(x) = \textstyle\int_x^\infty u^2, \tag{5.3}$$

$$E(x) = \exp\big(-\textstyle\int_x^\infty u\big), \qquad F(x) = \exp\big(-\textstyle\int_x^\infty v\big). \tag{5.4}$$

Next define two functions $f(x,w)$, $g(x,w)$ on $\mathbb{R}^2$, analytic in w for each fixed x, by the first order linear ODEs

$$\frac{\partial}{\partial w}\begin{pmatrix} f \\ g \end{pmatrix} = \begin{pmatrix} u^2 & -wu - u' \\ -wu + u' & w^2 - x - u^2 \end{pmatrix}\begin{pmatrix} f \\ g \end{pmatrix} \tag{5.5}$$

and the initial conditions

$$f(x,0) \;=\; E(x) \;=\; g(x,0). \tag{5.6}$$

Equation (5.5) is one member of the Lax pair for the Painlevé II equation. The functions f, g can also be defined in terms of the solution of the associated Riemann-Hilbert problem; analysis of the latter yields some information about u, f, g summarized in Facts 5.1.5 and 5.1.6 below. The following theorem expresses the relationship between the objects just defined and the general β characterization at $\beta = 2, 4$.

Theorem 5.1.1. *The identities*

$$F_{2,w}(x) = f(x,w)F(x), \tag{5.7}$$

$$F_{4,w}(x) = \left(\frac{(f+g)E^{-1/2} + (f-g)E^{1/2}}{2}\right) F^{1/2}\Big|_{(2^{2/3}x,\, 2^{1/3}w)} \tag{5.8}$$

hold and follow directly from Theorem 2.1.7 and Facts 5.1.5 and 5.1.6.

The formula for $F_{2,w}$ is given by Baik (2006), although it appeared earlier in work of Baik and Rains (2000, 2001) in a very different context. The formula for $F_{4,w}$ appears in Baik and Rains (2000, 2001) in a disguised form; the $w = 0$ case is obtained by Wang (2008), but it is a new result in this context for $w \neq 0, \infty$.

In particular, we recover the Painlevé II representations of Tracy and Widom at these β in a novel and simple way.

Corollary 5.1.2 (Tracy and Widom 1994, 1996, BBP 2005, Wang 2008)**.** *We have*

$$F_{2,\infty}(x) = F(x), \tag{5.9}$$

$$F_{4,\infty}(2^{-2/3}x) = \tfrac{1}{2}\big(E^{1/2}(x) + E^{-1/2}(x)\big)F^{1/2}(x), \tag{5.10}$$

$$F_{2,0}^{1/2}(x) = F_{4,0}(2^{-2/3}x) = E^{1/2}(x)F^{1/2}(x). \tag{5.11}$$

Remark 5.1.3. The latter distribution is known to be $F_{1,\infty}(x)$ (Tracy and Widom 1996). Unfortunately we lack an independent proof.

Remark 5.1.4. With $\beta = 4$ one often encounters an old convention in which the scaling of the independent variable is "off" by a factor $2^{1/6}$.

We now prove Theorem 5.1.1 and Corollary 5.1.2. We will need some standard facts about the function $u(x)$ defined by (5.1, 5.2) and the derived functions $v(x),\ E(x),\ F(x)$ defined in (5.3, 5.4).

Fact 5.1.5. *The following hold:*

(i) $u > 0$ *on* $\mathbb{R}$ *and* $u'/u \sim -\sqrt{x}$ *as* $x \to +\infty$.

(ii) E *and* F *are distribution functions.*

(iii) $E(x) = O(e^{-cx^{3/2}})$ *for some* $c > 0$ *as* $x \to +\infty$.

We will also take for granted some additional information about the functions $f(x, w)$, $g(x, w)$ defined by (5.5, 5.6).

Fact 5.1.6. *The following hold.*

(i) For each $x \in \mathbb{R}$,

$$\lim_{w\to+\infty} \begin{pmatrix} f \\ g \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{5.12}$$

$$\lim_{w\to-\infty} \begin{pmatrix} f \\ g \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{5.13}$$

(ii) For each $w \in \mathbb{R}$,

$$\frac{\partial}{\partial x}\begin{pmatrix} f \\ g \end{pmatrix} = \begin{pmatrix} 0 & u(x) \\ u(x) & -w \end{pmatrix}\begin{pmatrix} f \\ g \end{pmatrix}. \tag{5.14}$$

(iii) There is the identity

$$g(x,w) = f(x,-w)e^{\frac{1}{3}w^3 - xw}. \tag{5.15}$$

(iv) For fixed $w \in \mathbb{R}$,

$$f(x,w) \to 1 \quad \textit{as } x \to +\infty; \tag{5.16}$$

$$f(x,w) > 0 \quad \textit{for } x \textit{ sufficiently negative.} \tag{5.17}$$

These properties follow from an analysis of the associated Riemann-Hilbert problem with the special monodromy data corresponding to the Hastings-McLeod solution (see Fokas, Its, Kapaev and Novokshenov 2006). They are proved in Baik and Rains (2001) except for (iv) which goes back to Deift and Zhou (1995). Interestingly (5.6) and (5.12) are interchangeable in that the latter also uniquely determines a solution of (5.5); this fact does not depend on the specific solution of (5.1) specified by (5.2). By contrast, (5.13) does depend on (5.2). Equations (5.5, 5.14) constitute a so-called *Lax pair* for the Painlevé II equation (5.1). (It is in fact a simple transformation of the standard Flaschka-Newell Lax pair.) The consistency condition of this overdetermined system—i.e. that the partials commute—is (5.1).

Proof of Theorem 5.1.1, $\beta = 2$ case. Let $\tilde{F}_2(x,w)$ denote the right-hand side of (5.7). Using (5.4), (5.5) and (5.14), we check that that $\tilde{F}_2$ solves the PDE (2.9) with $\beta = 2$: compute

$$\begin{aligned}
\frac{\partial \tilde{F}_2}{\partial x} &= \Big\{ vf + ug \Big\} F \\
\frac{\partial \tilde{F}_2}{\partial w} &= \Big\{ u^2 f + \big(-wu - u'\big) g \Big\} F \\
\frac{\partial^2 \tilde{F}_2}{\partial w^2} &= \Big\{ \big(u^4 + w^2u^2 - (u')^2\big) f + \big(-u + (wu + u')(x - w^2)\big) g \Big\} F
\end{aligned}$$

and substitute. The coefficient of g vanishes and the coefficient of f is

$$v + u^4 - (u')^2 + xu^2.$$

Differentiating, we see that this quantity is constant by (5.1). As all terms vanish in the limit as $x \to \infty$, the constant is zero.

