

OUTLIERS IN THE SPECTRUM OF LARGE DEFORMED UNITARILY INVARIANT MODELS

BY SERBAN T. BELINSCHI^{*, †, ‡, 1}, HARI BERCOVICI^{§, 2},
 MIREILLE CAPITAINE^{*} AND MAXIME FÉVRIER[¶]

CNRS—Institut de Mathématiques de Toulouse^{}, Queen’s University[†],
 Institute of Mathematics of the Romanian Academy[‡], Indiana University[§]
 and Université Paris-Sud[¶]*

We characterize the possible outliers in the spectrum of large deformed unitarily invariant additive and multiplicative models, as well as the eigenvectors corresponding to them. We allow both the nondeformed unitarily invariant model and the perturbation matrix to have nontrivial limiting eigenvalue distributions and spiked outliers in their spectrum. The free subordination functions play a key role in this analysis.

CONTENTS

1.	Introduction	3572
1.1.	Statement of the problem	3572
1.2.	Heuristics	3575
2.	Notation and statements of the main results	3576
2.1.	Additive perturbations	3578
2.2.	Multiplicative perturbations of nonnegative matrices	3582
2.3.	Multiplicative perturbations of unitary matrices	3584
3.	Free convolutions	3585
3.1.	Additive free convolution	3586
3.2.	Multiplicative free convolution on $[0, +\infty)$	3588
3.3.	Multiplicative free convolution on \mathbb{T}	3589
3.4.	Analytic subordination	3590
3.4.1.	The subordination functions equations for free additive convolution	3590
3.4.2.	The subordination functions equations for multiplicative free convolution on $[0, +\infty)$	3592
3.4.3.	The subordination functions equations for multiplicative free convolution on \mathbb{T}	3592
4.	Preliminary results	3594
4.1.	Boundary behavior and convergence of some families of analytic functions	3594
4.2.	Matrix-valued functions and maps	3596
5.	Proofs of the main results	3603

Received May 2015; revised April 2016.

¹Supported in part by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada and a RIG from Queen’s University.

²Supported in part by NSF grants.

MSC2010 subject classifications. Primary 60B20; secondary 46L54.

Key words and phrases. Random matrices, deformed unitarily invariant models, spikes and outliers, free probability and free subordination.

5.1. The additive model $X_N = A_N + U_N B_N U_N^*$ 3603
 5.1.1. Reduction to the almost sure convergence of a $p \times p$ matrix 3604
 5.1.2. Convergence of F_N 3605
 5.1.3. Proofs of the main results for the additive model 3607
 5.2. The multiplicative model $X_N = A_N^{1/2} U_N B_N U_N^* A_N^{1/2}$ 3616
 5.2.1. Reduction to the almost sure convergence of a $p \times p$ matrix 3616
 5.2.2. Convergence of F_N 3617
 5.2.3. Proofs of the main results for the positive multiplicative model 3619
 5.3. The unitary multiplicative model $X_N = A_N U_N B_N U_N^*$ 3622
 5.3.1. Proofs of the main results for the unitary multiplicative model 3622
 Acknowledgments 3623
 References 3623

1. Introduction.

1.1. *Statement of the problem.* The set of possible spectra for the sum of two deterministic Hermitian matrices A_N and B_N depends in complicated ways on the spectra of A_N and B_N (see [23]). Nevertheless, if one adds some randomness to the eigenspaces of B_N then, as N becomes large, free probability provides a good understanding of the behavior of the spectrum of this sum. More precisely, set $X_N = A_N + U_N B_N U_N^*$, where U_N is a random unitary matrix distributed according to the Haar measure on the unitary group $U(N)$, and suppose that the empirical eigenvalue distributions of A_N and B_N converge weakly to compactly supported distributions μ and ν , respectively. Building on the groundbreaking result of Voiculescu [39], Speicher proved in [36] the almost sure weak convergence of the empirical eigenvalue distribution of X_N to the free additive convolution $\mu \boxplus \nu$. This convolution is again a compactly supported probability measure on \mathbb{R} . A similar result holds for products of matrices: if A_N, B_N are in addition assumed to be nonnegative definite, then the empirical eigenvalue distribution of $A_N^{1/2} U_N B_N U_N^* A_N^{1/2}$ converges to the free multiplicative convolution $\mu \boxtimes \nu$, a compactly supported probability measure on $[0, +\infty)$. (We recall that $A_N^{1/2} U_N B_N U_N^* A_N^{1/2}$ and $B_N^{1/2} U_N A_N U_N^* B_N^{1/2}$ have the same eigenvalue distribution, and that \boxtimes is a commutative operation.) Finally, if A_N and B_N are deterministic unitary matrices, their empirical eigenvalue distributions are supported on the unit circle $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$ and the empirical eigenvalue distribution of $A_N U_N B_N U_N^*$ converges to the free multiplicative convolution $\mu \boxtimes \nu$, a probability measure supported on \mathbb{T} . (We refer the reader to [42] for an introduction to free probability theory. We describe later the mechanics of calculating the free convolutions \boxplus and \boxtimes .)

The fact that the empirical eigenvalue distribution of X_N converges weakly to $\mu \boxplus \nu$ does not mean that all the eigenvalues of X_N are close to the support of this measure. There can be outliers, though they must not affect the limiting empirical eigenvalue distribution. Sometimes one can argue that these outliers must in fact

exist. For instance, the case when the rank r of A_N and its nonzero eigenvalues are fixed is investigated by Benaych-Georges and Nadakuditi in [12]. Denote by

$$\gamma_1 \geq \dots \geq \gamma_s > 0 > \gamma_{s+1} \geq \dots \geq \gamma_r$$

these fixed eigenvalues. Of course, in this case $\mu = \delta_0$ is a point mass at 0, so the *limiting* behavior of the empirical eigenvalue distribution of X_N is not affected by such a matrix A_N . More precisely, the empirical eigenvalue distribution of $X_N = A_N + U_N B_N U_N^*$ converges weakly almost surely to the limiting empirical eigenvalue distribution ν of B_N . One can however detect, among the outlying eigenvalues of X_N , the influence of the eigenvalues of A_N above a certain critical threshold. We use the notation

$$\lambda_1(X) \geq \dots \geq \lambda_N(X)$$

for the eigenvalues of an $N \times N$ matrix X , repeated according to multiplicity. The Cauchy–Stieltjes transform of a finite positive Borel measure ν on \mathbb{R} is given by

$$G_\nu(z) = \int_{\mathbb{R}} \frac{d\nu(t)}{z - t}$$

for z outside the support of ν , and G_ν^{-1} denotes the inverse of this function relative to composition. When the support of ν is contained in the compact interval $[a, b]$, the branch of the inverse function G_ν^{-1} that satisfies $G_\nu^{-1}(0) = \infty$ is defined and real-valued on the interval (α, β) , where

$$\alpha = \lim_{x \uparrow a} G_\nu(x), \quad \beta = \lim_{x \downarrow b} G_\nu(x).$$

The following result is proved in [12], Theorems 2.1 and 2.2.

THEOREM 1.1. (1) *Denote by a and b the infimum and supremum of the support of ν , respectively. Assume that the smallest and largest eigenvalues of B_N converge almost surely to a and b , respectively. Then, almost surely for $1 \leq i \leq s$,*

$$\lim_{N \rightarrow \infty} \lambda_i(X_N) = \begin{cases} G_\nu^{-1}(1/\gamma_i), & \gamma_i > 1/\beta, \\ b, & \text{otherwise.} \end{cases}$$

Similarly, almost surely for $0 \leq j \leq r - s - 1$,

$$\lim_{N \rightarrow \infty} \lambda_{N-j}(X_N) = \begin{cases} G_\nu^{-1}(1/\gamma_{r-j}), & \gamma_{r-j} < 1/\alpha, \\ a, & \text{otherwise.} \end{cases}$$

(2) *Fix $i_0 \in \{1, \dots, r\}$ such that $1/\gamma_{i_0} \in (\alpha, \beta)$. For each N , define*

$$\lambda(N) = \begin{cases} \lambda_{i_0}(X_N), & \gamma_{i_0} > 0, \\ \lambda_{N-r+i_0}(X_N), & \gamma_{i_0} < 0 \end{cases}$$

and let u_N be a unit-norm eigenvector of X_N associated to the eigenvalue $\lambda(N)$. Then the following almost sure limits hold:

$$\lim_{N \rightarrow \infty} \|P_{\ker(\gamma_{i_0} I_N - A_N)} u_N\|^2 = \frac{-1}{\gamma_{i_0}^2 G'_v(G_v^{-1}(1/\gamma_{i_0}))},$$

and

$$\lim_{N \rightarrow \infty} \|P_{\ker(\gamma_i I_N - A_N)} u_N\|^2 = 0, \quad i \neq i_0.$$

This result lies in the lineage of recent, and not so recent, works studying the influence of fixed rank additive or multiplicative perturbations on the extremal eigenvalues of classical random matrix models, the seminal paper being [5], where the so-called BBP phase transition was observed. See also [5, 6, 29] for sample covariance matrices, [19, 22, 24, 32, 33] for deformed Wigner models and [31] for information-plus-noise models. These investigations were first extended to perturbations of arbitrary rank in [34] and [4] for sample covariance matrices, in [20] for deformed Wigner models and in [17] for information-plus-noise models. It is pointed out in [20] that the subordination function (relative to the free additive convolution of a semicircular distribution with the limiting empirical eigenvalue distribution of the perturbation) plays an important role in the fact that some eigenvalues of the deformed Wigner model separate from the bulk. In [16], it is explained how the results of [34] and [4] in the setting of sample covariance matrices can also be described in terms of the subordination function related to the free multiplicative convolution of a Marchenko–Pastur distribution with the limiting empirical eigenvalue distribution of the multiplicative perturbation.

Similar results have been obtained in [12], Theorems 2.7 and 2.8, for multiplicative perturbations of the type $A_N^{1/2} U_N B_N U_N^* A_N^{1/2}$, where $A_N - I_N \geq 0$ is of fixed rank $p \in \mathbb{N}$ and $B_N \geq 0$.

In this paper, we investigate the asymptotic behavior of the eigenvalues and corresponding eigenvectors of the following models:

- $X_N = A_N + U_N B_N U_N^*$, where $A_N = A_N^*$, $B_N = B_N^*$ are deterministic, and U_N is a Haar-distributed unitary random matrix;
- $X_N = A_N^{1/2} U_N B_N U_N^* A_N^{1/2}$, where $A_N, B_N \geq 0$ are deterministic, and U_N is a Haar-distributed unitary random matrix;
- $X_N = A_N U_N B_N U_N^*$, where $A_N, B_N \in U(N)$ are deterministic unitary matrices and U_N is a Haar-distributed unitary random matrix.

In the first two models, we assume that A_N and B_N have compactly supported limiting empirical eigenvalue distributions μ and ν , respectively. In the third model, we assume that the supports of μ and ν are not equal to the entire unit circle. A fixed number $p \in \mathbb{N}$ of eigenvalues of A_N —the *spikes*—lay outside the support of μ for all $N \in \mathbb{N}$, but all other eigenvalues of A_N converge uniformly to the support of μ . A similar hypothesis is made about B_N and ν . We answer the following questions:

- When are some of the eigenvalues of X_N almost surely located outside the support of the limiting empirical eigenvalue distribution $\mu \boxplus \nu$ (resp., $\mu \boxtimes \nu$) of X_N ? In other words, when does the spectrum of X_N have *outliers* almost surely?
- What is the behavior of the eigenvectors corresponding to the outliers of X_N ?

The interesting question of the nature of the fluctuations of the outlying eigenvalues around their theoretical limit is beyond the scope of this paper. Although the spiked models first appeared in the statistical literature [29], they have been studied from a theoretical probability point of view since then. This is the point of view we adopt: therefore, we do not claim a statistical relevance for the models under study, we do not have statistical application in mind and we do not examine statistical questions such as the estimation of spikes [3].

When there are no spikes, it was shown in [21] that, given a neighborhood V of the support of $\mu \boxplus \nu$, the eigenvalues of X_N are almost surely contained in V for large $N \in \mathbb{N}$. Our paper extends the results of [12] to perturbations of arbitrary rank, and it also extends the free probabilistic interpretation of outlier phenomena in terms of subordination functions as described in [20] for deformations of Wigner models. Indeed, the occurrence and role of Biane’s subordination functions [14] in the analysis of the interaction spikes/outliers is quite natural. We clarify this in the additive case through the following heuristics, leaving the precise statements to Section 2.

1.2. *Heuristics.* Let \mathbf{a} and \mathbf{b} be two free self-adjoint random variables in a tracial W^* -probability space. Biane showed ([14], Theorem 3.1) that there exists an analytic self-map $\omega: \mathbb{C}^+ \rightarrow \mathbb{C}^+$ of the upper half-plane $\mathbb{C}^+ = \{x + iy: y > 0\}$ (depending on \mathbf{a}, \mathbf{b}) such that

$$(1.1) \quad \mathbb{E}_{\mathbb{C}[\mathbf{a}]}[(z - (\mathbf{a} + \mathbf{b}))^{-1}] = (\omega(z) - \mathbf{a})^{-1}, \quad z \in \mathbb{C}^+.$$

Here, $\mathbb{E}_{\mathbb{C}[\mathbf{a}]}$ denotes the conditional expectation onto the von Neumann algebra generated by \mathbf{a} and 1. The function ω is referred to as a *subordination function*. (This formula is a particular case of Biane’s result. Formula (1.1) appears in this form in [41], Appendix.) The subordination function continues analytically via Schwarz reflection through the complement in \mathbb{R} of the spectrum of $\mathbf{a} + \mathbf{b}$. If $A_N \rightarrow \mathbf{a}$ in distribution as $N \rightarrow \infty$, while a single eigenvalue θ , common to all of the matrices A_N , stays outside the spectrum of \mathbf{a} , this eigenvalue will disappear in the large N limit, in the sense that it will not influence the spectrum of \mathbf{a} . Thus, the analytic function $\omega(z)$ will not be prevented a priori from taking the value θ .

However, if relation (1.1) were true with \mathbf{a}, \mathbf{b} and $\mathbb{E}_{\mathbb{C}[\mathbf{a}]}$ replaced by $A_N, U_N B_N U_N^*$ and \mathbb{E} , respectively, and with the *same* subordination function ω , then any number ρ in the domain of analyticity of ω such that $\omega(\rho) = \theta$ must generate a polar singularity for the right-hand side of (1.1). Therefore, each such ρ must generate a similar singularity for the term on the left-hand side of the same equality, thus necessarily producing an eigenvalue of $A_N + U_N B_N U_N^*$. While this scenario

is not exactly true, we prove that an approximate version does hold. Namely, we show that a compression of the matrix

$$\mathbb{E}[(z - (A_N + U_N B_N U_N^*))^{-1}]^{-1} + A_N$$

to a certain subspace V_N is close to $\omega(z)I_{V_N}$, almost surely as $N \rightarrow \infty$. This insight is crucial in our arguments.