We must check that $\tilde{F}_2$ is bounded and that it has the boundary behaviour (2.10). To this end we claim $f, g > 0$ on $\mathbb{R}^2$. Fixing w, (5.17, 5.15) cover x sufficiently negative. Now (5.14) shows f increases at least until $x_0 = \min\{x : g(x, w) = 0\}$. But if x_0 exists then (5.14) shows $\frac{\partial g}{\partial x}(x_0) > 0$, a contradiction. This proves the claim. It now follows from (5.14) that $\frac{\partial f}{\partial x} > 0$. From (5.16) we deduce that $f \leq 1$; in particular f is bounded, and hence so is $\tilde{F}_2$. Furthermore, for a given $x \in \mathbb{R}$ and $\varepsilon > 0$, (5.12) yields w_+ such that $f > 1 - \varepsilon$ on $[x, \infty) \times [w_+, \infty)$, and (5.13) yields w_- such that $f < \varepsilon$ on $(-\infty, x] \times (-\infty, w_-]$. Using that $F(x)$ is a distribution function, (2.10) follows. □

Proof of Theorem 5.1.1, $\beta = 4$ case. That the right-hand side $\tilde{F}_4$ of (5.8) satisfies the PDE (2.9) with $\beta = 4$ may be verified just as in the $\beta = 2$ case; the computation is more tedious but the result is very similar and the final step is the same.

It is a little more work to get boundedness and the boundary behaviour (2.10) this time. Dropping the scale factors on x, w, consider

$$G = F^{-1/2}\tilde{F}_4 = \tfrac{1}{2}\big(E^{-1/2} + E^{1/2}\big)f + \tfrac{1}{2}\big(E^{-1/2} - E^{1/2}\big)g.$$

Clearly $G > 0$. For fixed w, $G \to 1$ as $x \to \infty$ by (5.16) and the fact that $E^{-1/2} - E^{1/2} = O(e^{-cx^{3/2}})$ while $g = O(e^{wx})$ from (5.15). Now by (5.14) we have

$$\frac{\partial G}{\partial x} = \tfrac{1}{2}\big(E^{-1/2} + E^{1/2}\big)\big(\tfrac{1}{2}ug\big) + \tfrac{1}{2}\big(E^{-1/2} - E^{1/2}\big)\big(\tfrac{1}{2}uf - wg\big),$$

which is positive for $w \leq 0$. Boundedness in the lower half-plane $\{w \leq 0\}$ follows, as does the lower boundary behaviour using (5.13).

From (5.15) we immediately see $g \leq 1$ on $\{x \geq 0,\, 0 \leq w \leq \sqrt{3x}\}$. By Lemma 5.1.2, $\frac{\partial}{\partial w}F_{\beta,w}(x) \geq 0$. The $\beta = 2$ case of the present theorem then implies that $\frac{\partial f}{\partial w} \geq 0$. From (5.5) we conclude $g \leq u/(w + u'/u)$ provided the denominator is positive. But $u'/u \sim -\sqrt{x}$ as $x \to +\infty$, so there is x_1 such that $u'/u \geq -\sqrt{2x}$ for $x \geq x_1$. The latter bound for g therefore implies that g is bounded on $\{x \geq x_1,\, w \geq \sqrt{3x}\}$. Moreover, for any $x_0 < x_1$ we have that u and u'/u are bounded on the interval $x_0 \leq x \leq x_1$, so g is bounded uniformly over these x for all w sufficiently large. Putting these bounds together we conclude g is bounded on all right half-planes $\{x \geq x_0\}$, and the same then follows for $\tilde{F}_4$.

The upper boundary behaviour follows as well. Indeed, as $x, w \to \infty$ together the coefficient of g vanishes while the coefficient of f tends to 1; the g-term then vanishes while the f-term tends to 1 as in the $\beta = 2$ case.

It remains to show $\tilde{F}_4$ is bounded on the whole plane; it suffices to bound $\tilde{F}_4$ on the upper-left quadrant $Q = \{x \leq 0,\, w \geq 0\}$. Here we can use the fact that $\tilde{F}_4$ solves the PDE. With notation as in Theorem 2.1.7 we have that $\tilde{F}_4(x, p_x)$ is a local martingale under $\mathbf{P}_{(x_0,w_0)}$. By boundedness on right half-planes, it is in fact a bounded martingale. Using that paths explode only to $-\infty$, optional stopping gives the representation $\tilde{F}_4(x_0, w_0) = \mathbf{E}_{(x_0,w_0)}\, \tilde{F}_4(T, p_T)$ where $T = \inf\{x : (x, p_x) \notin Q\}$. The bound thus extends to Q. □

Proof of Corollary 5.1.2. These identities are straightforward consequences of the theorem, (5.6) and (5.12). □

5.2 Subsequent eigenvalue laws

In the previous section we demonstrated an explicit connection between our boundary value problem for the top eigenvalue law at $\beta = 2$ and the Hastings-McLeod solution of Painlevé II. In this section, based on conversations with Alexander Its, we outline how the connection extends to one between the boundary value problem for the subsequent eigenvalue laws and the Ablowitz-Segur solutions.

Taking up the notation of Theorem 2.4.3, let $F_{(k)}(x; w)$ and $F_{(k)}(x) = F_{(k)}(x; +\infty)$ be the limiting distributions for the w-deformed and un-deformed $(k+1)$st largest eigenvalue with $\beta = 2$. Suppose $0 \leq \lambda \leq 1$. Tracy and Widom (1994) proved that

$$\sum_{k=0}^{\infty} (1-\lambda)^k \lambda\, F_{(k)}(x) = \exp\left(-\int_x^{\infty} (s-x)\, u^2(s,\lambda)\, ds \right) \tag{5.18}$$

where $u(\cdot, \lambda)$ solves (5.1) subject to

$$u(x,\lambda) \sim \sqrt{\lambda}\, \mathrm{Ai}(x) \qquad \text{as } x \to \infty. \tag{5.19}$$

Ablowitz and Segur (1977) show that the latter function is unique with no singularities on $\mathbb{R}$ for $0 \leq \lambda < 1$; Hastings and McLeod (1980) showed this statement continues to hold for $\lambda = 1$ (but fails for $\lambda > 1$). The subsequent eigenvalue laws are therefore expressed, via the generating function on the left side of (5.18), in terms of the **Ablowitz-Segur solutions**. Of course one could differentiate repeatedly at $\lambda = 1$ and solve for $F_{(1)}, F_{(2)}, \ldots$ to obtain the formulas in Tracy and Widom (1994).