It follows from our results that a remarkable phenomenon (first noted without proof in [12], Remark 2.12, for finite-rank perturbations) occurs: a single spike of A_N can generate asymptotically a finite, possibly arbitrarily large, set of outliers for X_N . This arises from the fact that the restriction to the real line of some subordination functions may be many-to-one. In other words, with the above notation, the set $\omega^{-1}(\{\theta\})$ may have cardinality strictly greater than 1, unlike the subordination function related to free convolution with a semicircular distribution that was used in [20].

The case of multiplicative perturbations is based on similar ideas, with the subordination function ω replaced by its multiplicative counterparts ([14], Theorems 3.5 and 3.6).

In addition to outliers, we investigate the corresponding eigenspaces of X_N . It turns out that the angle between the outlying eigenvectors and the eigenvectors associated to the original spikes is determined by Biane’s subordination function, this time via its derivative.

The paper is organized as follows. In Section 2, we describe in detail the matrix models to be analysed, and state the main results of the paper. In Section 3, we introduce free convolutions and the analytic transforms involved in their study. We give the functional equations characterising the subordination functions. In Section 4, we collect and prove a number of auxiliary results, and in Section 5 we prove the main results.

2. Notation and statements of the main results. We denote by $\mathbb{C}^+ = \{z \in \mathbb{C} : \Im z > 0\}$ the upper half-plane, by $\mathbb{C}^- = \{z \in \mathbb{C} : \Im z < 0\}$ the lower half-plane, and by $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ the unit disc in \mathbb{C} . The topological boundary of the unit disc is denoted by $\mathbb{T} = \partial\mathbb{D}$. For any vector subspace E of \mathbb{C}^m , P_E denotes the orthogonal projection onto E . $M_m(\mathbb{C})$ stands for the set of $m \times m$ matrices with complex entries, $GL_m(\mathbb{C})$ for the subset of invertible matrices, and $U(m) \subset GL_m(\mathbb{C})$ for the unitary group. The operator norm of a matrix X is $\|X\|$, its spectrum is $\sigma(X)$, its kernel is $\ker X$, its trace is $\text{Tr}_m(X)$ and its normalized trace is $\text{tr}_m(X) = \frac{1}{m}\text{Tr}_m(X)$. The eigenvalues of a Hermitian matrix X are denoted

$$\lambda_1(X) \geq \dots \geq \lambda_m(X),$$

and the probability measure

$$\mu_X = \frac{1}{m} \sum_{i=1}^m \delta_{\lambda_i(X)}$$

is the empirical eigenvalue distribution of X . When X is unitary, its eigenvalues are ordered decreasingly according to the size of their principal arguments in $[0, 2\pi)$. If $X \in M_m(\mathbb{C})$ is a normal matrix, we denote by E_X its spectral measure. Thus, if $S \subseteq \mathbb{C}$ is a Borel set, then $E_X(S)$ is the orthogonal projection onto the linear span of all eigenvectors of X corresponding to eigenvalues in S . The support of a measure ρ on \mathbb{C} is denoted $\text{supp}(\rho)$. Given any set $K \subseteq \mathbb{R}$, we define its ε -neighborhood by

$$K_\varepsilon = \{x \in \mathbb{R} : \inf\{|x - y| : y \in K\} < \varepsilon\}.$$

As long as there is no risk of confusion, the same notation will be used when K and K_ε are subsets of \mathbb{T} . Open intervals in \mathbb{R} and open arcs in \mathbb{T} are denoted (a, b) .

As already seen in Section 1, the Cauchy (or Cauchy–Stieltjes) transform of a Borel probability measure ρ on \mathbb{C} is an analytic function defined by

$$(2.1) \quad G_\rho(z) = \int_{\mathbb{C}} \frac{1}{z - t} d\rho(t), \quad z \in \mathbb{C} \setminus \text{supp}(\rho).$$

We only consider finite measures ρ supported on \mathbb{R} or \mathbb{T} . We denote by F_ρ the reciprocal of G_ρ , that is, $F_\rho(z) = 1/G_\rho(z)$. The moment generating function for ρ is

$$(2.2) \quad \psi_\rho(z) = \int_{\mathbb{C}} \frac{tz}{1 - tz} d\rho(t), \quad z \in \mathbb{C} \setminus \left\{z \in \mathbb{C} : \frac{1}{z} \in \text{supp}(\rho)\right\}.$$

The η -transform of ρ is defined as

$$\eta_\rho(z) = \frac{\psi_\rho(z)}{1 + \psi_\rho(z)}.$$

The relevant analytic properties of the transforms above are presented in Sections 3.1–3.3.

The *free additive convolution* of the Borel probability measures μ and ν on \mathbb{R} is denoted by $\mu \boxplus \nu$ and the *free multiplicative convolution* of the Borel probability measures μ and ν either on $[0, +\infty)$ or on \mathbb{T} is denoted by $\mu \boxtimes \nu$. Given μ, ν , denote by ω_1 and ω_2 the *subordination functions* corresponding to the free convolution $\mu \boxplus \nu$. They are known to be meromorphic on the complement of $\text{supp}(\mu \boxplus \nu)$. As the name suggests, they satisfy an analytic subordination property in the sense of Littlewood:

$$(2.3) \quad G_{\mu \boxplus \nu}(z) = G_\mu(\omega_1(z)) = G_\nu(\omega_2(z)).$$

A similar result holds for the multiplicative counterparts of the subordination functions. We have

$$(2.4) \quad \psi_{\mu \boxtimes \nu}(z) = \psi_\mu(\omega_1(z)) = \psi_\nu(\omega_2(z)),$$

where ω_1 and ω_2 are analytic on $\{z \in \mathbb{C} : 1/z \notin \text{supp}(\mu \boxtimes \nu)\}$. Free convolutions are defined in Sections 3.1–3.3, and the subordination functions are defined via functional equations in Sections 3.4.1–3.4.3.

In the following three subsections, we describe our models and state the main results. To avoid dealing with too many special cases, we make the following technical assumption, which will be in force for the remainder of the paper, except for Remark 2.6.

(2.5) Neither of the two limiting measures μ, ν is a point mass.

Under this assumption, the subordination functions extend continuously to the boundary of their domains (see Lemmas 3.1, 3.2 and 3.3). Our results, however, remain substantially valid without this assumption, and we discuss in Remark 2.6 the relevant modifications.

2.1. *Additive perturbations.* Here are the ingredients for constructing the additive matrix model $X_N = A_N + U_N B_N U_N^*$:

- Two compactly supported Borel probability measures μ and ν on \mathbb{R} .
- A positive integer p and fixed real numbers

$$\theta_1 \geq \theta_2 \geq \dots \geq \theta_p$$

which do not belong to $\text{supp}(\mu)$.

- A sequence $(A_N)_{N \in \mathbb{N}}$ of deterministic Hermitian matrices of size $N \times N$ such that:
 - μ_{A_N} converges weakly to μ as $N \rightarrow \infty$;
 - for $N \geq p$ and $\theta \in \{\theta_1, \dots, \theta_p\}$, the sequence $(\lambda_n(A_N))_{n=1}^N$ satisfies

$$\text{card}(\{n : \lambda_n(A_N) = \theta\}) = \text{card}(\{i : \theta_i = \theta\});$$

- the eigenvalues of A_N which are not equal to some θ_i converge uniformly to $\text{supp}(\mu)$ as $N \rightarrow \infty$, that is,

$$\lim_{N \rightarrow \infty} \max_{\lambda_n(A_N) \notin \{\theta_1, \dots, \theta_p\}} \text{dist}(\lambda_n(A_N), \text{supp}(\mu)) = 0.$$

- A positive integer q and fixed real numbers

$$\tau_1 \geq \tau_2 \geq \dots \geq \tau_q$$

which do not belong to $\text{supp}(\nu)$.

- A sequence $(B_N)_{N \in \mathbb{N}}$ of deterministic Hermitian matrices of size $N \times N$ such that:
 - μ_{B_N} converges weakly to ν as $N \rightarrow \infty$;
 - for $N \geq q$ and $\tau \in \{\tau_1, \dots, \tau_q\}$, the sequence $(\lambda_n(B_N))_{n=1}^N$ satisfies

$$\text{card}(\{n : \lambda_n(B_N) = \tau\}) = \text{card}(\{j : \tau_j = \tau\});$$

- the eigenvalues of B_N which are not equal to some τ_j converge uniformly to $\text{supp}(\nu)$ as $N \rightarrow \infty$.

- A sequence $(U_N)_{N \in \mathbb{N}}$ of unitary random matrices such that the distribution of U_N is the normalized Haar measure on the unitary group $U(N)$.

It is known from [39] that the asymptotic empirical eigenvalue distribution of X_N is $\mu \boxplus \nu$. In the following statement, we take advantage of the fact, discussed later, that the functions ω_1 and ω_2 extend continuously to the real line. The points $\rho \in \mathbb{R} \setminus \text{supp}(\mu \boxplus \nu)$ satisfying $\omega_1(\rho) = \theta$ are isolated but they may accumulate to $\text{supp}(\mu \boxplus \nu)$. We denote by P_N and Q_N the projection onto the space generated by the eigenvectors corresponding to the spikes of A_N and B_N , respectively. These can also be written as

$$(2.6) \quad P_N = E_{A_N}(\{\theta_1, \dots, \theta_p\}), \quad Q_N = E_{B_N}(\{\tau_1, \dots, \tau_q\}),$$

in terms of the spectral measures of A_N and B_N . We are now ready to state our result for the additive model.

THEOREM 2.1. *With the above notation, set $K = \text{supp}(\mu \boxplus \nu)$, and*

$$K' = K \cup \left[\bigcup_{i=1}^p \omega_1^{-1}(\{\theta_i\}) \right] \cup \left[\bigcup_{j=1}^q \omega_2^{-1}(\{\tau_j\}) \right],$$

where ω_1, ω_2 are the subordination functions satisfying (2.3).

(1) *Given $\varepsilon > 0$, we have $\mathbb{P}(\exists N_0 \forall N, N > N_0, \sigma(X_N) \subset K'_\varepsilon) = 1$.*

(2) *Fix a number $\rho \in K' \setminus K$, let $\varepsilon > 0$ be such that $(\rho - 2\varepsilon, \rho + 2\varepsilon) \cap K' = \{\rho\}$, and set $k = \text{card}(\{i : \omega_1(\rho) = \theta_i\})$, $\ell = \text{card}(\{j : \omega_2(\rho) = \tau_j\})$. Then*

$$\mathbb{P}(\exists N_0 \forall N, N > N_0, \text{card}(\{\sigma(X_N) \cap (\rho - \varepsilon, \rho + \varepsilon)\}) = k + \ell) = 1.$$

(3) *With ρ and ε as in (2), we have*

$$\lim_{N \rightarrow \infty} \left\| P_N E_{X_N}((\rho - \varepsilon, \rho + \varepsilon)) P_N - \frac{1}{\omega'_1(\rho)} E_{A_N}(\{\omega_1(\rho)\}) \right\| = 0$$

and

$$\lim_{N \rightarrow \infty} \left\| Q_N U_N E_{X_N}((\rho - \varepsilon, \rho + \varepsilon)) U_N^* Q_N - \frac{1}{\omega'_2(\rho)} E_{B_N}(\{\omega_2(\rho)\}) \right\| = 0$$

almost surely.

(4) *With ρ and ε as in (2), suppose in addition that $\ell = 0$. Then almost surely,*

$$\lim_{N \rightarrow \infty} \sup \left\{ \left\| E_{A_N}(\{\omega_1(\rho)\}) \xi \right\|^2 - \frac{1}{\omega'_1(\rho)} \right\} : \xi \in E_{X_N}((\rho - \varepsilon, \rho + \varepsilon)) \mathbb{C}^N, \|\xi\|_2 \leq 1 \Big\} = 0.$$

Analogously, if $k = 0$, then almost surely

$$\lim_{N \rightarrow \infty} \sup \left\{ \left\| E_{U_N B_N U_N^*}(\{\omega_2(\rho)\}) \xi \right\|^2 - \frac{1}{\omega'_2(\rho)} \right\} : \xi \in E_{X_N}((\rho - \varepsilon, \rho + \varepsilon)) \mathbb{C}^N, \|\xi\|_2 \leq 1 \Big\} = 0.$$

REMARK 2.2. In case $k > 0$ and $\ell > 0$, the result of (3) above implies the following. Almost surely for ε small enough, if, for N large enough, $\{\xi_1, \dots, \xi_{k+\ell}\}$ is an orthonormal basis of $E_{X_N}((\rho - \varepsilon, \rho + \varepsilon))$, then

$$\lim_{N \rightarrow +\infty} \sum_{n=1}^{k+\ell} \|E_{A_N}(\theta)\xi_n\|_2^2 = \frac{\delta_{\omega_1(\rho), \theta} k}{\omega'_1(\rho)}$$

and

$$\lim_{N \rightarrow +\infty} \sum_{n=1}^{k+\ell} \|E_{U_N B_N U_N^*}(\tau)\xi_n\|_2^2 = \frac{\delta_{\omega_2(\rho), \tau} \ell}{\omega'_2(\rho)},$$

for $\theta, \tau \in \mathbb{R}$, where $\delta_{\omega_2(\rho), \tau}$ is the usual Kronecker symbol. Thus, assertion (4) is a strengthening of (3) in the special case $k\ell = 0$.

We give below two simple examples to illustrate the correspondence between spikes and outliers for unitarily invariant additive models, emphasizing the phenomenon (first noted without proof in [12], Remark 2.12, for finite rank perturbations) of a single spike generating multiple outliers.

EXAMPLE 2.3. The numerical simulation presented in Figure 1, due to Charles Bordenave, illustrates the appearance of two outliers arising from a single spike. We take $N = 1000$ and $X_N = A_N + U_N B_N U_N^*$, where $A_N = 2p - I_N$,

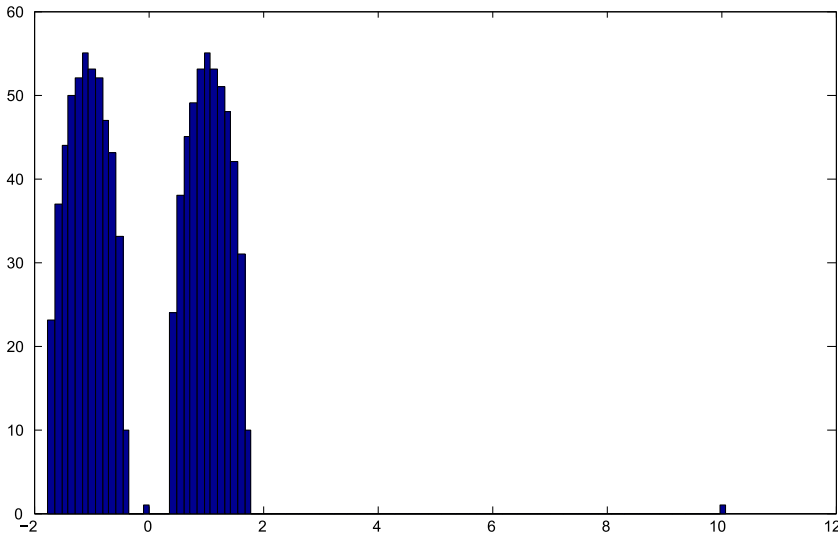


FIG. 1. The horizontal grid elements have a width of 2 units, starting at the left end with coordinate -2 . The two outliers appear at approximately -0.07420064427396 and 10.09926345874086 .

with p an orthogonal projection of rank $N/2 = 500$. The matrix B_N is given by the formula

$$B_N = \begin{bmatrix} \frac{W}{2\sqrt{N-1}} & 0_{(N-1) \times 1} \\ 0_{1 \times (N-1)} & 10 \end{bmatrix},$$

with W being sampled from a standard 999×999 GUE. A few elementary computations based on the results mentioned in Sections 3.1 and 3.4.1, as well as the fact that the R -transform of the standard Wigner distribution is $R(z) = z$ indicate that the support of the empirical eigenvalue distribution of X_N tends as $N \rightarrow \infty$ to $[-a, -b] \cup [b, a]$ where $a = \frac{\sqrt{18+2\sqrt{33}}}{4} + \frac{\sqrt{18+2\sqrt{33}}}{2+2\sqrt{33}} \simeq 1.760172593$, $b = \frac{\sqrt{18-2\sqrt{33}}}{4} + \frac{\sqrt{18-2\sqrt{33}}}{2-2\sqrt{33}} \simeq 0.369008729828$. Theorem 2.1 indicates the presence of two outliers at $\rho_1 = \frac{15}{2} - \frac{3\sqrt{11}}{4} - \frac{\sqrt{215+60\sqrt{11}}}{4} \simeq -0.07420064427396$ and $\rho_2 = \frac{15}{2} - \frac{3\sqrt{11}}{4} + \frac{\sqrt{215+60\sqrt{11}}}{4} \simeq 10.09926345874086$ (some computations/verifications were performed using Maple).

EXAMPLE 2.4. As noted above, the occurrence of multiple outliers generated by a single spike may happen also in the case of finite rank perturbations (see [12], Remark 2.12). In fact, asymptotically even countably many outliers may appear. The following example, due to Charles Bordenave, illustrates to a certain extent this phenomenon. In the graph from Figure 2, we present the cumulative distribution function for the empirical eigenvalue distribution of the matrix $U_N D_N U_N^* + 3e_1 e_1^*$, where $N = 2^{11} - 1$, $D_N = \text{Diag}(d_0, d_1, \dots, d_{10})$, with $d_j = 2^{-10+j+3} I_{2j}$, and e_1^* being the vector $(1, 0, \dots, 0) \in \mathbb{C}^N$. One outlier appears

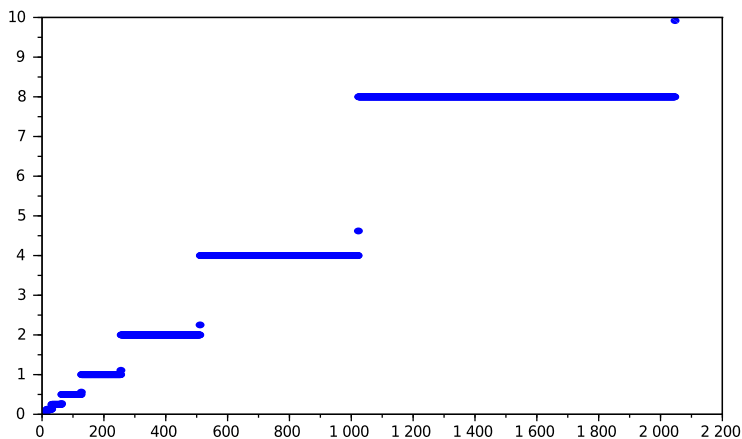


FIG. 2. This time, the vertical grid indicates the sizes of the eigenvalues.

in each interval of the complement in $\mathbb{R} \cup \{\infty\}$ of the support of μ_{D_N} , with the solution in the component containing infinity occurring, as expected, in $(0, +\infty)$. The reader is invited to imagine the case of $(d_0, d_1, \dots, d_{10})$ above being replaced by (d_0, \dots, d_m) , $m \rightarrow \infty$.

2.2. *Multiplicative perturbations of nonnegative matrices.* Here are the ingredients for constructing the multiplicative model $X_N = A_N^{1/2} U_N B_N U_N^* A_N^{1/2}$:

- Two compactly supported Borel probability measures μ and ν on $[0, +\infty)$.
- A positive integer p , and fixed positive numbers

$$\theta_1 \geq \theta_2 \geq \dots \geq \theta_p > 0$$

which do not belong to $\text{supp}(\mu)$.

- A sequence $(A_N)_{N \in \mathbb{N}}$ of deterministic nonnegative matrices of size $N \times N$ such that:
 - μ_{A_N} converges weakly to μ as $N \rightarrow \infty$;
 - for $N \geq p$ and $\theta \in \{\theta_1, \dots, \theta_p\}$, the sequence $\{\lambda_n(A_N)\}_{n=1}^N$ satisfies $\text{card}(\{n: \lambda_n(A_N) = \theta\}) = \text{card}(\{i: \theta_i = \theta\})$;
 - the eigenvalues of A_N which are not equal to some θ_i converge uniformly to $\text{supp}(\mu)$ as $N \rightarrow \infty$.
- A positive integer q , and fixed positive numbers

$$\tau_1 \geq \tau_2 \geq \dots \geq \tau_q > 0$$

which do not belong to $\text{supp}(\nu)$.

- A sequence $(B_N)_{N \in \mathbb{N}}$ of deterministic nonnegative matrices of size $N \times N$ such that:
 - μ_{B_N} converges weakly to ν as $N \rightarrow \infty$;
 - for $N \geq q$ and $\tau \in \{\tau_1, \dots, \tau_q\}$, the sequence $\{\lambda_n(B_N)\}_{n=1}^N$ satisfies $\text{card}(\{n: \lambda_n(B_N) = \tau\}) = \text{card}(\{j: \tau_j = \tau\})$;
 - the eigenvalues of B_N which are not equal to some τ_j converge uniformly to $\text{supp}(\nu)$ as $N \rightarrow \infty$.
- A sequence $(U_N)_{N \in \mathbb{N}}$ of unitary random matrices such that the distribution of U_N is the normalized Haar measure on the unitary group $U(N)$.

It is known from [39] that the asymptotic empirical eigenvalue distribution of X_N is $\mu \boxtimes \nu$. The projections P_N and Q_N used in the following statement were defined in (2.6).

THEOREM 2.5. *With the above notation, let ω_1, ω_2 be the subordination functions satisfying (2.4), set $K = \text{supp}(\mu \boxtimes \nu)$, $v_j(z) = \omega_j(1/z)$, $j = 1, 2$, and*

$$K' = K \cup \left[\bigcup_{i=1}^p v_1^{-1}(\{1/\theta_i\}) \right] \cup \left[\bigcup_{j=1}^q v_2^{-1}(\{1/\tau_j\}) \right].$$

(1) Given $\varepsilon > 0$, we have $\mathbb{P}(\exists N_0 \forall N, N > N_0, \sigma(X_N) \subset K'_\varepsilon) = 1$.

(2) Fix a positive number $\rho \in K' \setminus K$, let $\varepsilon > 0$ be such that $(\rho - 2\varepsilon, \rho + 2\varepsilon) \cap K' = \{\rho\}$ and set $k = \text{card}(\{i : v_1(\rho) = 1/\theta_i\})$, $\ell = \text{card}(\{j : v_2(\rho) = 1/\tau_j\})$. Then

$$\mathbb{P}(\exists N_0 \forall N, N > N_0, \text{card}(\{\sigma(X_N) \cap (\rho - \varepsilon, \rho + \varepsilon)\}) = k + \ell) = 1.$$

(3) With ρ and ε as in (2), we have

$$\lim_{N \rightarrow \infty} \left\| P_N E_{X_N}((\rho - \varepsilon, \rho + \varepsilon)) P_N - \frac{\rho \omega_1(1/\rho)}{\omega'_1(1/\rho)} E_{A_N}(\{1/\omega_1(1/\rho)\}) \right\| = 0$$

and

$$\begin{aligned} \lim_{N \rightarrow \infty} \left\| Q_N U_N E_{X_N}((\rho - \varepsilon, \rho + \varepsilon)) U_N^* Q_N \right. \\ \left. - \frac{\rho \omega_2(1/\rho)}{\omega'_2(1/\rho)} E_{B_N}(\{1/\omega_2(1/\rho)\}) \right\| = 0 \end{aligned}$$

almost surely.

(4) With ρ and ε as in (2), suppose in addition that $\ell = 0$. Then almost surely,

$$\begin{aligned} \lim_{N \rightarrow \infty} \sup \left\{ \left\| E_{A_N}(\{1/\omega_1(1/\rho)\}) \xi \right\|_2^2 - \frac{\rho \omega_1(1/\rho)}{\omega'_1(1/\rho)} \right\} : \\ \left. \xi \in E_{X_N}((\rho - \varepsilon, \rho + \varepsilon)) \mathbb{C}^N, \|\xi\|_2 \leq 1 \right\} = 0. \end{aligned}$$

Analogously, if $k = 0$, then almost surely

$$\begin{aligned} \lim_{N \rightarrow \infty} \sup \left\{ \left\| E_{U_N B_N U_N^*}(\{1/\omega_2(1/\rho)\}) \xi \right\|_2^2 - \frac{\rho \omega_2(1/\rho)}{\omega'_2(1/\rho)} \right\} : \\ \left. \xi \in E_{X_N}((\rho - \varepsilon, \rho + \varepsilon)) \mathbb{C}^N, \|\xi\|_2 \leq 1 \right\} = 0. \end{aligned}$$

REMARK 2.6. The analogue of Theorem 2.1 when $\mu = \delta_0$ was proved in [12] under the additional assumption that all eigenvalues of A except for the spikes are equal to zero. Our arguments below provide a proof of this result without this additional assumption. Similar observations apply to Theorem 2.5 when either μ or ν is a point mass. The only case in which one needs to be more careful is that of Theorem 2.5 for the positive half-line when μ or ν is equal to δ_0 . Suppose, for instance, that $\nu = \delta_0 \neq \mu$. The eigenvalues of $X_N = A_N^{1/2} U_N B_N U_N^* A_N^{1/2}$ are uniformly approximated arbitrarily well by the eigenvalues of

$$X_{N,\varepsilon} = A_N^{1/2} U_N (B_N + \varepsilon I_N) U_N^* A_N^{1/2} = X_N + \varepsilon A_N,$$

and our methods do apply to the perturbed model $X_{N,\varepsilon}$, whose asymptotic eigenvalue distribution is $\mu \boxtimes \delta_\varepsilon$. The outliers are calculated explicitly by noting that $\eta_{\mu \boxtimes \delta_\varepsilon}(z) = \eta_\mu(\varepsilon z)$, so $\omega_1(z) = \varepsilon z$, $\omega_2(z) = \eta_\mu(\varepsilon z)/\varepsilon$. Thus, $v_1(z) = \varepsilon/z$

and $v_2(z) = \eta_\mu(\varepsilon/z)/\varepsilon$. The outliers of $X_{N,\varepsilon}$ are the solutions of the equations $v_1(\rho) = 1/\theta_i, i = 1, \dots, p$ and $v_2(\rho) = 1/(\tau_j + \varepsilon), j = 1, \dots, q$. The first set of equations yields the outliers $\varepsilon\theta_i, i = 1, \dots, p$, while the second set of equations can be rewritten as

$$\rho = (\tau_j + \varepsilon) \left[\frac{\rho}{\varepsilon} \eta_\mu \left(\frac{\varepsilon}{\rho} \right) \right], \quad j = 1, \dots, q.$$

As $\varepsilon \rightarrow 0$, we conclude that the outliers of X_N are the numbers $\tau_j \gamma, j = 1, \dots, q$, where $\gamma = \eta'_\mu(0) = \int_0^\infty t d\mu(t)$ is the first moment of μ . If $\mu = \nu = \delta_0$, a similar argument shows that X_N has no outliers at all, that is, $\lim_{N \rightarrow \infty} \|X_N\| = 0$ almost surely.

2.3. *Multiplicative perturbations of unitary matrices.* Finally, we describe the ingredients for the construction of the multiplicative matrix model $X_N = A_N U_N B_N U_N^*$ with unitary A_N and B_N :

- Two Borel probability measures μ and ν on \mathbb{T} with nonzero first moments such that $\text{supp}(\mu \boxtimes \nu) \neq \mathbb{T}$.
- A positive integer p and fixed complex numbers $\theta_1, \dots, \theta_p \in \mathbb{T}$ which do not belong to $\text{supp}(\mu)$ and such that

$$2\pi > \arg \theta_1 \geq \dots \geq \arg \theta_p \geq 0.$$

- A sequence $(A_N)_{N \in \mathbb{N}}$ of deterministic unitary matrices of size $N \times N$ such that:
 - μ_{A_N} converges weakly to μ as $N \rightarrow \infty$;
 - for $N \geq p$ and $\theta \in \{\theta_1, \dots, \theta_p\}$, the sequence $\{\lambda_n(A_N)\}_{n=1}^N$ satisfies

$$\text{card}(\{n : \lambda_n(A_N) = \theta\}) = \text{card}(\{i : \theta_i = \theta\});$$

- the eigenvalues of A_N which are not equal to some θ_i converge uniformly to $\text{supp}(\mu)$ as $N \rightarrow \infty$.
- A positive integer q and fixed complex numbers $\tau_1, \dots, \tau_q \in \mathbb{T}$ which do not belong to $\text{supp}(\nu)$ and such that

$$2\pi > \arg \tau_1 \geq \dots \geq \arg \tau_q \geq 0.$$

- A sequence $(B_N)_{N \in \mathbb{N}}$ of deterministic unitary matrices of size $N \times N$ such that:
 - μ_{B_N} converges weakly to ν as $N \rightarrow \infty$;
 - for $N \geq q$ and $\tau \in \{\tau_1, \dots, \tau_q\}$, the sequence $\{\lambda_n(B_N)\}_{n=1}^N$ satisfies

$$\text{card}(\{n : \lambda_n(B_N) = \tau\}) = \text{card}(\{j : \tau_j = \tau\});$$

- the eigenvalues of B_N which are not equal to some τ_j converge uniformly to $\text{supp}(\nu)$ as $N \rightarrow \infty$.
- A sequence $(U_N)_{N \in \mathbb{N}}$ of unitary random matrices such that the distribution of U_N is the normalized Haar measure on the unitary group $U(N)$.

It is known from [39] that the asymptotic empirical eigenvalue distribution of X_N is $\mu \boxtimes \nu$. When $\rho \in \mathbb{T}$ and $\varepsilon > 0$, the interval $(\rho - \varepsilon, \rho + \varepsilon)$ consists of those numbers in \mathbb{T} whose argument differs from $\arg \rho$ by less than ε . With this convention, Theorem 2.5 holds *verbatim* in the unitary case as well.

REMARK 2.7. It is easy to see that our results hold equally well when A_N is random, independent of U_N , and has spikes $\theta_1(N), \dots, \theta_p(N)$ with the property that $\lim_{N \rightarrow \infty} \theta_i(N) = \theta_i, 1 \leq i \leq p$, almost surely. Similarly, B_N can be taken to be random, independent of A_N and U_N , and with spikes $\tau_1(N), \dots, \tau_q(N)$ that converge almost surely to τ_1, \dots, τ_q . The proofs use the general form of Propositions 5.1 and 5.7, respectively.