Remark 5.2.1. Tracy and Widom (1994) actually showed the two sides of (5.18) are both equal to a certain Fredholm determinant. We briefly explain the situation by

introducing the determinantal representation at $\beta = 2$; while not neeeded here, it is somewhat illuminating. The $\beta = 2$ soft edge limit is a determinantal point process with the so-called **Airy kernel** given by

$$\mathbf{A}(x, y) = \frac{\mathrm{Ai}(x)\,\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\,\mathrm{Ai}(y)}{x - y}$$

(Forrester 1993). See Hough, Krishnapur, Peres and Virág (2009) for definitions and facts about determinantal processes. The gap probabilities can be represented by Fredholm determinants: for example,

$$F_{(0)}(x) = \det(1 - \mathbf{A}_x)$$

where $\mathbf{A}_x$ represents the (trace class) integral operator with Airy kernel acting on $L^2(x, \infty)$. The subsequent eigenvalue laws are encoded as follows. For $0 \le \lambda \le 1$, the determinantal process with kernel $\lambda \mathbf{A}_x$ can be described probabilistically by sampling from the original $\lambda = 1$ process and then independently deleting each point with probability $1 - \lambda$. For the gap probability of this "incomplete spectrum" we then have the geometric mixture

$$\sum_{k=0}^{\infty} (1 - \lambda)^k \lambda\, F_{(k)}(x) = \det(1 - \lambda \mathbf{A}_x). \tag{5.20}$$

Tracy and Widom (1994) prove the identity of the right hand sides of (5.18) and (5.20).

Turning to our PDE characterization Theorem 2.4.3, consider the geometric mixture

$$F_\lambda(x; w) = \sum_{k=0}^{\infty} (1 - \lambda)^k \lambda\, F_{(k)}(x; w).$$

Recall that $F_{(k)}$ solves the usual PDE (2.9); by linearity, so does F_λ. Writing the boundary conditions informally, $F_{(k)}$ also solves the usual $F_{(k)}(+\infty, +\infty) = 1$ but now the recursive condition

$$F_{(k)}(x, -\infty) = F_{(k-1)}(x, +\infty).$$

It follows that F_λ solves $F_\lambda(+\infty, +\infty) = 1$ but now the "periodic" condition

$$F_\lambda(x, -\infty) = (1 - \lambda)\, F_\lambda(x, +\infty). \tag{5.21}$$

One can show that there is a unique bounded solution of the PDE with these boundary conditions; argue as in Section 2.4 but using a diffusion path that explodes and restarts a random number of times with Geometric(λ) distribution.

Now, one can also define λ-dependent "Ablowitz-Segur versions" of $v(x)$, $E(x), F(x)$, $f(x;w), g(x;w)$ by (5.3)–(5.6) but using $u(x,\lambda)$ characterized by (5.1),(5.19). Once again (5.14) also holds, and therefore $\tilde{F}_\lambda(x;w) = F(x)f(x,w)$ satisfies the PDE (2.9) by the same independent verification as in Section 5.1. Remarkably, it is also possible to verify independently that this f and hence $\tilde{F}_\lambda$ satisfy the same boundary conditions as those satisfied by F_λ, especially the periodic boundary conditon (5.21)! We omit the details, deferring to a personal communication (Its 2011).

5.3 Higher-rank deformations

Baik (2006) also gives a formula for the multi-parameter function $F_2(x;w_1,\ldots,w_r)$ which appeared originally in BBP and which we also characterize in Theorem 3.1.6. While we do not have an independent proof of Baik's formula at present, we used the computer algebra system Maple to verify symbolically that it does indeed satisfy our PDE (3.6) at $\beta = 2$ for $r = 2, 3, 4, 5$. Of course, a pencil-and-paper proof for all r would be much more satisfying. (It would then remain to verify boundedness and the boundary conditions (3.7),(3.8).)

Baik's formula is

$$F_2(x;w_1,\ldots,w_r) = F(x)\frac{\det\Big(\big(w_i + \frac{\partial}{\partial x}\big)^{j-1} f(x,w_i)\Big)_{1\le i,j\le r}}{\prod_{1\le i<j\le r}(w_j - w_i)}. \tag{5.22}$$

Our symbolic verification for small values of r consisted of the following steps. The differential relations given by (5.1), (5.3), (5.4), (5.5) and (5.14) were encoded as formal substitution rules. The determinant in (5.22) was expanded (this step becomes problematic for larger r!) and the result plugged into our PDE (3.6). The substitution rules were then applied repeatedly. Finally, the result was factored using Maple's built-in command. Each time, the output contained the factor

$$v + u^4 - (u')^2 + xu^2,$$

which vanishes identically. (Once again, differentiate and apply (5.1) to see it is constant and then take $x \to \infty$ to see the constant is zero.)

Chapter 6

A note on numerics

Consider the problem of evaluating the Tracy-Widom distributions $F_\beta(x)$ numerically. One use for such values (especially with $\beta = 1$) is in hypothesis testing on high-dimensional data, e.g. with the so-called largest root test (Johnstone 2001, 2007, 2008, 2009). Another use has been in high-precision experimental verification of mathematical predictions about growth interface fluctuations (Takeuchi and Sano 2010). A third use is in "experimental mathematics": good numerics allow one to test proposed relations or identities before attempting a proof, and even sometimes to discover new identities (Bornemann 2010a).

If there were no structure at all, the only recourse would be to Monte Carlo simulation with large random matrices. Of course, the Fredholm determinant and Painlevé II representations suggest better alternatives. For a long time, the common point of view was that the Painlevé representations offered the most straightforward method. As this method essentially only requires solving an ODE/initial value problem on the line, one might reasonably expect this to be the case. Unfortunately, naïve solution of the Painlevé II equation (5.1) with Hastings-Mcleod asymptotic data (5.2) is well-known to be unstable and inevitably leads to large errors. Deeper knowledge of this nonlinear special function (the so-called "connection formulas" giving the left tail asymptotics) allowed Tracy and Widom and especially Prähofer and Spohn (2004) to obtain good results. The latter authors obtained tables with 16-digit accuracy (Prähofer 2003) but used variable precision software arithmetic with up to 1500 digits. More recently, Bornemann (2010b) demonstrated that the Fredholm determinant representations actually offer the better route to efficient and accurate numerical evaluation. See Bornemann (2010a) for a review of these developments.

In this short note we suggest a third approach, under development with Brian Sutton, based on numerical solution of the boundary value problem (2.9),(2.10). While numerical solution of PDE can in general be computationally demanding and fraught with potential problems, preliminary results using a standard solution scheme are very promising.

The main novelty of the approach is that it is "general β": in stark contrast to the other approaches, one can switch from F_1 to F_2 to F_4 by changing a single parameter in the code. Of course one can also evaluate F_β for nonclassical β, and here our method seems far superior to the one suggested in RRV, namely Monte-Carlo simulation based on the diffusion representation. Another new feature is that one automatically computes an entire table of values at once, rather than evaluating one argument at a time. One further obtains all the rank one deformed distributions $F_\beta(x; w)$ for free along the way, and for $\beta = 1$ this is so far the only reasonable way to evaluate these distributions numerically. (Using (3.6), it should be possible to obtain the higher rank deformations as well.) Finally, extension to subsequent eigenvalue distributions is completely straightforward. A current limitation, however, is the lack of access to joint distributions of the eigenvalues (i.e. multi-point correlation functions of the soft edge limit).

Turning to the details of the implementation, the PDE (2.9) is a fairly standard diffusion-advection equation with space variable w and time variable $-x$. The main apparent difficulty with the formulation of the boundary value problem (2.9), (2.10) is that the boundary conditions and the desired slice of the solution are all at infinity. There is a natural way around this problem for the w variable: in deriving the PDE, instead of using the affine Riccatti coordinate $w = f'/f$ in the (f, f') phase plane of the stochastic Airy equation, use the angular Prüfer coordinate $\theta = \arg(f, f')$. This amounts to simply making a tangent substitution; we actually use $w = -\cot\theta$ to put the bottom boundary $w = -\infty$ at $\theta = 0$.

Abusing notation we still write $F(x, \theta)$ and the equation becomes

$$\frac{\partial F}{\partial x} + \left(\frac{2}{\beta}\sin^4\theta\right)\frac{\partial^2 F}{\partial \theta^2} + \left(\left(x + \frac{2}{\beta}\sin 2\theta\right)\sin^2\theta - \cos^2\theta\right)\frac{\partial F}{\partial \theta} = 0.$$

Note that the coefficients are smooth: there is in fact no singularity at $\theta = k\pi$. The boundary conditions become

$$\begin{aligned} F &\to 1 \qquad \text{as } x \to \infty \text{ with } \theta \geq \theta_0 > 0 \\ F &= 0 \qquad \text{on } \theta = 0 \end{aligned}$$

and one finds the undeformed Tracy-Widom(β) distribution at the $\theta = \pi$ slice, i.e.

$F_\beta = F(\cdot, \pi)$. The equation makes sense for all $\theta > 0$, however, and by Theorem 2.4.3, one finds the k'th largest eigenvalue law at $\theta = k\pi$.

To deal with the boundary condition at $x = +\infty$, we cut the domain off at, say, $x = x_0 = 10$ and use the Gaussian asymptotics obtained in Theorem 4.1.1. Specifically, we use the approximately valid initial data

$$F(x_0, \theta) = \begin{cases} \Phi\left(\dfrac{x_0 - \cot^2\theta}{\sqrt{(4/\beta)\cot\theta}}\right) & 0 \le \theta \le \pi/2 \\ 1 & \theta \ge \pi/2 \end{cases}$$

where Φ is the standard normal distribution function. (Plots confirm this initial data is correct. However, there appears to be some inherent stability in the sense that using "wrong" initial data at $x = 10$ barely affects the results for, say, $x \le 9$.)

At this stage we can almost plug the boundary value problem into an off-the-shelf numerical solution package such as Mathematica's `NDSolve`. We still need an upper boundary, so we cut the domain off arbitrarily at $\theta = \theta_1$. Mathematically there is no need for an upper boundary condition: at $\theta = k\pi$ the equation has vanishing diffusivity and a unit upward drift (going in the positive time or $-x$ direction). One might incorporate this property into a custom numerical scheme using upwind differencing, but off-the-shelf schemes require a boundary condition and we simply use $F(x, \theta_1) = 1$. If we are interested in values $\theta \le k\pi$, it seems wise to use $\theta_1 = (k+1)\pi$ at least.

Choosing any final time, say $x = x_1 = -10$, it is now possible to plug our linear parabolic initial/boundary value problem on the box $[x_1, x_0] \times [0, \theta_1]$ into Mathematica's `NDSolve`. Here is a sample implementation (with `b` $= \beta = 1$ and `z` $= \theta$):

```
b = 1;
h[x_, z_] :=
  Piecewise[{{0, z == 0},
    {CDF[NormalDistribution[Cot[z]^2,Sqrt[4/b*Cot[z]]],x], 0<z<Pi/2},
    {1, z>=Pi/2}}]
x0 = 10;  xm = -10; zm = 3 Pi;
sol = NDSolve[{D[F[x,z],x] + (2/b)Sin[z]^4*D[F[x,z],{z,2}]
          +((x+(2/b)Sin[2*z])*Sin[z]^2-Cos[z]^2) D[F[x,z],z] == 0,
          F[x0,z] == h[x0,z], F[x,0] == 0, F[x,zm] == 1},
        F, {z,0,zm}, {x,x0,xm}, MaxStepSize->.005, AccuracyGoal->6];
```

This simple code executes in around 10 seconds on a laptop. The resulting object, essentially a table of values, is of Mathematica type `InterpolatingFunction`. It can therefore be conveniently evaluated at any point; slices can be numerically integrated, differentiated or even inverted to solve for quantiles. We run a very simple test:

```
Fb[x_] := F[x, Pi] /. sol[[1]];
NumberForm[{x,Fb[x],Log[Fb'[x]]},8]/.x->2
NumberForm[{x,Fb[x],Log[Fb'[x]]},8]/.x->-2
```

These commands define F_β (recall we put $\beta = 1$) and evaluate $F_1(x)$ and $\log F_1'(x)$ at $x = 2, -2$ with the output displayed to 8 digits:

```
{ 2, 0.98959757, -4.0472670}
{-2, 0.27432019, -1.2680192}
```

These values are in agreement up to the last digit with those found in Prähofer (2003). One finds similarly good results over a range of x-values and for $\beta = 2$ as well. Numerically integrating to find the mean and variance also agrees well with published results for $\beta = 1, 2, 4$. (Note that for $\beta = 4$ published results generally use an old scaling convention whereby the independent variable is off by a factor $2^{1/6}$.)

If one wants fast results for plotting or experimenting, it is possible to tweak `MaxStepSize` and `AccuracyGoal` to make the code yield 5–6 digits in less than a second. At this point, attempting to push for more than 8–9 digits seems to increase compute time significantly without giving uniformly better results. One should remember, however, that the schemes used here are not tailored to the problem at hand. We are hopeful that the method presented could be improved and optimized in many respects.

Appendix A

Stochastic Airy is a classical Sturm-Liouville problem

The spectral theory of the Stochastic Airy operator and its Riccati diffusion representation are developed in Ramírez, Rider and Virág (2011) (**RRV**). It is possible to recast this development to take advantage of classical Sturm-Liouville theory. The point is that continuous integrated potentials actually fall well within the scope of the standard theory. One rewrites the eigenvalue equation as a pair of first order linear ODEs with continuous coefficients; this system is in fact equivalent to the integrated form of the eigenvalue equation in the paper. A further change of variables reduces the equation to standard Sturm-Liouville form with coefficients satisfying the most classical hypotheses.