REMARK 2.8. The above remark allows us to treat sums or products of more than two spiked matrices. More precisely, let $k \geq 3$ be an integer, let $A_N^{(1)}, \dots, A_N^{(k)} \in M_N(\mathbb{C})$ be deterministic Hermitian matrices and let $U_N^{(1)}, \dots, U_N^{(k)} \in U(N)$ be independent Haar-distributed random matrices. Suppose that the eigenvalue distribution of $A_N^{(j)}$ tends weakly to μ_j and $A_N^{(j)}$ has spikes subject to the hypotheses of Section 2.1. Then $X_N^{(k)} = U_N^{(1)} A_N^{(1)} U_N^{(1)*} + \dots + U_N^{(k-1)} A_N^{(k-1)} U_N^{(k-1)*} + A_N^{(k)}$ has asymptotic eigenvalue distribution equal to $\mu_1 \boxplus \dots \boxplus \mu_k$, and the outliers in the spectrum of $X_N^{(k)}$ are described by an appropriate reformulation of Theorem 2.1. The result can be proved by induction on k if we observe that $X_N^{(k+1)}$ has the same distribution as $A_N^{(k+1)} + U_N B_N U_N^*$, where $B_N = X_N^{(k)}$ and U_N is a Haar-distributed unitary independent from $U_N^{(1)}, \dots, U_N^{(k-1)} \in U(N)$. A similar remark applies to Theorem 2.5 in the case of the circle. For the multiplicative model on $[0, +\infty)$, the corresponding generalization applies to models of the form $A_k^{1/2} U_{k-1} A_{k-1}^{1/2} \dots U_2 A_2^{1/2} U_1 A_1 U_1^* A_2^{1/2} U_2^* \dots A_{k-1}^{1/2} U_{k-1}^* A_k^{1/2}$.

REMARK 2.9. While we assume unitary invariance for our models, we would like to emphasize that our results apply to models for which the concentration results from Lemma 4.11, respectively, Corollary 4.12, the strong asymptotic freeness results of [21] and an asymptotic version of Lemma 4.7 hold.

3. Free convolutions. Free convolutions arise as natural analogues of classical convolutions in the context of free probability theory. For two Borel probability measures μ and ν on the real line, one defines the free additive convolution $\mu \boxplus \nu$ as the distribution of $a + b$, where a and b are free self-adjoint random variables with distributions μ and ν , respectively. Similarly, if both μ, ν are supported on $[0, +\infty)$ or on \mathbb{T} , their free multiplicative convolution $\mu \boxtimes \nu$ is the distribution of the product ab , where, as before, a and b are free, positive in the first case, unitary in the second, random variables with distributions μ and ν , respectively. The product ab of two free positive random variables is usually not positive, but it has

the same moments as the positive random variables $a^{1/2}ba^{1/2}$ and $b^{1/2}ab^{1/2}$. We refer to [42] for an introduction to free probability theory and to [13, 37, 38] for the definitions and main properties of free convolutions. In this section, we recall the analytic approach developed in [37, 38] to calculate the free convolutions of measures, as well as the analytic subordination property [14, 40, 41] and related results.

3.1. *Additive free convolution.* Recall from (2.1) the definition of the Cauchy–Stieltjes transform of a finite positive Borel measure μ on the real line:

$$G_\mu(z) = \int_{\mathbb{R}} \frac{1}{z - t} d\mu(t), \quad z \in \mathbb{C} \setminus \text{supp}(\mu).$$

This function maps \mathbb{C}^+ to \mathbb{C}^- and $\lim_{y \uparrow +\infty} iyG_\mu(iy) = \mu(\mathbb{R})$. Conversely, any analytic function $G : \mathbb{C}^+ \rightarrow \mathbb{C}^-$ for which $\lim_{y \uparrow +\infty} iyG(iy)$ is finite is of the form $G = G_\mu|_{\mathbb{C}^+}$ for some finite positive Borel measure μ on \mathbb{R} . When μ has compact support, the function G_μ is also analytic at ∞ and $G_\mu(\infty) = 0$ (see [1], Chapter 3, for these results). The measure μ can be recovered from its Cauchy–Stieltjes transform as a weak limit

$$(3.1) \quad d\mu(x) = \lim_{y \downarrow 0} \frac{-1}{\pi} \Im G_\mu(x + iy) dx.$$

This is the Stieltjes inversion formula and it holds for signed measures as well. The density of (the absolutely continuous part of) μ relative to Lebesgue measure is calculated as

$$\frac{d\mu}{dx}(x) = \lim_{y \downarrow 0} \frac{-1}{\pi} \Im G_\mu(x + iy)$$

for almost every x relative to the Lebesgue measure. In particular, $\mathbb{R} \setminus \text{supp}(\mu)$ can be described as the set of those points $x \in \mathbb{R}$ with the property that $G_\mu|_{\mathbb{C}^+}$ can be continued analytically to an open interval $I \ni x$ such that $G_\mu|_I$ is real-valued. On the other hand,

$$\lim_{y \downarrow 0} \frac{-1}{\pi} \Im G_\mu(x + iy) = +\infty$$

almost everywhere relative to the singular part of μ . Indeed, these facts follow from the straightforward observation that $(-\pi)^{-1} \Im G_\mu(x + iy)$, $y > 0$, is the Poisson integral of μ . See [25] for these aspects of harmonic analysis.

It is often convenient to work with the reciprocal Cauchy–Stieltjes transform $F_\mu(z) = 1/G_\mu(z)$, which defines an analytic self-map of the upper half-plane. This function enjoys the following properties:

- (a) For any $z \in \mathbb{C}^+$, $\Im F_\mu(z) \geq \mu(\mathbb{R})^{-1} \Im z$. If equality holds at one point of \mathbb{C}^+ , then it holds at all points, and $\mu = \mu(\mathbb{R}) \delta_{-\mu(\mathbb{R}) \Re F_\mu(i)}$.

(b) In particular, the function

$$(3.2) \quad h_\mu(z) = F_\mu(z) - \mu(\mathbb{R})^{-1}z, \quad z \in \mathbb{C}^+,$$

is a self-map of \mathbb{C}^+ unless μ is a point mass, in which case h_μ is a real constant.

(c) If μ is compactly supported, there exist a real number α and a finite positive Borel measure ρ on \mathbb{R} with $\text{supp}(\rho)$ included in the convex hull of $\text{supp}(\mu)$ such that

$$(3.3) \quad F_\mu(z) = \alpha + \mu(\mathbb{R})^{-1}z + \int_{\mathbb{R}} \frac{1}{t-z} d\rho(t), \quad z \in \mathbb{C} \setminus \text{supp}(\rho).$$

Conversely, if $F: \mathbb{C}^+ \rightarrow \mathbb{C}^+$ extends to an analytic real-valued function to the complement in \mathbb{R} of a compact set, and if $\lim_{y \rightarrow +\infty} F(iy) = \infty$, then there exists a compactly supported positive Borel measure μ on \mathbb{R} satisfying $F = F_\mu$. The value $\mu(\mathbb{R})$ is determined by $\mu(\mathbb{R}) = \lim_{y \rightarrow +\infty} iy/F(iy)$.

(d) If $\mu(\mathbb{R}) = 1$ and ρ is as in (3.3), then $\rho(\mathbb{R}) = \int_{\mathbb{R}} (t - \int_{\mathbb{R}} s d\mu(s))^2 d\mu(t)$ and $\alpha = -\int_{\mathbb{R}} t d\mu(t)$.

Equation (3.3) is a special case of the Nevanlinna representation of analytic self-maps of the upper half-plane ([1], Chapter 3):

$$(3.4) \quad F(z) = a + bz + \int_{\mathbb{R}} \frac{1 + tz}{t - z} d\Omega(t), \quad z \in \mathbb{C}^+,$$

where $a \in \mathbb{R}, b \geq 0$ and Ω is a finite positive Borel measure on \mathbb{R} . We identify $a = \Re F(i)$, $b = \lim_{y \rightarrow +\infty} F(iy)/iy$, $\Omega(\mathbb{R}) = \Im F(i) - b$. If $\int_{\mathbb{R}} t^2 d\Omega(t) < +\infty$, then (3.4) reduces to (3.3), with $b = \mu(\mathbb{R})^{-1}$ and $d\rho(t) = (1 + t^2) d\Omega(t), \alpha = a - \int_{\mathbb{R}} t d\Omega(t)$.

The Cauchy–Stieltjes transform of a compactly supported probability measure μ is conformal in the neighborhood of ∞ , and its functional inverse G_μ^{-1} is meromorphic at zero with principal part $1/z$. The R -transform [37] of μ is the convergent power series defined by

$$R_\mu(z) = G_\mu^{-1}(z) - \frac{1}{z}.$$

The free additive convolution of two compactly supported probability measures μ and ν is another compactly supported probability measure characterized by the identity

$$R_{\mu \boxplus \nu} = R_\mu + R_\nu$$

satisfied by these convergent power series.

3.2. *Multiplicative free convolution on $[0, +\infty)$.* Recall from (2.2) the definition of the moment-generating function of a Borel probability measure μ on $[0, +\infty)$:

$$\psi_\mu(z) = \int_{[0,+\infty)} \frac{zt}{1-zt} d\mu(t), \quad z \in \mathbb{C} \setminus \left\{ z \in \mathbb{C} : \frac{1}{z} \in \text{supp}(\mu) \right\}.$$

This function is related to the Cauchy–Stieltjes transform of μ via the relation

$$\psi_\mu(z) = \frac{1}{z} G_\mu\left(\frac{1}{z}\right) - 1.$$

It satisfies the following properties, for which we refer to [13], Section 6:

- $\psi_\mu(\mathbb{C}^+) \subseteq \mathbb{C}^+$.
- $\psi_\mu((-\infty, 0)) \subseteq (\mu(\{0\}) - 1, 0)$ and

$$\psi_\mu(i\mathbb{C}^+) \subseteq \left\{ z \in \mathbb{C} : \left| z - \frac{\mu(\{0\}) - 1}{2} \right| < \frac{1 - \mu(\{0\})}{2} \right\}.$$

In addition,

$$\lim_{x \downarrow -\infty} \psi_\mu(x) = \mu(\{0\}) - 1, \quad \lim_{x \uparrow 0} \psi_\mu(x) = 0,$$

$$\lim_{x \uparrow 0} \psi'_\mu(x) = \int_{[0,+\infty)} t d\mu(t).$$

- In particular, if $\text{supp}(\mu)$ is compact and not equal to $\{0\}$, then ψ_μ is injective on some neighborhood of zero in \mathbb{C} .
- ψ_μ is injective on $i\mathbb{C}^+$.

It is often convenient to work with the eta transform, or Boolean cumulant function,

$$\eta_\mu(z) = \frac{\psi_\mu(z)}{1 + \psi_\mu(z)}.$$

It inherits from ψ the following properties:

(a) $\pi > \arg \eta_\mu(z) \geq \arg z$ for all $z \in \mathbb{C}^+$, where \arg takes values in $(0, \pi)$ on \mathbb{C}^+ . Moreover, if equality holds for one point in \mathbb{C}^+ , it holds for all points in \mathbb{C}^+ , and $\mu = \delta_{\eta_\mu(z)/z} = \delta_{\eta'_\mu(0)}$ for any $z \in \mathbb{C}^+$.

(b) $\lim_{x \uparrow 0} \eta_\mu(x) = 0$ and $\lim_{x \uparrow 0} \eta'_\mu(x) = \int_{[0,+\infty)} t d\mu(t)$. In particular, if $\text{supp}(\mu)$ is compact and different from $\{0\}$, then η_μ is injective on some neighborhood of zero in \mathbb{C} .

(c) If $\mu \neq \delta_0$, η_μ is strictly increasing from $(-\infty, 0]$ to $(\mu(\{0\})^{-1}(\mu(\{0\}) - 1), 0]$, where $\mu(\{0\})^{-1}(\mu(\{0\}) - 1)$ should be replaced by $-\infty$ if $\mu(\{0\}) = 0$. Moreover, η_μ is injective on $i\mathbb{C}^+$.

(d) Conversely, if an analytic function $\eta: \mathbb{C}^+ \rightarrow \mathbb{C}^+$ satisfies $\pi > \arg \eta(z) \geq \arg z$ for all $z \in \mathbb{C}^+$ and $\lim_{x \uparrow 0} \eta(x) = 0$, then $\eta = \eta_\mu$ for some Borel probability measure on $[0, +\infty)$ ([9], Proposition 2.2).

The Σ -transform [13, 38] of a compactly supported Borel probability measure $\mu \neq \delta_0$ is the convergent power series defined by

$$\Sigma_\mu(z) = \frac{\eta_\mu^{-1}(z)}{z},$$

where η_μ^{-1} is the inverse of η_μ relative to composition. The free multiplicative convolution of two compactly supported probability measures $\mu \neq \delta_0 \neq \nu$ is another compactly supported probability measure characterized by the identity

$$\Sigma_{\mu \boxtimes \nu}(z) = \Sigma_\mu(z) \Sigma_\nu(z)$$

in a neighborhood of 0.

3.3. *Multiplicative free convolution on \mathbb{T} .* The analytic transforms involved in the study of multiplicative free convolution on \mathbb{T} are formally the same ones as in Section 3.2, but their analytical properties are different. Thus,

$$\psi_\mu(z) = \int_{\mathbb{T}} \frac{zt}{1-zt} d\mu(t), \quad z \in \mathbb{C} \setminus \left\{ z \in \mathbb{C} : \frac{1}{z} \in \text{supp}(\mu) \right\}.$$

It satisfies $\Re \psi_\mu(z) > -\frac{1}{2}$ for all $|z| < 1$. We work almost exclusively with the eta transform, or Boolean cumulant function,

$$\eta_\mu(z) = \frac{\psi_\mu(z)}{1 + \psi_\mu(z)}.$$

The following properties of η_μ are relevant to our study:

(a) For any $z \in \mathbb{D}$, we have $|\eta_\mu(z)| \leq |z|$. If equality holds at one point in $\mathbb{D} \setminus \{0\}$, it holds at all points in \mathbb{D} , and $\mu = \delta_{\eta_\mu(z)/z} = \delta_{\eta'_\mu(0)}$ for any $z \in \mathbb{D} \setminus \{0\}$.

(b) $\eta_\mu(0) = 0$ and $\eta'_\mu(0) = \int_{\mathbb{T}} t d\mu(t)$. In particular η_μ is injective on a neighborhood of zero in \mathbb{C} if and only if $\int_{\mathbb{T}} t d\mu(t) \neq 0$.

(c) The function η_μ continues via Schwarz reflection through the set $\{z \in \mathbb{T} : \bar{z} \notin \text{supp}(\mu)\}$, that is,

$$\eta_\mu(z) = \frac{1}{\eta_\mu(\frac{1}{\bar{z}})}, \quad |z| > 1.$$

(d) For almost all points $1/x$ with respect to the absolutely continuous part of μ (relative to the Haar measure on \mathbb{T}), the nontangential limit of η_μ at x [denoted by $\triangleleft \lim_{z \rightarrow x} \eta_\mu(z)$] belongs to \mathbb{D} , and for almost all points $1/x$ in the complement of the support of the absolutely continuous part of μ , the nontangential limit of η_μ at x belongs to \mathbb{T} . Moreover, if μ has a singular component, then for almost all points $1/x$ with respect to this component, the nontangential limit of η_μ at x equals one.