The main advantage is that various standard results become available. For the problem on a finite interval, the classical Sturm oscillation and comparison theorems apply directly. One also more easily gets a full picture of the spectral theory, including for example the completeness of the eigenfunctions in $L^2(\mathbb{R}_+)$.

After the fact, one can apply the Riccati transformation to the linear system, obtaining a nonlinear ODE with continuous coefficients whose solutions are already understood. A final change of variables recovers the SDE from the ODE. This formulation has the further advantage of avoiding infinity: "blowups and restarts" only appear once the work is already done. Of course, they could be avoided altogether by working with the angular Prüfer coordinate instead of the affine Riccati coordinate in the phase plane; the resulting ODE and SDE have trigonometric rather than quadratic nonlinearity, however, and are somewhat more complicated (see Chapter 6).

The linear system

The formal second order linear differential equation

$$f''(x) = (x - \lambda + b'_x)f(x) \tag{A.1}$$

on $x \in \mathbb{R}_+$ with initial condition

$$f'(0) = wf(0) \tag{A.2}$$

has a standard classical interpretation. Note that the Dirichlet case $f(0) = 0$ is included by formally allowing $w = +\infty$, and we omit the $\frac{2}{\sqrt{\beta}}$ factor in the white noise. Following Section V.1 of Reid (1971), begin by rewriting (A.1) in the form

$$\big(f' - b_x f\big)' = (x - \lambda)f - b_x f'.$$

Now let $g = f' - b_x f$. The equation becomes

$$\begin{aligned} g' &= (x - \lambda)f - b_x f' \\ &= (x - \lambda - b_x^2)f - b_x g. \end{aligned}$$

In other words, the pair $\big(f(x), g(x)\big) \in \mathbb{R}^2$ formally satisfies the first order linear system

$$\begin{pmatrix} f' \\ g' \end{pmatrix} = \begin{pmatrix} b_x & 1 \\ x - \lambda - b_x^2 & -b_x \end{pmatrix} \begin{pmatrix} f \\ g \end{pmatrix}. \tag{A.3}$$

The initial condition becomes $\big(f(0), g(0)\big) = (1, w)$ up to a factor, (0,1).

Following Carathéodory one can allow general measurable coefficients and define a solution to be a pair of absolutely continuous functions (f, g) satisfying (A.3) Lebesgue a.e. This definition, equivalent to writing (A.3) in integrated form, is easily seen to coincide with the definition in RRV. Here the coefficients are continuous, however; such solutions may therefore be taken to satisfy (A.3) everywhere and are in fact continuously differentiable. It is classical that the initial value problem has a unique solution, which exists for all $x \in \mathbb{R}_+$ and depends smoothly on the parameter λ and the initial condition w (see e.g. Sections 1.5 and 1.7 of Coddington and Levinson 1955).

The Sturm-Liouville equation

While Reid (1971, 1980) works in a framework sufficiently general to include (A.3), he also discusses a standard reduction to a simpler form. It consists in the integrating factor substitution

$$\tilde{f} = e^{-\int_0^x b} f, \qquad \tilde{g} = e^{\int_0^x b} g,$$

which transforms (A.3) into

$$\begin{pmatrix} \tilde{f}' \\ \tilde{g}' \end{pmatrix} = \begin{pmatrix} 0 & e^{-2\int_0^x b} \\ (x-\lambda-b_x^2)e^{2\int_0^x b} & 0 \end{pmatrix} \begin{pmatrix} \tilde{f} \\ \tilde{g} \end{pmatrix}. \tag{A.4}$$

The auxilliary function $\tilde{g} = e^{2\int_0^x b}\tilde{f}'$ is sometimes called the "quasi-derivative". It is common to abbreviate (A.4) as

$$\left(e^{2\int_0^x b}\tilde{f}'\right)' = (x-\lambda-b_x^2)e^{2\int_0^x b}\tilde{f}. \tag{A.5}$$

The latter equation is in standard Sturm-Liouville form. As such it satisfies the most classical hypotheses: the coefficients are continuous, the coefficient of $\tilde{f}'$ is positive and continuously differentiable, and the "weight function" $e^{2\int_0^x b}$ that multiplies λ is positive.

In the language of functional analysis, the mapping $f \mapsto \tilde{f} = e^{-\int_0^x b} f$ is an isometry of $L^2(\mathbb{R}_+)$ onto the weighted space $L^2(\mathbb{R}_+, e^{2\int_0^x b}dx)$. It transforms the stochastic Airy operator

$$\mathcal{H} = -\frac{d^2}{dx^2} + x + b_x'$$

which is self-adjoint on $L^2(\mathbb{R}_+, dx)$ into the classical Sturm-Liouville operator

$$\widetilde{\mathcal{H}} = -e^{-2\int_0^x b}\frac{d}{dx}e^{2\int_0^x b}\frac{d}{dx} + x - b_x^2$$

which is self-adjoint on $L^2(\mathbb{R}_+, e^{2\int_0^x b}dx)$. Denoting the weighted inner product with a tilde and neglecting boundary terms, the associated quadratic form is simply

$$\begin{aligned} \left\langle \tilde{f}, \widetilde{\mathcal{H}}\tilde{f}\right\rangle_\sim &= \left\langle \tilde{f}', \tilde{f}'\right\rangle_\sim + \left\langle \tilde{f}, (x-b_x^2)\tilde{f}\right\rangle_\sim \\ &= \left\langle f' - b_x f,\ f' - b_x f\right\rangle + \left\langle f, (x-b_x^2)f\right\rangle \\ &= \left\langle f', f'\right\rangle + \left\langle f, xf\right\rangle - 2\left\langle f', b_x f\right\rangle \end{aligned}$$

which is precisely the integrated-by-parts definition of $\left\langle f, \mathcal{H}f\right\rangle$ given in RRV!

Sturm-Liouville theory

In the 1830s Sturm and Liouville studied so-called regular self-adjoint second order boundary value problems, marking the first systematic investigation of an important class of equations that generally lack explicit solutions. Here "regular" basically means that the problem is posed on a compact interval. In 1910 Weyl initiated the study of the much more complicated singular case, which includes the present problem.

Weyl's theory begins with the "limit-point/limit-circle" dichotomy, explained in Section 9.2 of Teschl (2009). In functional analytic language, the point is to understand the self-adjoint versions of the operator. In our case $\widetilde{\mathcal{H}}$ is regular and hence limit-circle at the left endpoint 0. At the right endpoint ∞, $\widetilde{\mathcal{H}}$ is limit-point by Weyl's criterion (Theorem 9.9 in Teschl 2009) since there are many solutions of (A.5) that are not square integrable near infinity. It follows that, for each $w \in (-\infty, \infty]$, the operator $\widetilde{\mathcal{H}}$ with domain

$$\left\{ f \in L^2(\mathbb{R}_+) \,:\, f', \left(e^{2\int_0^x b} f'\right)' \in L^1_{\text{loc}},\ \widetilde{\mathcal{H}} f \in L^2,\ f'(0) = w f'(0) \right\}$$

is self-adjoint (Theorem 9.6 in Teschl 2009).