(e) Conversely, if an analytic function $\eta: \mathbb{D} \rightarrow \mathbb{D}$ satisfies $\eta(0) = 0$, then $\eta = \eta_\mu$ for a unique Borel probability measure on \mathbb{T} ([9], Proposition 3.2).

When $\int_{\mathbb{T}} t d\mu(t) \neq 0$, we define the Σ -transform [13, 38] of μ as the convergent power series

$$\Sigma_\mu(z) = \frac{\eta_\mu^{-1}(z)}{z}.$$

Again, the free multiplicative convolution of two probability measures μ and ν with nonzero first moments is another probability measure with nonzero first moment characterized by the identity

$$\Sigma_{\mu \boxtimes \nu}(z) = \Sigma_\mu(z) \Sigma_\nu(z)$$

in a neighborhood of 0.

If both of μ and ν have zero first moment, then $\mu \boxtimes \nu$ is the Haar (uniform) distribution on \mathbb{T} ; see [42]. From now on, we always assume that all our probability measures on \mathbb{T} have nonzero first moments.

3.4. *Analytic subordination.* The analytic subordination phenomenon for free convolutions, as seen in (2.3) and (2.4), was first noted by Voiculescu in [40] for free additive convolution of compactly supported probability measures. Later, Biane [14] extended the result to free additive convolutions of arbitrary probability measures on \mathbb{R} , and also found a subordination result for multiplicative free convolution. More importantly, he proved the stronger result (see heuristics in the Introduction) that the conditional expectation of the resolvent of a sum or product of free random variables onto the algebra generated by one of them is in fact also a resolvent. In [41], Voiculescu deduced this property from the fact that such a conditional expectation is a coalgebra morphism for certain coalgebras, and through this observation he extended the subordination property to free convolutions of operator-valued distributions. For our purposes, considerably less than that is required: we essentially only use the complex analytic properties of these functions.

3.4.1. *The subordination functions equations for free additive convolution.* Given Borel probability measures μ and ν on \mathbb{R} , there exist two unique analytic functions $\omega_1, \omega_2: \mathbb{C}^+ \rightarrow \mathbb{C}^+$ such that:

- (1) $\lim_{y \rightarrow +\infty} \omega_j(iy)/iy = 1, j = 1, 2.$
- (2) For every $z \in \mathbb{C}^+$, we have

$$(3.5) \quad \omega_1(z) + \omega_2(z) - z = F_\mu(\omega_1(z)) = F_\nu(\omega_2(z)) = F_{\mu \boxplus \nu}(z).$$

(3) In particular (see [10]), for any $z \in \mathbb{C}^+ \cup \mathbb{R}$ such that ω_1 is analytic at z , $\omega_1(z)$ is the attracting fixed point of the self-map of \mathbb{C}^+ defined by

$$w \mapsto F_\nu(F_\mu(w) - w + z) - (F_\mu(w) - w).$$

A similar statement, with μ, ν interchanged, holds for ω_2 .

We note that (3.5) implies that the functions ω_1, ω_2 continue analytically across an interval $(\alpha, \beta) \subseteq \mathbb{R}$ such that $\omega_1|_{(\alpha, \beta)}$ and $\omega_2|_{(\alpha, \beta)}$ are real-valued if and only if the same is true for $F_{\mu \boxplus \nu}$. For the sake of providing an intrinsic characterization of the correspondence between spikes and outliers, we formalize and slightly strengthen this remark in the following lemma. Here, we use the functions h_μ, h_ν defined by (3.2).

LEMMA 3.1. *Consider two compactly supported Borel probability measures μ and ν , neither of them a point mass. Then the subordination functions ω_1 and ω_2 have extensions to $\mathbb{C}^+ \cup \{\infty\}$ with the following properties:*

- (a) $\omega_1, \omega_2 : \overline{\mathbb{C}^+} \cup \{\infty\} \rightarrow \overline{\mathbb{C}^+} \cup \{\infty\}$ are continuous.
- (b) If $x \in \mathbb{R} \setminus \text{supp}(\mu \boxplus \nu)$, then the functions ω_1 and ω_2 continue meromorphically to a neighborhood of x , $\omega_1(x) = h_\nu(\omega_2(x)) + x \in (\mathbb{R} \cup \{\infty\}) \setminus \text{supp}(\mu)$, and $\omega_2(x) = h_\mu(\omega_1(x)) + x \in (\mathbb{R} \cup \{\infty\}) \setminus \text{supp}(\nu)$. If $\omega_1(x) = \infty$, then $\omega_2(x) = x - \int_{\mathbb{R}} t d\mu(t) \in \mathbb{R}$, and if $\omega_2(x) = \infty$, then $\omega_1(x) = x - \int_{\mathbb{R}} t d\nu(t) \in \mathbb{R}$.
- (c) Conversely, suppose that ω_1 continues meromorphically to a neighborhood of a point $x \in \mathbb{R}$ and that $\omega_1(y) \in \mathbb{R}$ when $y \in (x - \delta, x + \delta) \setminus \{x\}$ for some $\delta > 0$. If $x \in \text{supp}(\mu \boxplus \nu)$, then x is an isolated atom for $\mu \boxplus \nu$.

In the context of Part (b) of the above lemma, we note that h_μ is analytic around infinity, and $h_\mu(\infty) = -\int_{\mathbb{R}} t d\mu(t)$.

PROOF OF LEMMA 3.1. Part (a) was proved in [8], Theorem 3.3. Fix $x \in \mathbb{R} \setminus \text{supp}(\mu \boxplus \nu)$. Equation (3.5) indicates that ω_1 and ω_2 must take real values on $(x - \delta, x + \delta) \setminus \{x\}$ for some $\delta > 0$. Schwarz reflection implies that ω_1 and ω_2 have meromorphic continuations with real values on \mathbb{R} across the corresponding intervals.

The relation $G_{\mu \boxplus \nu}(z) = G_\mu(\omega_1(z))$ shows that the limit $\lim_{z \rightarrow y} G_\mu(\omega_1(z))$ is real for $y \in (x - \delta, x + \delta)$ and therefore $\mu(\{\omega_1(y) : y \in (x - \delta, x + \delta)\}) = 0$ by the Stieltjes inversion formula. In particular, $\omega_1(x) \notin \text{supp}(\mu)$. To conclude the proof of (b), suppose that $\omega_1(x) = \infty$. It follows from (3.5) in conjunction with items (c) and (d) of Section 3.1 that

$$\omega_2(x) = \lim_{z \rightarrow x} F_\mu(\omega_1(z)) - \omega_1(z) + z = x + \lim_{w \rightarrow \infty} F_\mu(w) - w = x - \int_{\mathbb{R}} t d\mu(t),$$

such that ω_2 is analytic at x . The statement for $\omega_2(x) = \infty$ follows by symmetry.

Finally, suppose that the hypotheses of (c) are satisfied. It was observed in [8] that $\omega_2(y)$ is also real for $y \in (x - \delta, x + \delta)$. [Indeed, if $\Im \omega_2(y) > 0$, relation (3.5) implies

$$(3.6) \quad \omega_1(y) + \omega_2(y) = y + F_{\mu \boxplus \nu}(y) = y + F_\nu(\omega_2(y))$$

and, therefore, $\Im F_\nu(\omega_2(y)) = \Im \omega_2(y)$. This relation can only hold when ν is a point mass, a case which we excluded.] Now, (3.6) implies that $F_{\mu \boxplus \nu}$ is continuous and real-valued on $(x - \delta, x + \delta) \setminus \{x\}$, and this yields the desired conclusion via the Stieltjes inversion formula. \square

3.4.2. *The subordination functions equations for multiplicative free convolution on $[0, +\infty)$.* Given Borel probability measures μ, ν on $[0, +\infty)$, there exist two unique analytic functions $\omega_1, \omega_2 : \mathbb{C} \setminus [0, +\infty) \rightarrow \mathbb{C} \setminus [0, +\infty)$ with the following properties:

- (1) $\pi > \arg \omega_j(z) \geq \arg z$ for $z \in \mathbb{C}^+$ and $j = 1, 2$.
- (2) For every $z \in \mathbb{C} \setminus [0, +\infty)$, we have

$$(3.7) \quad \frac{\omega_1(z)\omega_2(z)}{z} = \eta_\mu(\omega_1(z)) = \eta_\nu(\omega_2(z)) = \eta_{\mu \boxtimes \nu}(z).$$

(3) In particular (see [10]), for any $z \in \mathbb{C}^+ \cup \mathbb{R}$ such that ω_1 is analytic at z , the point $h_1(z) := \omega_1(z)/z$ is the attracting fixed point of the self-map of $\mathbb{C} \setminus [0, +\infty)$ defined by

$$w \mapsto \frac{w}{\eta_\mu(zw)} \eta_\nu\left(\frac{\eta_\mu(zw)}{w}\right).$$

A similar statement, with μ, ν interchanged, holds for ω_2 .

A version of Lemma 3.1 holds for multiplicative free convolution on $[0, +\infty)$. Since the proof is similar to the proof of Lemma 3.1 and of Lemma 3.3 below, we omit it. Item (a) appears in the proof of [7], Theorem 3.2.

LEMMA 3.2. *Consider two compactly supported Borel probability measures μ, ν on $[0, +\infty)$, neither of them a point mass. Then the restrictions of the subordination functions ω_1 and ω_2 to \mathbb{C}^+ have extensions to $\overline{\mathbb{C}^+} \cup \{\infty\}$ with the following properties:*

- (a) $\omega_1, \omega_2 : \overline{\mathbb{C}^+} \cup \{\infty\} \rightarrow \overline{\mathbb{C}^+} \cup \{\infty\}$ are continuous.
- (b) If $1/x \in \mathbb{R} \setminus \text{supp}(\mu \boxtimes \nu)$ then the functions ω_1 and ω_2 continue analytically to a neighborhood of x , $1/\omega_1(x) = \omega_2(x)/x\eta_\nu(\omega_2(x)) \in \mathbb{R} \setminus \text{supp}(\mu)$, and $1/\omega_2(x) = \omega_1(x)/x\eta_\mu(\omega_1(x)) \in \mathbb{R} \setminus \text{supp}(\nu)$.

3.4.3. *The subordination functions equations for multiplicative free convolution on \mathbb{T} .* Given Borel probability measures μ, ν on \mathbb{T} with nonzero first moments, there exist unique analytic functions $\omega_1, \omega_2 : \mathbb{D} \rightarrow \mathbb{D}$ such that:

- (1) $|\omega_j(z)| \leq |z|$, $z \in \mathbb{D}$, $j = 1, 2$.
- (2) For every $z \in \mathbb{D}$, we have

$$(3.8) \quad \frac{\omega_1(z)\omega_2(z)}{z} = \eta_\mu(\omega_1(z)) = \eta_\nu(\omega_2(z)) = \eta_{\mu \boxtimes \nu}(z).$$

(3) In particular (see [10]), if $z \in \mathbb{D} \cup \mathbb{T}$ and ω_1 is analytic at z , then the point $h_1(z) := \omega_1(z)/z$ is the attracting fixed point of the self-map of \mathbb{D}

$$\mathbb{D} \ni w \mapsto \frac{w}{\eta_\mu(zw)} \eta_\nu\left(\frac{\eta_\mu(zw)}{w}\right) \in \mathbb{D}.$$

A similar statement, with μ, ν interchanged, holds for ω_2 .

LEMMA 3.3. Consider two Borel probability measures μ, ν on \mathbb{T} with nonzero first moments, neither of them a point mass. Suppose that $\mathbb{T} \setminus \text{supp}(\mu \boxtimes \nu) \neq \emptyset$. Then the subordination functions ω_1 and ω_2 have extensions to \mathbb{T} with the following properties:

(a) $\omega_1, \omega_2 : \mathbb{D} \cup \mathbb{T} \rightarrow \mathbb{D} \cup \mathbb{T}$ are continuous.

(b) If $1/x \in \mathbb{T} \setminus \text{supp}(\mu \boxtimes \nu)$ then the functions ω_1 and ω_2 continue analytically to a neighborhood of x , $1/\omega_1(x) = \omega_2(x)/x\eta_\nu(\omega_2(x)) \in \mathbb{T} \setminus \text{supp}(\mu)$, and $1/\omega_2(x) = \omega_1(x)/x\eta_\mu(\omega_1(x)) \in \mathbb{T} \setminus \text{supp}(\nu)$.

PROOF. Part (a) can be found in [7], Theorem 3.6. Fix $1/x \in \mathbb{T} \setminus \text{supp}(\mu \boxtimes \nu)$. Equation (3.8) coupled with items (d) and (e) of Section 3.3 indicate clearly that ω_1, ω_2 must take values in \mathbb{T} at least a.e. on a neighborhood of x . As proved in [7], Proposition 1.9(a), if, say, ω_1 does not reflect analytically through a neighborhood of x , then for any $\varepsilon > 0$ the set of nontangential limits $\{\triangleleft \lim_{z \rightarrow c} \omega_1(z) : \arg(xe^{-i\varepsilon}) < \arg(c) < \arg(xe^{i\varepsilon})\}$ of ω_1 around x is dense in \mathbb{T} . As $\mathbb{T} \setminus \text{supp}(\mu)$ is nonempty, many of these limits will fall in the domain of analyticity of η_μ . In particular, we may choose an arbitrary interval $I = \{e^{it} : t \in [s_1, s_2]\}$ strictly included in the domain of analyticity of η_μ and we will be able to find points $1/c_n \in \mathbb{T} \setminus \text{supp}(\mu \boxtimes \nu)$ tending to $1/x$ so that $\triangleleft \lim_{z \rightarrow c_n} \omega_1(z) = d_n \in I$. Obviously, in that case any limit point of $\{d_n\}_{n \in \mathbb{N}}$ will still belong to I , and hence be in the domain of analyticity of η_μ . Pick such a limit point d_0 . Note that, as a trivial consequence of the Julia–Carathéodory theorem ([25], Chapter I, Exercises 6 and 7), $\eta'_\mu(w) > 0$ for any $w \in \mathbb{T}$ in the domain of analyticity of η_μ , and thus $\eta'_\mu(d_0) > 0$, which implies that η_μ is conformal on a neighborhood U of d_0 (in \mathbb{C}). Now recall that

$$\eta_{\mu \boxtimes \nu}(c_n) = \triangleleft \lim_{z \rightarrow c_n} \eta_{\mu \boxtimes \nu}(z) = \triangleleft \lim_{z \rightarrow c_n} \eta_\mu(\omega_1(z)) = \eta_\mu(d_n).$$

Letting n go to infinity (along a subsequence, if necessary), and recalling that $1/x \notin \text{supp}(\mu \boxtimes \nu)$, we obtain $\eta_{\mu \boxtimes \nu}(x) = \eta_\mu(d_0)$. Both functions are analytic around the two respective points from \mathbb{T} , so the conformality of η_μ on U allows us to find $\eta_\mu(U)$ as a neighborhood of $\eta_{\mu \boxtimes \nu}(x)$ on which the compositional inverse η_μ^{-1} can be defined. We write $\eta_\mu^{-1} \circ \eta_{\mu \boxtimes \nu}$ on some convex neighborhood W of x which is small enough so $\eta_{\mu \boxtimes \nu}(W) \subset \eta_\mu(U)$ (the existence of W is guaranteed by the continuity of $\eta_{\mu \boxtimes \nu}$ around x). Pick points $z_n \in \mathbb{D}$ such that $|z_n - c_n| < \frac{1}{n}$. Clearly, $\lim_{n \rightarrow \infty} z_n = x$, so that for all $n \in \mathbb{N}$ large enough, $z_n \in W$. Pick a piecewise linear path going consecutively through the z_n 's, so, by the convexity of W , this path stays in W and necessarily converges to x . For any z in this path, we have

$$\eta_{\mu \boxtimes \nu}(z) = \eta_\mu(\omega_1(z)) \implies \omega_1(z) = (\eta_\mu^{-1} \circ \eta_{\mu \boxtimes \nu})(z),$$

which, by analytic continuation and analyticity of $\eta_\mu^{-1} \circ \eta_{\mu \boxtimes \nu}$ on W , forces ω_1 to be analytic on W , providing a contradiction. Thus, ω_1 extends analytically

through x . The same argument shows that ω_2 extends analytically around x . The last statement of (b) above follows again from the simple remark that $\omega_2(x) = x\eta_\mu(\omega_1(x))/\omega_1(x)$. \square

Unlike the case of free additive convolution, the functions ω_1 and ω_2 are bounded on \mathbb{D} , and hence do not have pole singularities on \mathbb{T} .

4. Preliminary results. The proofs of our main results will be based largely on both scalar- and matrix-valued analytic function methods, as well as on some elementary results from operator theory. We start by collecting some results which apply to both additive and multiplicative models. We use the notation introduced in Section 2.

4.1. *Boundary behavior and convergence of some families of analytic functions.* The following convergence result for sequences of Nevanlinna-type functions is necessary in the analysis of eigenvectors corresponding to outliers. $C(\mathbf{X})$ denotes the space of complex-valued continuous functions on a topological space \mathbf{X} . We use the notation $|\rho|$ for the total variation measure of a signed Borel measure ρ on \mathbb{R} . That is, $|\rho| = \rho^+ + \rho^-$, where $\rho = \rho^+ - \rho^-$ is the Hahn decomposition of ρ . The total variation of ρ is $\|\rho\| = |\rho|(\mathbb{R})$.

LEMMA 4.1. *Let $\{\rho_N\}_{N \in \mathbb{N}}$ be a sequence of signed Borel measures on \mathbb{R} satisfying the following properties:*

- *There exists $m \in \mathbb{R}$ such that $\text{supp}(\rho_N) \subseteq [-m, m]$ for all $N \in \mathbb{N}$;*
- *$\rho_N \rightarrow 0$ in the weak*-topology as $N \rightarrow \infty$, that is,*

$$\lim_{N \rightarrow \infty} \int_{\mathbb{R}} f(t) d\rho_N(t) = 0, \quad f \in C([-m, m]).$$

Then there exists a sequence $\{v_N\}_{N \in \mathbb{N}} \subset [0, +\infty)$ converging to zero, independent of z , such that

$$\left| \int_{\mathbb{R}} \frac{1}{z - t} d\rho_N(t) \right| < \left(1 + \frac{1}{(\Im z)^2} \right) v_N, \quad z \in \mathbb{C}^+, N \in \mathbb{N}.$$

PROOF. The number $M = \sup_{N \in \mathbb{N}} \|\rho_N\|$ is finite by the uniform boundedness theorem. Suppose, to get a contradiction, that the conclusion of the lemma does not hold. Passing if necessary to a subsequence, we deduce the existence of $v > 0$ and of numbers $z_N \in \mathbb{C}^+$ such that

$$v \left(1 + \frac{1}{(\Im z_N)^2} \right) \leq \left| \int_{\mathbb{R}} \frac{1}{z_N - t} d\rho_N(t) \right|, \quad N \in \mathbb{N}.$$

The inequality

$$\left| \int_{\mathbb{R}} \frac{1}{z_N - t} d\rho_N(t) \right| \leq M \text{dist}(z_N, [-m, m])^{-1} \leq \frac{M}{\Im z_N}$$

implies that the sequence $\{|z_N|\}_{N \in \mathbb{N}}$ is bounded (by $m + M/v$) and the sequence $\{\Im z_N\}_{N \geq 0}$ is bounded away from zero (by v/M). Passing to a further subsequence, we may assume that the limit $w = \lim_{N \rightarrow \infty} z_N$ exists in \mathbb{C}^+ . The uniform boundedness of ρ_N shows that

$$\lim_{N \rightarrow \infty} \left| \int_{\mathbb{R}} \frac{1}{z_N - t} d\rho_N(t) - \int_{\mathbb{R}} \frac{1}{w - t} d\rho_N(t) \right| = 0.$$

We conclude that the numbers $\int_{\mathbb{R}} (1/(w - t)) d\rho_N(t)$ do not converge to zero, contrary to the hypothesis. The lemma follows. \square

An analogous result holds for \mathbb{T} .

LEMMA 4.2. *Let $\{\omega_N\}_{N \in \mathbb{N}}$ be a sequence of analytic self-maps of the unit disc such that the limit $\omega(z) = \lim_{N \rightarrow \infty} \omega_N(z)$ exists for all $z \in \mathbb{D}$. Then there exists a sequence $\{v_N\}_{N \in \mathbb{N}} \subset [0, +\infty)$ converging to zero, independent of z , such that*

$$|\omega_N(z) - \omega(z)| < \frac{v_N}{1 - |z|}, \quad z \in \mathbb{D}, N \in \mathbb{N}.$$

PROOF. Suppose, to get a contradiction, that the conclusion of the lemma is not true. Passing if necessary to a subsequence, we deduce the existence of $v > 0$ and of a sequence $\{z_N\}_{N \in \mathbb{N}} \subset \mathbb{D}$ such that

$$(4.1) \quad v \leq \frac{v}{1 - |z_N|} \leq |\omega_N(z_N) - \omega(z_N)|, \quad N \in \mathbb{N}.$$

Combining this with the obvious inequality

$$|\omega_N(z) - \omega(z)| \leq 2, \quad z \in \mathbb{D},$$

we see that the sequence $\{z_N\}_{N \in \mathbb{N}}$ is contained in a compact subset K of \mathbb{D} . Montel's theorem implies that the convergence of ω_N to ω is uniform on K , and this clearly contradicts (4.1). The lemma follows. \square

The following lemma from [18], Appendix, is proved using ideas from [28]. We use the notation $\mathcal{D}(\mathbb{R})$ for the space of infinitely differentiable, compactly supported functions $\varphi : \mathbb{R} \rightarrow \mathbb{C}$.

LEMMA 4.3. *Let Δ be an analytic function on $\mathbb{C} \setminus \mathbb{R}$ which satisfies*

$$|\Delta(z)| \leq (|z| + K)^\alpha P(|\Im z|^{-1})$$

for some numbers $\alpha \geq 1, K \geq 0$, and polynomial P with nonnegative coefficients. Then for every $h \in \mathcal{D}(\mathbb{R})$ there exists a constant $c > 0$ depending only on α, K, P, h such that

$$\limsup_{y \rightarrow 0^+} \left| \int_{\mathbb{R}} h(x) \Delta(x + iy) dx \right| \leq c.$$

We also record a result from [16], Lemma 6.3, on the boundary behavior of a certain Poisson kernel convolution [see (3.1) and the comments following it]. We use \mathbb{E} to denote the expectation. If $v: \mathbb{R} \rightarrow \mathbb{C}$ is a continuous function and A is a self-adjoint matrix, then $v(A)$ is constructed using the continuous functional calculus.

LEMMA 4.4. *Given a deterministic $N \times N$ matrix C_N , a random $N \times N$ Hermitian matrix X_N , and a continuous function $h: \mathbb{R} \rightarrow \mathbb{R}$ with compact support, we have*

$$\mathbb{E}[\text{Tr}_N[h(X_N)C_N]] = \lim_{\gamma \downarrow 0} \frac{1}{\pi} \Im \int_{\mathbb{R}} \mathbb{E}[\text{Tr}_N[(X_N - (t + iy)I_N)^{-1}C_N]]h(t) dt.$$

4.2. *Matrix-valued functions and maps.* An essential ingredient in our analysis is the resolvent of A_N and of the matrices X_N (depending on the model considered, $X_N = A_N + U_N B_N U_N^*$, $X_N = A_N^{1/2} U_N^* B_N U_N A_N^{1/2}$ or $X_N = A_N U_N B_N U_N^*$). We denote by

$$(4.2) \quad R_N(z) = (zI_N - X_N)^{-1}, \quad z \notin \sigma(X_N)$$

the resolvent of X_N . It is a random matrix-valued rational function with poles in \mathbb{R} for the first two models and \mathbb{T} for the third. For the first two models, it has the following properties:

- (1) $R_N(\bar{z}) = R_N(z)^*$. In particular, $R_N(x)$ is self-adjoint if $x \in \mathbb{R} \setminus \sigma(X_N)$.
- (2) R_N is analytic at ∞ , and $\lim_{z \rightarrow \infty} zR_N(z) = I_N$, where I_N denotes the $N \times N$ identity matrix. The limit is in the norm topology of $M_N(\mathbb{C}) \otimes L^\infty(U(N), m_N)$, where m_N denotes the Haar measure on $U(N)$.
- (3) With the notation $\Re T = (T + T^*)/2$ and $\Im T = (T - T^*)/2i$ for the real and imaginary parts of T , respectively, we have for $z \in \mathbb{C}^+$

$$\begin{aligned} -\Im R_N(z) &= \Im z ((\Im z)^2 I_N + (\Re z I_N - X_N)^2)^{-1} \\ &\geq \frac{\Im z}{|z|^2 + 2|\Re z| \|X_N\| + \|X_N\|^2} I_N. \end{aligned}$$

This last quantity is uniformly bounded below in N for z in any fixed compact set $K \subseteq \mathbb{C}^+$. In particular, if $C > 0$ is such that $\sup_N \|X_N\| \leq C$,

$$(4.3) \quad -\Im \mathbb{E}[(zI_N - X_N)^{-1}] \geq \frac{\Im z}{|z|^2 + 2|\Re z|C + C^2} I_N.$$

For the unitary model, a slightly different property is needed.

- (a) If $z \in \mathbb{D}$, we have $\sigma(zX_N) \subset \mathbb{D}$, and thus $\sigma((I_N - zX_N)^{-1}) \subset \{w \in \mathbb{C}: \Re w > 1/2\}$. Therefore,

$$\Re \left[\frac{1}{z} R_N \left(\frac{1}{z} \right) \right] > \frac{1}{2} I_N, \quad z \in \mathbb{D}.$$

(This observation uses the fact that X_N is unitary, and hence normal.)

The following lemma is a fairly straightforward generalization of a result of Hurwitz. A similar result appears in [12]. In the statement, we use K_δ to denote the subset of γ consisting of all points at distance strictly less than δ from K . In the special case $K = \{\rho\}$, we write $(\rho - \delta, \rho + \delta)$ instead of K_δ .

LEMMA 4.5. *Let γ be either \mathbb{R} or \mathbb{T} , let $K \subsetneq \gamma$ be compact, and let r be a positive integer. Consider an analytic function $F : \overline{\mathbb{C}} \setminus K \rightarrow M_r(\mathbb{C})$ such that $F(z)$ is diagonal for each $z \in \mathbb{C} \setminus K$, $F(\infty) = I_r$ and $z \mapsto (F(z))_{ii} \in \mathbb{C}$ has only simple zeros, all of which are contained in $\gamma \setminus K$, $1 \leq i \leq r$. Fix $\delta > 0$ such that $\det(F)$ has no zeros on the boundary of K_δ relative to γ , and let $\rho_1, \dots, \rho_s \in \gamma$ be a list of those points $z \in \mathbb{C} \setminus K_\delta$ for which $F(z)$ is not invertible.*

Suppose that there exist positive numbers $\{\delta_N\}_{N \in \mathbb{N}}$ and analytic maps $F_N : \overline{\mathbb{C}} \setminus K_{\delta_N} \rightarrow M_r(\mathbb{C})$, $N \in \mathbb{N}$, such that:

- (1) $\lim_{N \rightarrow \infty} \delta_N = 0$;
- (2) $F_N(z)$ is invertible for $z \in \mathbb{C} \setminus \gamma$ and $N \in \mathbb{N}$; and
- (3) F_N converges to F uniformly on compact subsets of $\overline{\mathbb{C}} \setminus K$.

Then:

- (i) $\dim(\ker(F(\rho_j)))$ equals the order of ρ_j as a zero of $z \mapsto \det(F(z))$;
- (ii) Given $\varepsilon > 0$ such that

$$\varepsilon < \frac{1}{2} \min\{|\rho_i - \rho_j|, \text{dist}(\rho_i, K_\delta) : 1 \leq i \neq j \leq s\},$$

there exists an integer N_0 such that for $N \geq N_0$, we have:

- counting multiplicities, $\det(F_N)$ has exactly $\dim(\ker(F(\rho_j)))$ zeros in $(\rho_j - \varepsilon, \rho_j + \varepsilon) \subset \gamma$, $j = 1, \dots, s$, and
- $\{z \in \mathbb{C} \setminus K_\delta : \det(F_N(z)) = 0\} \subset \bigcup_{j=1}^s (\rho_j - \varepsilon, \rho_j + \varepsilon)$.

PROOF. Assertion (i) is obvious. The functions $f_N(z) = \det(F_N(z))$ converge to $f(z) = \det(F(z))$ uniformly on compact subsets of $\overline{\mathbb{C}} \setminus K$. The theorem of Hurwitz (see [35], Kapitel 8.5) guarantees that, for sufficiently large N , f_N has (counting multiplicities) exactly as many zeros as f in $\overline{\mathbb{C}} \setminus K_\delta$. All the zeros of f_N were assumed to be in γ and, therefore, these zeros cluster around $\{\rho_1, \dots, \rho_s\}$ in the following sense: for any given $\varepsilon > 0$, there exists an $N_\varepsilon \in \mathbb{N}$ such that

$$\{z \in \mathbb{C} \setminus K_\delta : \det(F_N(z)) = 0\} \subset \bigcup_{j=1}^s (\rho_j - \varepsilon, \rho_j + \varepsilon)$$

when $N \geq N_\varepsilon$. When $\varepsilon > 0$ is small enough, there are (counting multiplicities) exactly $\dim(\ker(F(\rho_j)))$ zeros of f_N in $(\rho_j - \varepsilon, \rho_j + \varepsilon)$. \square

Later, we apply this lemma in order to control the behavior of functions related to the resolvent R_N .