The spectral theorem and the min-max theorem now apply (Theorems 3.7 and 4.10 in Teschl 2009). The variational characterization in RRV, together with the fact that eigenvalues accumulate only at infinity (proved very simply just before Lemma 2.2.7), implies rigorously that there is no essential spectrum. The eigenfunctions form a complete orthonormal basis for $L^2(\mathbb{R}_+ e^{2\int_0^x b} dx)$, and the resolvents are compact operators on the latter space. The completeness transfers to the eigenfunctions of $\mathcal{H}$ in $L^2(\mathbb{R}_+)$, implying the expansions

$$\mathcal{H} f = \sum_{k=0}^{\infty} \lambda_k \langle f_k, f \rangle f_k,$$
$$\left\langle f, \mathcal{H} f \right\rangle = \sum_{k=0}^{\infty} \lambda_k \left| \langle f, f_k \rangle \right|^2.$$

While there are oscillation theorems for singular Sturm-Liouville problems, they are far less straightforward than the one for the classical regular case. Here it still seems best to proceed as in RRV, truncating to approximate the singular problem with regular problems that converge to it in norm resolvent sense.

We therefore consider the truncated eigenvalue problem $\mathcal{H}^L$ on a compact interval $[0, L]$ with Dirichlet condition $f(L) = 0$ at the right endpoint. In the form (A.5), we can use the most classical oscillation theorem in Section 8.2 of Coddington and Levinson (1955); see also V.7 of Reid (1971) or II.5 of Reid (1980). The theorem states that *the spectrum of $\mathcal{H}^L$ is purely discrete, the eigenvalues are simple and form a sequence $\lambda_0^L < \lambda_1^L < \ldots$ with $\lambda_k^L \to \infty$ as $k \to \infty$, and the eigenfunction f_k^L corresponding to λ_k^L has exactly k zeros on $(0, L)$.*

Comparison theorems in 8.1 of Coddington and Levinson (1955) or V.7 of Reid (1971) state that, for each λ, the kth largest zero of the solution $f(x; \lambda)$ of the initial value

problem on $\mathbb{R}_+$ is a continuous decreasing function of λ as soon as it exists. We conclude that *for each λ, the number of zeros of $f(\cdot;\lambda)$ in $(0,L)$ equals the number of eigenvalues of $\mathcal{H}^L$ strictly below λ.*

Lemma 3.3 of RRV and its proof give that $\lambda_k^L \to \lambda_k$ and $f_k^L \to_{L^2} f_k$ as $L \to \infty$, where λ_k, f_k are the eigenvalues and eigenfunctions of $\mathcal{H}$ on $L^2(\mathbb{R}_+)$. (The initial condition at 0 can be general.) Equivalently, the spectral projections $\mathbf{1}_{(-\infty,\lambda)}\mathcal{H}^L \to \mathbf{1}_{(-\infty,\lambda)}\mathcal{H}$ in norm for all $\lambda \in \mathbb{R} \setminus \{\lambda_0, \lambda_1, \ldots\}$; the operators converge in norm resolvent sense. Taking $L \to \infty$ in the claim of the previous paragraph, we see that *for each $\lambda \in \mathbb{R} \setminus \{\lambda_0, \lambda_1, \ldots\}$, the number of zeros of $f(\cdot;\lambda)$ in $(0,\infty)$ equals the number of eigenvalues of $\mathcal{H}$ strictly below λ.* As $\lambda \searrow \lambda_k$ the last zero must tend to ∞ by monotonicity and continuity, so the claim actually holds for all $\lambda \in \mathbb{R}$.

Riccati equation and diffusion

We return to (A.3). The right-hand side describes a two-dimensional vector field. Because it is linear, it factors through the projective quotient into a one-dimensional vector field on the projective line; this is the essence of the Riccati transformation. (These vector fields integrate respectively to flows of linear transformations and fractional linear transformations.)

By uniqueness, a nontrivial solution (f, g) never passes through $(0, 0)$. At a zero of f we have $f' = g$, and it follows that the zeros of f are isolated. On an interval where f does not vanish, the projective coordinate $q = g/f$ is continuously differentiable and satsifies the Riccati equation

$$q' = x - \lambda - (q + b)^2.$$

At a zero of f, q explodes to $-\infty$ and restarts at $+\infty$. To see the evolution through such an explosion, switch to the other coordinate $\tilde{q} = f/g = 1/q$, which satisfies

$$\tilde{q}' = (1 + \tilde{q}s)^2 - (x - \lambda)\tilde{q}^2.$$

Notice how $\tilde{q}' = 1$ when $\tilde{q} = 0$.

Now let $p = q + b = f'/f$. While p is not differentiable, it certainly satisfies the integral equation

$$p_x - p_0 = b_x + \int_0^x (y - \lambda - p_y^2)\, dy.$$

In other words, *p_x is a strong solution of the Itō equation*

$$dp_x = db_x + (x - \lambda - p_x^2)dx$$

To see the behavior through ∞, let $\tilde{p} = 1/p = f/f'$. By Itō's lemma, $\tilde{p} = \tilde{q}/(1 + b\tilde{q})$ solves

$$d\tilde{p}_x = -\tilde{p}_x^2 db_x + (1 - (x - \lambda)\tilde{p}_x^2 + \tilde{p}_x^3)dx.$$

Inverse questions

It would be interesting to obtain a better understanding of how the spectrum of the operator and its rank one perturbations are related to the Brownian path in the stochastic Airy potential. It is possible that certain "inverse questions" could be answered, see e.g. Pöschel and Trubowitz (1987).

For a preliminary question, the stochastic Airy spectrum is (pathwise almost surely) monotone in the rank one perturbation w; it would be nice to know that it is strictly monotone and analytic. In particular, given $\lambda \in \mathbb{R}$, is it true that there is exactly one $w \in (-\infty, \infty]$ such that λ is an eigenvalue? The analogous statement in the finite dimensional setting is that, for a given matrix A, cyclic vector v and λ with $A - \lambda \not\geq 0$ there is exactly one $t \in [0, \infty)$ so that $\lambda \in \text{spec}(A + tvv^*)$. This fact follows from Lemma A.1 of Stolz (2011) and it seems that there should be a fully analogous picture for stochastic Airy.

Bibliography

Ablowitz, M. J. and Segur, H. (1977). Exact linearization of a Painlevé transcendent, *Phys. Rev. Lett.* **38**: 1103–1106.