Next, we collect some facts about matrix functions and maps on matrix spaces which commute with the operation of conjugation by unitary matrices. First, an analog of the Nevanlinna representation for matrix-valued functions ([26], Section 5). Let $m > 0$ be fixed and let $F : \mathbb{C} \setminus [-m, m] \rightarrow M_N(\mathbb{C})$ be an analytic function. Assume that $\Im F(z) = (F(z) - F(z)^*)/2i$ is nonnegative definite for $z \in \mathbb{C}^+$, and $F(x) = F(x)^*$ for $x \in \mathbb{R} \setminus [-m, m]$. Then F can be represented as

$$(4.4) \quad F(z) = A + Bz - \int_{[-m,m]} \frac{d\rho(t)}{z-t}, \quad z \in \mathbb{C} \setminus [-m, m],$$

where A is a self-adjoint matrix, $B \geq 0$, and ρ is a measure with values in $M_N(\mathbb{C})$ such that $\rho(S) \geq 0$ for every Borel set $S \subset \mathbb{R}$. Observe that

$$\rho(\mathbb{R}) = \lim_{z \rightarrow \infty} z(A + Bz - F(z)).$$

The norm of such a function can obviously be estimated as

$$\|F(z)\| \leq \|A\| + \|B\||z| + \frac{\|\rho(\mathbb{R})\|}{\Im z}, \quad z \in \mathbb{C}^+.$$

The specific situation we have in mind is as follows. Let X be a random self-adjoint matrix in $M_N(\mathbb{C})$ such that $\|X\| \leq m$ almost surely. Pick $b \in M_N(\mathbb{C})$ such that $\Im b := (b - b^*)/2i > 0$ (that is, $\Im b$ is positive definite). The matrix $\mathbb{E}[(\Re b + z\Im b - X)^{-1}]$ is analytic in z , it is invertible for

$$z \in \mathbb{C} \setminus [-(m + \|\Re b\|)\|(\Im b)^{-1}\|, (m + \|\Re b\|)\|(\Im b)^{-1}\|],$$

and it is self-adjoint for

$$z \in \mathbb{R} \setminus [-(m + \|\Re b\|)\|(\Im b)^{-1}\|, (m + \|\Re b\|)\|(\Im b)^{-1}\|].$$

Moreover,

$$\Im \mathbb{E}[(\Re b + z\Im b - X)^{-1}] < 0, \quad z \in \mathbb{C}^+.$$

It follows that the function $F(z) = (\mathbb{E}[(\Re b + z\Im b - X)^{-1}])^{-1}$ satisfies the properties required for it to have a representation of the form (4.4). The matrices A, B and $\rho(\mathbb{R})$ are easily determined. Indeed, we have

$$\begin{aligned} (\mathbb{E}[(\Im b - \varepsilon(X - \Re b))^{-1}])^{-1} &= \Im b - \varepsilon \mathbb{E}[X - \Re b] - \varepsilon^2 [\mathbb{E}[X(\Im b)^{-1}X] \\ &\quad - \mathbb{E}[X](\Im b)^{-1}\mathbb{E}[X]] + O(\varepsilon^3) \end{aligned}$$

as $\varepsilon \rightarrow 0$. Substituting $\varepsilon = 1/z$, we obtain

$$(4.5) \quad \begin{aligned} F(z) &= z\Im b - \mathbb{E}[X] + \Re b \\ &\quad - \frac{\mathbb{E}[(X - \mathbb{E}[X])(\Im b)^{-1}(X - \mathbb{E}[X])]}{z} + O\left(\frac{1}{z^2}\right) \end{aligned}$$

as $z \rightarrow \infty$. This yields the three equalities below:

$$(4.6) \quad \begin{aligned} A &= -\mathbb{E}[X] + \Im b, & B &= \Im b, \\ \rho(\mathbb{R}) &= \mathbb{E}[(X - \mathbb{E}[X])(\Im b)^{-1}(X - \mathbb{E}[X])]. \end{aligned}$$

We are mostly, but not exclusively, interested in the case $\Im b = I_N$.

These observations apply to the variables X_N from our models. We begin with $X_N = A_N + U_N B_N U_N^*$, where (A_N) and (B_N) are any sequences of deterministic real diagonal matrices of size $N \times N$ with uniformly bounded norms and limiting distributions μ and ν , respectively. As before, $R_N(z) = (zI_N - X_N)^{-1}$. More generally, if $b \in M_N(\mathbb{C})$ satisfies $\Im b > 0$, then $R_N(b) = (b - X_N)^{-1}$.

LEMMA 4.6. *The function $b \mapsto \mathbb{E}[R_N(b)]$ takes values in $GL(N)$ whenever $\Im b > 0$. Moreover,*

$$\begin{aligned} \Im \mathbb{E}[R_N(b)]^{-1} &\geq \Im b \quad \text{and} \\ \|\mathbb{E}[R_N(b)]^{-1}\| &\leq \|b\| + C_1 + 4C_2 \|(\Im b)^{-1}\|, \quad \Im b > 0, \end{aligned}$$

where $C_1 = \sup_N (\|A_N\| + \|B_N\|)$, and $C_2 = \sup_N (\text{tr}_N(B_N^2) - [\text{tr}_N(B_N)]^2)$. In particular,

$$(4.7) \quad \begin{aligned} \Im \mathbb{E}[R_N(z)]^{-1} &\geq \Im z I_N \quad \text{and} \\ \|\mathbb{E}[R_N(z)]^{-1}\| &\leq |z| + C_1 + \frac{4C_2}{|\Im z|}, \quad z \in \mathbb{C}^+. \end{aligned}$$

PROOF. It is well known that an element of a C^* -algebra is invertible if its imaginary part is strictly positive or strictly negative. Since $\Im R_N(b) < 0$ for any b with $\Im b > 0$, and \mathbb{E} is completely positive, it follows that $\Im \mathbb{E}[R_N(b)] < 0$, so $\mathbb{E}[R_N(b)] \in GL(N)$. The relation $\Im \mathbb{E}[R_N(b)]^{-1} \geq \Im b$ follows from [11], Remark 2.5. The second inequality follows immediately from the observations preceding the lemma, and from the fact that for any deterministic matrix Z ,

$$\begin{aligned} &\mathbb{E}[U_N B_N U_N^* Z U_N B_N U_N^*] - \mathbb{E}[U_N B_N U_N^*] Z \mathbb{E}[U_N B_N U_N^*] \\ &= (\text{tr}_N(B_N^2) - [\text{tr}_N(B_N)]^2) \left(\frac{N^2}{N^2 - 1} \text{tr}_N(Z) I_N - \frac{1}{N^2 - 1} Z \right). \quad \square \end{aligned}$$

In some situations, it is convenient to see how $\mathbb{E}[(zI_N - X_N)^{-1}]$ depends on A_N ; recall that $X_N = A_N + U_N B_N U_N^*$. This is achieved to some extent by the following lemma (see also [30]).

LEMMA 4.7. *Fix a matrix $B_N \in M_N(\mathbb{C})$. Let $b \in M_N(\mathbb{C})$ be such that $b - U B_N U^*$ is invertible for every $U \in U(N)$, consider the random matrix $R(b) = (b - U_N B_N U_N^*)^{-1}$ and its expected value $G(b) = \mathbb{E}[(b - U_N^* B_N U_N)^{-1}]$. Then:*

(1) For every $Y \in M_N(\mathbb{C})$, we have

$$(4.8) \quad G(b)Y - YG(b) = \mathbb{E}[R(b)(Yb - bY)R(b)].$$

If $G(b)$ is invertible, we also have

$$(4.9) \quad \begin{aligned} & Y(G(b)^{-1} - b) - (G(b)^{-1} - b)Y \\ &= G(b)^{-1} \mathbb{E}[(R(b) - G(b))(Yb - bY)(R(b) - G(b))]G(b)^{-1}. \end{aligned}$$

(2) $G(b) \in \{b\}''$, where $\{b\}''$ denotes the double commutant of b in $M_N(\mathbb{C})$.

REMARK 4.8. The conclusion of item (2). of the above lemma applies to any complex differentiable map f defined on an open set in $M_N(\mathbb{C})$ with the property that $f(V^*bV) = V^*f(b)V$ for all $V \in U(N)$.

PROOF OF LEMMA 4.7. The analytic function

$$H(Y) = \mathbb{E}[(b - e^{iY}U_N B_N U_N^* e^{-iY})^{-1}]$$

is defined in an open set containing the self-adjoint matrices. Moreover, the invariance of the Haar measure under multiplication implies that H is constant on the self-adjoint matrices. Since the self-adjoint matrices form a uniqueness set for analytic functions, we deduce that H is constant in a neighborhood of the self-adjoint matrices. In particular, given $Y \in M_N(\mathbb{C})$, the function

$$\mathbb{E}[(b - e^{\varepsilon Y}U_N B_N U_N^* e^{-\varepsilon Y})^{-1}]$$

does not depend on ε for small $\varepsilon \in \mathbb{C}$. Differentiation at $\varepsilon = 0$ yields the identity

$$\mathbb{E}[R(b)(U_N B_N U_N^* Y - Y U_N B_N U_N^*)R(b)] = 0.$$

Using now the fact that $R(b)U_N B_N U_N^* = -I_N + R(b)b$ and $U_N B_N U_N^* R(b) = -I_N + bR(b)$, we obtain

$$\mathbb{E}[-YR(b) + R(b)bYR(b) + R(b)Y - R(b)YbR(b)] = 0$$

which is equivalent to (4.8) because $\mathbb{E}[R(b)Y] = G(b)Y$ and $\mathbb{E}[YR(b)] = YG(b)$.

To prove the second identity in (1), observe that

$$\begin{aligned} & \mathbb{E}[R(b)(Yb - bY)R(b)] \\ &= \mathbb{E}[(R(b) - G(b))(Yb - bY)(R(b) - G(b))] + G(b)(Yb - bY)G(b) \end{aligned}$$

so (4.8) implies

$$\begin{aligned} & G(b)Y - YG(b) - G(b)(Yb - bY)G(b) \\ &= \mathbb{E}[(R(b) - G(b))(Yb - bY)(R(b) - G(b))]. \end{aligned}$$

Relation (4.9) is now obtained multiplying this relation by $G(b)^{-1}$ on both sides.

To verify (2), we need to show that $G(b)$ commutes with any matrix $Y \in \{b\}'$. This follows immediately from (4.8). \square

The preceding lemma shows that $G(b)$ must be of the form $u(b)$ for some rational function u of a complex variable, and (4.9) allows us to show that in fact $G(b)^{-1}$ is close to a matrix of the form $b + wI_N$ when the variance of $R(b)$ is small. This follows from the next result.

LEMMA 4.9. *Assume that $\varepsilon > 0$, and $T \in M_N(\mathbb{C})$ satisfies the inequality*

$$|k^*(TY - YT)h| \leq \varepsilon \|Y\|$$

for every rank one matrix $Y \in M_N(\mathbb{C})$ and all unit vectors $h, k \in \mathbb{C}^N$. Then for any w in the numerical range $W(T) = \{h^*Th : \|h\| = 1\}$, we have $\|T - wI_N\| \leq 2\varepsilon$.

PROOF. Given two unit (column) vectors $h, k \in \mathbb{C}^N$, consider the—necessarily rank one—matrix $Y = kh^* \in M_N(\mathbb{C})$. The hypothesis implies that

$$|k^*Tk - h^*Th| = |k^*(TY - YT)h| \leq \varepsilon.$$

We deduce that the numerical range $W(T) = \{h^*Th : \|h\| = 1\}$ has diameter at most ε and, therefore, there $W(T - wI_N) \subset \{\lambda \in \mathbb{C} : |\lambda| \leq \varepsilon\}$ for any $w \in W(T)$. Thus, any $w \in W(T)$ satisfies the conclusion because the norm of an operator is at most twice its numerical radius (see [27], Theorem 1.3-1). \square

A further property of eigenvectors of Hermitian matrices which are close in norm to each other appears in the analysis of the behavior of the eigenvectors of our matrix models. The following lemma appears already in [16]; we offer a proof for the reader’s convenience.

LEMMA 4.10. *Let X and X_0 be Hermitian $N \times N$ matrices. Assume that $\alpha, \beta, \delta \in \mathbb{R}$ are such that $\alpha < \beta, \delta > 0$, and neither X nor X_0 has any eigenvalues in $[\alpha - \delta, \alpha] \cup [\beta, \beta + \delta]$. Then*

$$\|E_X((\alpha, \beta)) - E_{X_0}((\alpha, \beta))\| < \frac{4(\beta - \alpha + 2\delta)}{\pi\delta^2} \|X - X_0\|.$$

In particular, for any unit vector $\xi \in E_{X_0}((\alpha, \beta))(\mathbb{C}^N)$,

$$\|(I_N - E_X((\alpha, \beta)))\xi\|_2 < \frac{4(\beta - \alpha + 2\delta)}{\pi\delta^2} \|X - X_0\|.$$

PROOF. Consider the rectangle γ having as corners the complex points $\alpha - (1 \pm i)\delta/2$ and $\beta + (1 \pm i)\delta/2$. By assumptions, we have $\sigma(X) \cap ([\alpha - \delta, \alpha] \cup [\beta, \beta + \delta]) = \emptyset$ and $\sigma(X_0) \cap ([\alpha - \delta, \alpha] \cup [\beta, \beta + \delta]) = \emptyset$. Thus, the spectral projections can be obtained by analytic functional calculus:

$$E_X((\alpha, \beta)) - E_{X_0}((\alpha, \beta)) = \frac{1}{2\pi i} \int_{\gamma} [(\lambda - X)^{-1} - (\lambda - X_0)^{-1}] d\lambda.$$

An application of the resolvent equation and elementary norm estimates yield

$$\begin{aligned}
 & \|E_X((\alpha, \beta)) - E_{X_0}((\alpha, \beta))\| \\
 &= \frac{1}{2\pi} \left\| \int_{\gamma} (\lambda - X)^{-1} (X_0 - X) (\lambda - X_0)^{-1} d\lambda \right\| \\
 &\leq \frac{1}{2\pi} \int_{\gamma} \|(\lambda - X)^{-1} (X_0 - X) (\lambda - X_0)^{-1}\| d\lambda \\
 &\leq (\beta - \alpha + 2\delta) \frac{\|X - X_0\|}{\pi} \sup_{\lambda \in \gamma} \frac{1}{\|\lambda - X\|} \sup_{\lambda \in \gamma} \frac{1}{\|\lambda - X_0\|} \\
 &< \frac{4(\beta - \alpha + 2\delta)}{\pi \delta^2} \|X - X_0\|.
 \end{aligned}$$

The lemma follows. \square

For the following concentration of measure result, it is convenient to identify \mathbb{C}^N with the subspace of \mathbb{C}^{N+1} consisting of all vectors whose last component is zero. Similarly, $M_N(\mathbb{C})$ is identified with those matrices in $M_{N+1}(\mathbb{C})$ whose last column and row are zero. We use the notation \mathbb{V} for variance.

LEMMA 4.11. *Fix a positive integer r , a projection P of rank r and a scalar $z \in \mathbb{C} \setminus \mathbb{R}$. Then:*

- (i) $\lim_{N \rightarrow \infty} \|P R_N(z) P^* - P \mathbb{E}[R_N(z)] P^*\| = 0$ almost surely.
- (ii) Given unit vectors $h, k \in \mathbb{C}^N$, $\mathbb{V}(k^* R_N(z) h) \leq C/[N|\Im z|^4]$.