Adler, M., Delépine, J. and van Moerbeke, P. (2009). Dyson's nonintersecting Brownian motions with a few outliers, *Comm. Pure Appl. Math.* **62**: 334–395.

Aldous, D. and Steele, J. M. (2004). The objective method: probabilistic combinatorial optimization and local weak convergence, *Probability on discrete structures*, Springer, pp. 1–72.

Anderson, G., Guionnet, A. and Zeitouni, O. (2009). *An Introduction to Random Matrices*, Cambridge University Press.

Anderson, T. W. (1963). Asymptotic theory for principal component analysis, *Ann. Math. Statist.* **34**: 122–148.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, third edn, Wiley-Interscience.

Bai, Z. D. (1999). Methodologies in spectral analysis of large-dimensional random matrices, a review, *Statist. Sinica* **9**: 611–677.

Bai, Z. D. and Silverstein, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices, *Ann. Probab.* **26**: 316–345.

Bai, Z. D. and Silverstein, J. W. (1999). Exact separation of eigenvalues of large-dimensional sample covariance matrices, *Ann. Probab.* **27**: 1536–1555.

Bai, Z. and Yao, J.-f. (2008). Central limit theorems for eigenvalues in a spiked population model, *Ann. Inst. Henri Poincaré Probab. Stat.* **44**: 447–474.

Baik, J. (2006). Painlevé formulas of the limiting distributions for nonnull complex sample covariance matrices, *Duke Math. J.* **133**: 205–235.

Baik, J., Ben Arous, G. and Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices, *Ann. Probab.* **33**: 1643–1697.

Baik, J., Deift, P. and Johansson, K. (1999). On the distribution of the length of the longest increasing subsequence of random permutations, *J. Amer. Math. Soc.* **12**: 1119–1178.

Baik, J. and Rains, E. M. (2000). Limiting distributions for a polynuclear growth model with external sources, *J. Statist. Phys.* **100**: 523–541.

Baik, J. and Rains, E. M. (2001). The asymptotics of monotone subsequences of involutions, *Duke Math. J.* **109**: 205–281.

Baik, J. and Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models, *J. Multivariate Anal.* **97**: 1382–1408.

Baik, J. and Wang, D. (2011). On the largest eigenvalue of a Hermitian random matrix model with spiked external source II. Higher rank case, *arXiv:1104.2915*.

Bassler, K. E., Forrester, P. J. and Frankel, N. E. (2010). Edge effects in some perturbations of the Gaussian unitary ensemble, *J. Math. Phys.* **51**: 123305, 16.

Baur, G. and Kratz, W. (1989). A general oscillation theorem for selfadjoint differential systems with applications to Sturm-Liouville eigenvalue problems and quadratic functionals, *Rend. Circ. Mat. Palermo (2)* **38**: 329–370.

Ben Arous, G. and Corwin, I. (2011). Current fluctuations for TASEP: a proof of the Prähofer-Spohn conjecture, *Ann. Probab.* **39**: 104–138.

Benaych-Georges, F. and Nadakuditi, R. R. (2009). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices, *arXiv:0910.2120v2*.

Bloemendal, A. and Virág, B. (2010). Limits of spiked random matrices I, *arXiv:1011.1877v1*.

Bornemann, F. (2010a). On the numerical evaluation of distributions in random matrix theory: A Review, *Markov Processes Relat. Fields* **16**: 803–866.

Bornemann, F. (2010b). On the numerical evaluation of Fredholm determinants, *Math. Comp.* **79**: 871–915.

Bourgade, P., Erdős, L. and Yau, H.-T. (2011). Universality of general β-ensembles, *arXiv:1104.2272v2*.

Coddington, E. A. and Levinson, N. (1955). *Theory of Ordinary Differential Equations*, McGraw-Hill.

Deift, P. (2007). Universality for mathematical and physical systems, *International Congress of Mathematicians. Vol. I*, Eur. Math. Soc., Zürich, pp. 125–152.

Deift, P. A. (1999). *Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach*, Courant Lecture Notes in Mathematics, New York University.

Deift, P. A. and Zhou, X. (1995). Asymptotics for the Painlevé II equation, *Comm. Pure Appl. Math.* **48**: 277–337.

Desrosiers, P. and Forrester, P. J. (2006). Asymptotic correlations for Gaussian and Wishart matrices with external source, *Int. Math. Res. Not.* **2006**: Art. ID 27395, 43 pp.

Dieng, M. (2005). Distribution functions for edge eigenvalues in orthogonal and symplectic ensembles: Painlevé representations, *Int. Math. Res. Not.* pp. 2263–2287.

Dumitriu, I. and Edelman, A. (2002). Matrix models for beta ensembles, *J. Math. Phys.* **43**: 5830–5847.

Dyson, F. J. (1962). A Brownian-motion model for the eigenvalues of a random matrix, *J. Mathematical Phys.* **3**: 1191–1198.

Edelman, A. and Sutton, B. D. (2007). From random matrices to stochastic operators, *J. Stat. Phys.* **127**: 1121–1165.

El Karoui, N. (2003). On the largest eigenvalue of Wishart matrices with identity covariance when n, p, and $p/n \to \infty$, *arXiv:math/0309355v1*.

El Karoui, N. (2007). Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices, *Ann. Probab.* **35**: 663–714.

Erdős, L. and Yau, H.-T. (2011). Universality of local spectral statistics of random matrices, *arXiv:1106.4986*.

Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence*, John Wiley & Sons, Inc.

Féral, D. and Péché, S. (2007). The largest eigenvalue of rank one deformation of large Wigner matrices, *Comm. Math. Phys.* **272**: 185–228.

Féral, D. and Péché, S. (2009). The largest eigenvalues of sample covariance matrices for a spiked population: diagonal case, *J. Math. Phys.* **50**: 073302, 33 pp.

Fokas, A. S., Its, A. R., Kapaev, A. A. and Novokshenov, V. Y. (2006). *Painlevé Transcendents: The Riemann-Hilbert Approach*, American Mathematical Society.

Forrester, P. J. (1993). The spectrum edge of random matrix ensembles, *Nuclear Phys. B* **402**: 709–728.

Forrester, P. J. (2010). *Log-gases and Random Matrices*, Princeton University Press.

Forrester, P. J. (2011). Probability densities and distributions for spiked Wishart β-ensembles, *arXiv:1101.2261v1*.

Geman, S. (1980). A limit theorem for the norm of random matrices, *Ann. Probab.* **8**: 252–261.

Halmos, P. (1957). *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*, Chelsea Publishing Co.

Harding, M. (2008). Explaining the single factor bias of arbitrage pricing models in finite samples, *Economics Letters* **99**: 85–88.

Hastings, S. P. and McLeod, J. B. (1980). A boundary value problem associated with the second Painlevé transcendent and the Korteweg-de Vries equation, *Arch. Rational Mech. Anal.* **73**: 31–51.

Hough, J. B., Krishnapur, M., Peres, Y. and Virág, B. (2009). *Zeros of Gaussian analytic functions and determinantal point processes*, American Mathematical Society.

Its, A. (2011). Personal communication.

Johansson, K. (2000). Shape fluctuations and random matrices, *Comm. Math. Phys.* **209**: 437–476.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis, *Ann. Statist.* **29**: 295–327.

Johnstone, I. M. (2007). High dimensional statistical inference and random matrices, *International Congress of Mathematicians. Vol. I*, Eur. Math. Soc., Zürich, pp. 307–333.

Johnstone, I. M. (2008). Multivariate analysis and Jacobi ensembles: largest eigenvalue, Tracy-Widom limits and rates of convergence, *Ann. Statist.* **36**: 2638–2716.

Johnstone, I. M. (2009). Approximate null distribution of the largest root in multivariate analysis, *Ann. Appl. Stat.* **3**: 1616–1633.

Killip, R. and Stoiciu, M. (2009). Eigenvalue statistics for CMV matrices: from Poisson to clock via random matrix ensembles, *Duke Math. J.* **146**: 361–399.

Krishnapur, M., Rider, B. and Virág, B. (2011+). In preparation.

Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices, *Mat. Sb. (N.S.)* **72 (114)**: 507–536.

Mehta, M. L. (2004). *Random Matrices*, third edn, Elsevier/Academic Press.

Mo, M. Y. (2011). The rank 1 real Wishart spiked model, *arXiv:1101.5144v1*.

Morse, M. (1932). *The calculus of variations in the large*, American Mathematical Society Colloquium Publications, Vol. 18 (1996 reprint of the original).

Morse, M. (1973). *Variational analysis: critical extremals and Sturmian extensions*, Interscience Publishers, John Wiley & Sons, Inc.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, John Wiley & Sons Inc.

Onatski, A. (2008). The Tracy-Widom limit for the largest eigenvalues of singular complex Wishart matrices, *Ann. Appl. Probab.* **18**: 470–490.

Patterson, N., Price, A. L. and Reich, D. (2006). Population structure and eigenanalysis, *PLoS Genetics* **2**: e190.

Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model, *Statist. Sinica* **17**: 1617–1642.

Péché, S. (2006). The largest eigenvalue of small rank perturbations of Hermitian random matrices, *Probab. Theory Related Fields* **134**: 127–173.

Péché, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles, *Probab. Theory Related Fields* **143**: 481–516.

Pöschel, J. and Trubowitz, E. (1987). *Inverse Spectral Theory*, Academic Press Inc.

Prähofer, M. (2003). Tables to: Exact scaling functions for one-dimensional stationary KPZ growth, `http://www-m5old.ma.tum.de/KPZ/`.

Prähofer, M. and Spohn, H. (2004). Exact scaling functions for one-dimensional stationary KPZ growth, *J. Statist. Phys.* **115**: 255–279. See numerical tables at *http://www-m5old.ma.tum.de/KPZ/*.

Ramírez, J. A. and Rider, B. (2009). Diffusion at the random matrix hard edge, *Comm. Math. Phys.* **288**: 887–906.

Ramírez, J. A., Rider, B. and Virág, B. (2011). Beta ensembles, stochastic Airy spectrum, and a diffusion, *J. Amer. Math. Soc.* **24**: 919–944.

Reed, M. and Simon, B. (1978). *Methods of Modern Mathematical Physics. IV. Analysis of operators*, Academic Press.

Reed, M. and Simon, B. (1980). *Methods of Modern Mathematical Physics. I. Functional analysis*, Academic Press.

Reid, W. T. (1971). *Ordinary Differential Equations*, John Wiley & Sons Inc., New York.

Reid, W. T. (1972). *Riccati Differential Equations*, Academic Press.

Reid, W. T. (1980). *Sturmian Theory for Ordinary Differential Equations*, Springer-Verlag.

Savchuk, A. M. and Shkalikov, A. A. (1999). Sturm-Liouville operators with singular potentials, *Math. Notes* **66**: 897–912.

Silverstein, J. W. and Bai, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices, *J. Multivariate Anal.* **54**: 175–192.

Soshnikov, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices, *J. Statist. Phys.* **108**: 1033–1056.

Stolz, G. (2011). An introduction to the mathematics of Anderson localization, *arXiv:1104.2317v1*.

Sutton, B. D. (2005). *The Stochastic Operator Approach to Random Matrix Theory*, PhD thesis, Massachusetts Institute of Technology.

Takeuchi, K. and Sano, M. (2010). Universal fluctuations of growing interfaces: Evidence in turbulent liquid crystals, *Phys. Rev. Lett.* **104**: 230601.

Takeuchi, K., Sano, M., Sasamoto, T. and Spohn, H. (2011). Growing interfaces uncover universal fluctuations behind scale invariance, *Sci Rep [Nature]* **1**.

Telatar, E. (1999). Capacity of multi-antenna Gaussian channels, *Europ. Trans. Telecom.* **10**: 585–595.

Teschl, G. (2009). *Mathematical Methods in Quantum Mechanics*, Graduate Studies in Mathematics, American Mathematical Society. Available on the author's website.

Tracy, C. A. and Widom, H. (1993). Level-spacing distributions and the Airy kernel, *Phys. Lett. B* **305**: 115–118.

Tracy, C. A. and Widom, H. (1994). Level-spacing distributions and the Airy kernel, *Comm. Math. Phys.* **159**: 151–174.

Tracy, C. A. and Widom, H. (1996). On orthogonal and symplectic matrix ensembles, *Comm. Math. Phys.* **177**: 727–754.

Tracy, C. A. and Widom, H. (2002). Distribution functions for largest eigenvalues and their applications, *Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002)*, Higher Ed. Press, pp. 587–596.

Trotter, H. F. (1984). Eigenvalue distributions of large Hermitian matrices; Wigner's semicircle law and a theorem of Kac, Murdock, and Szegő, *Adv. in Math.* **54**: 67–82.

Valkó, B. and Virág, B. (2009). Continuum limits of random matrices and the Brownian carousel, *Invent. Math.* **177**: 463–508.

Wang, D. (2008). *Spiked Models in Wishart Ensemble*, PhD thesis, Brandeis University. *arXiv:0804.0889v1.*

Weidmann, J. (1997). Strong operator convergence and spectral theory of ordinary differential operators, *Univ. Iagel. Acta Math.* **34**: 153–163.

Yin, Y. Q., Bai, Z. D. and Krishnaiah, P. R. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix, *Probab. Theory Related Fields* **78**: 509–521.