PROOF. Assertion (i) is equivalent to the statement that, given unit vectors $h, k \in \mathbb{C}^N$

$$(4.10) \quad \lim_{N \rightarrow \infty} k^*(R_N(z) - \mathbb{E}[R_N(z)])h = 0$$

almost surely. The random variable $k^* R_N(z) h$ is a Lipschitz function on the unitary group $U(N)$ with Lipschitz constant $C/|\Im z|^2$. An application of [2], Corollary 4.4.28, yields the inequality

$$\mathbb{P}\left(|k^*(R_N(z) - \mathbb{E}[R_N(z)])h| > \frac{\varepsilon}{N^{\frac{1}{2}-\alpha}}\right) \leq 2 \exp(-CN^{2\alpha}|\Im z|^4 \varepsilon^2)$$

for any $\alpha \in (0, 1/2)$, and (4.10) follows by an application of the Borel–Cantelli lemma. To prove (ii), apply the same inequality in the usual formula $\mathbb{E}[X] = \int_0^{+\infty} \mathbb{P}(X > t) dt$ for a positive random variable X . \square

In the following result, the coefficient t^4 can be replaced by t^2 if we estimate the operator norm of a matrix by its Hilbert–Schmidt norm.

COROLLARY 4.12. Fix a positive integer t , matrices Y, Z of rank at most t , and a scalar $z \in \mathbb{C} \setminus \mathbb{R}$. Then:

$$\mathbb{E}[\|Y(R_N(z) - \mathbb{E}[R_N(z)])Z\|^2] \leq Ct^4 \|Y\|^2 \|Z\|^2 / [N|\Im z|^4].$$

PROOF. Choose orthonormal vectors h_1, \dots, h_t whose span contains the range of Z and orthonormal vectors k_1, \dots, k_t whose span contains the range of Y^* . The corollary follows from the inequality

$$\|Y(R_N(z) - \mathbb{E}[R_N(z)])Z\| \leq \sum_{i,j=1}^t \|Y\| \|Z\| |k_j^*(R_N(z) - \mathbb{E}[R_N(z)])h_i|$$

and Part (ii) of the preceding lemma. \square

REMARK 4.13. We note for further use that Lemma 4.11 and Corollary 4.12 apply to the resolvent of any self-adjoint polynomial in $m + 1$ noncommuting variables $P(A_N^{(1)}, \dots, A_N^{(m)}, U_N B_N U_N^*)$ as long as the norms of $A_N^{(j)}$ and B_N are uniformly bounded in N .

5. Proofs of the main results. The three subsections below provide parallel treatments of the three models under consideration.

5.1. *The additive model* $X_N = A_N + U_N B_N U_N^*$. We use the notation from Section 2.1. Fix $\alpha \in \text{supp}(\mu)$ and $\beta \in \text{supp}(\nu)$. Due to the left and right invariance of the Haar measure on $U(N)$ we may, and do, assume without loss of generality that both A_N and B_N are diagonal matrices. More precisely, we let A_N be the diagonal matrix

$$A_N = \text{Diag}(\theta_1, \dots, \theta_p, \alpha_1^{(N)}, \dots, \alpha_{N-p}^{(N)}),$$

where $\alpha_1^{(N)} \geq \dots \geq \alpha_{N-p}^{(N)}$. We also have $\theta_1 \geq \dots \geq \theta_p$, but no order relation is assumed between θ_i and $\alpha_j^{(N)}$ other than $\theta_i \neq \alpha_j^{(N)}$. For $N \geq p$, we write $A_N = A'_N + A''_N$, where

$$A'_N = \text{Diag}(\underbrace{\alpha, \dots, \alpha}_p, \alpha_1^{(N)}, \dots, \alpha_{N-p}^{(N)}),$$

and

$$A''_N = \text{Diag}(\theta_1 - \alpha, \dots, \theta_p - \alpha, \underbrace{0, \dots, 0}_{N-p}).$$

We have $A''_N = P_N^* \Theta P_N$, where P_N is the $p \times N$ matrix representing the usual projection $\mathbb{C}^N \rightarrow \mathbb{C}^p$ onto the first p coordinates, and

$$\Theta = \text{Diag}(\theta_1 - \alpha, \dots, \theta_p - \alpha).$$

The operator P_N is precisely $E_{A_N}(\{\theta_1, \dots, \theta_p\})$ co-restricted to its range. Similarly, $B_N = B'_N + B''_N$, where

$$B'_N = \text{Diag}(\underbrace{\beta, \dots, \beta}_q, \beta_1^{(N)}, \dots, \beta_{N-q}^{(N)}),$$

$$B''_N = \text{Diag}(\tau_1 - \beta, \dots, \tau_q - \beta, \underbrace{0, \dots, 0}_{N-q}) = Q_N^* T Q_N,$$

$$T = \text{Diag}(\tau_1 - \beta, \dots, \tau_q - \beta),$$

and Q_N is the $q \times N$ matrix representing the usual projection $\mathbb{C}^N \rightarrow \mathbb{C}^q$.

5.1.1. *Reduction to the almost sure convergence of a $p \times p$ matrix.* Here, we explain how to reduce, in the spirit of [12], the problem of locating outliers of $A_N + U_N B'_N U_N^*$ to a convergence problem for a random matrix of fixed size $p \times p$. The matrices A'_N and B'_N have no spikes and, therefore, [21], Corollary 2.2, applies to the matrix $X'_N = A'_N + U_N B'_N U_N^*$. Recall that $K = \text{supp}(\mu \boxplus \nu)$. The corollary states that for any integer $k > 0$ there exists almost surely $N_k \in \mathbb{N}$ such that for any $N \geq N_k$, we have $\sigma(X'_N) \subset K_{\frac{1}{k}}$. We reformulate this result as follows: there exist positive random variables $(\delta_N)_{N \in \mathbb{N}}$, such that

$$\sigma(A'_N + U_N B'_N U_N^*) \subseteq K_{\delta_N}, \quad N \in \mathbb{N},$$

and $\lim_{N \rightarrow \infty} \delta_N = 0$ almost surely. Indeed, choose for instance $\delta_N = \frac{1}{k}$ for any $N_k \leq N < N_{k+1}$. Given $z \in \mathbb{C} \setminus K_{\delta_N}$, we have

$$\begin{aligned} zI_N - (A_N + U_N B'_N U_N^*) &= zI_N - X'_N - A''_N \\ &= (zI_N - X'_N)(I_N - (zI_N - X'_N)^{-1} A''_N) \end{aligned}$$

and, therefore,

$$\det(zI_N - (A_N + U_N B'_N U_N^*)) = \det(zI_N - X'_N) \det(I_N - (zI_N - X'_N)^{-1} P_N^* \Theta P_N).$$

Using the fact that $\det(I - XY) = \det(I - YX)$ when XY and YX are square matrices, we obtain

$$\det(I_N - (zI_N - X'_N)^{-1} P_N^* \Theta P_N) = \det(I_p - P_N(zI_N - X'_N)^{-1} P_N^* \Theta),$$

so

$$\begin{aligned} \det(zI_N - (A_N + U_N B'_N U_N^*)) &= \det(zI_N - X'_N) \det(I_p - P_N(zI_N - X'_N)^{-1} P_N^* \Theta). \end{aligned}$$

We conclude that the eigenvalues of $A_N + U_N B'_N U_N^*$ outside K_{δ_N} are precisely the zeros of the function $\det(F_N(z))$, where

$$(5.1) \quad F_N(z) = I_p - P_N(zI_N - X'_N)^{-1} P_N^* \Theta$$

in that open set. This is a random analytic function defined on $\overline{\mathbb{C}} \setminus K_{\delta_N}$, with values in $M_p(\mathbb{C})$. We argue next that the sequence $\{F_N(z)\}_N$ converges almost surely to the deterministic diagonal matrix function

$$F(z) = \text{Diag}\left(1 - \frac{\theta_1 - \alpha}{\omega_1(z) - \alpha}, \dots, 1 - \frac{\theta_p - \alpha}{\omega_1(z) - \alpha}\right),$$

where ω_1 is the subordination function from (2.3).

5.1.2. *Convergence of F_N .* We begin with a somewhat more general result.

PROPOSITION 5.1. *Fix a positive integer p , and let C_N and D_N be deterministic real diagonal $N \times N$ matrices whose norms are uniformly bounded and such that the limits*

$$\eta_i = \lim_{N \rightarrow \infty} (C_N)_{ii}$$

exist for $i = 1, 2, \dots, p$. Suppose that the empirical eigenvalue distributions of C_N and D_N converge weakly to μ and ν , respectively. Then the resolvent

$$R_N(z) = (zI_N - C_N - U_N D_N U_N^*)^{-1}, \quad z \in \mathbb{C} \setminus \mathbb{R},$$

satisfies

$$(5.2) \quad \lim_{N \rightarrow \infty} P_N \mathbb{E}[R_N(z)] P_N^* = \text{Diag}\left(\frac{1}{\omega_1(z) - \eta_1}, \dots, \frac{1}{\omega_1(z) - \eta_p}\right),$$

where ω_1 is the subordination function from (2.3).

PROOF. Since all functions involved satisfy $f(\bar{z}) = f(z)^*$, it suffices to consider the case of $z \in \mathbb{C}^+$. Fix such a scalar z and apply Lemma 4.6 and Lemma 4.7(2) to $b = zI_N - C_N$ to conclude that the $N \times N$ matrix $\mathbb{E}[R_N(z)]$ is invertible and diagonal. Set

$$(5.3) \quad \omega_{N,i}(z) = \frac{1}{\mathbb{E}[R_N(z)]_{ii}} + (C_N)_{ii}, \quad 1 \leq i \leq p,$$

and observe that $\Im \omega_{N,i}(z) \geq \Im z$ for $z \in \mathbb{C}^+$ by Lemma 4.6. We proceed to show that this function satisfies an approximate subordination relation. We state this separately for future reference.

LEMMA 5.2. *We have*

$$\lim_{N \rightarrow \infty} \|\mathbb{E}[R_N(z)] - (\omega_{N,i}(z)I_N - C_N)^{-1}\| = 0, \quad z \in \mathbb{C}^+, 1 \leq i \leq p.$$

PROOF. The existence of

$$\Omega_N(z) = \mathbb{E}[R_N(z)]^{-1} + C_N, \quad z \in \mathbb{C}^+,$$

is guaranteed by Lemma 4.6. We apply now Lemma 4.7 with $zI_N - C_N$ in place of b and D_N in place of B_N , so $\mathbb{E}[R_N(z)] = G(zI_N - C_N)$. Relation (4.9) shows that

$$\begin{aligned} & Y\Omega_N(z) - \Omega_N(z)Y \\ &= \mathbb{E}[R_N(z)]^{-1} \mathbb{E}[(R_N(z) - \mathbb{E}[R_N(z)]) \\ &\quad \times (YC_N - C_N Y)(R_N(z) - \mathbb{E}[R_N(z)])] \mathbb{E}[R_N(z)]^{-1}, \quad Y \in M_N(\mathbb{C}). \end{aligned}$$

Suppose that Y has rank one and $h, k \in \mathbb{C}^N$ are unit vectors. In this case, there exist rank one projections p_1, p_2 and rank two projections q_1, q_2 (depending on z) such that

$$\begin{aligned} & k^*(Y\Omega_N(z) - \Omega_N(z)Y)h \\ &= k^* \mathbb{E}[R_N(z)]^{-1} \\ &\quad \times \mathbb{E}[p_1(R_N(z) - \mathbb{E}[R_N(z)])q_1(YC_N - C_N Y)q_2(R_N(z) - \mathbb{E}[R_N(z)])p_2] \\ &\quad \times \mathbb{E}[R_N(z)]^{-1}h. \end{aligned}$$

Indeed, the third and first factors in the product above have rank one, while the rank of $YC_N - C_N Y$ is at most two. We deduce that

$$\begin{aligned} |k^*(\Omega_N(z)Y - Y\Omega_N(z))h| &\leq \|\mathbb{E}[R_N(z)]^{-1}\|^2 \|YC_N - C_N Y\| \\ &\quad \times \mathbb{E}[\|p_1(R_N(z) - \mathbb{E}[R_N(z)])q_1\|^2]^{1/2} \\ &\quad \times \mathbb{E}[\|q_2(R_N(z) - \mathbb{E}[R_N(z)])p_2\|^2]^{1/2}. \end{aligned}$$

We use now the estimates from Lemma 4.6 and Corollary 4.12 along with the inequality $\|YC_N - C_N Y\| \leq 2\|C_N\|\|Y\|$ to obtain a constant $C > 0$ (independent of N and z) such that

$$|k^*(\Omega_N(z)Y - Y\Omega_N(z))h| \leq C \frac{(|z| + 1 + (1/\Im z))^2}{N|\Im z|^4} \|Y\|.$$

The number $\omega_{N,i}(z)$ is precisely the (i, i) entry of the matrix $\Omega_N(z)$, and thus it belongs to the numerical range $W(\Omega_N(z))$; indeed it equals $e_i^* \Omega_N(z) e_i$, where e_1, \dots, e_N is the canonical basis in which C_N is diagonal. Lemma 4.9 yields the estimate

$$\|\Omega_N(z) - \omega_{N,i}(z)I_N\| \leq 2C \frac{(|z| + 1 + (1/\Im z))^2}{N|\Im z|^4},$$

which gives the desired result as $N \rightarrow \infty$. \square

Fix now $i \in \{1, \dots, p\}$ and observe that the family of functions $(\omega_{N,i})_N$ is normal on \mathbb{C}^+ . Voiculescu’s asymptotic freeness result shows that

$$\lim_{N \rightarrow \infty} \operatorname{tr}_N(\mathbb{E}[R_N(z)]) = G_{\mu_{\boxplus v}}(z) = G_v(\omega_1(z)),$$

so Lemma 5.2 implies that $\omega_{N,i}$ converges uniformly on compact subsets of \mathbb{C}^+ . This, together with a second application of Lemma 5.2, implies (5.2) and completes the proof of Proposition 5.1. \square

The convergence result for the functions F_N also uses a normal family argument, more specifically the fact that a normal sequence converges uniformly on compact sets if it converges pointwise on a set with an accumulation point which belongs to the domain.

PROPOSITION 5.3. *Almost surely, the sequence $\{F_N\}_N$ converges uniformly on compact subsets of $\overline{\mathbb{C}} \setminus K$ to the analytic function F defined by*

$$(5.4) \quad F(z) = \operatorname{Diag}\left(1 - \frac{\theta_1 - \alpha}{\omega_1(z) - \alpha}, \dots, 1 - \frac{\theta_p - \alpha}{\omega_1(z) - \alpha}\right), \quad z \in \overline{\mathbb{C}} \setminus K.$$

PROOF. Lemma 3.1, Part (b), and the hypothesis on α , show that the function $z \mapsto 1/(\omega_1(z) - \alpha)$ is analytic on $\overline{\mathbb{C}} \setminus K$. Define

$$\mathcal{D} = \{z \in \mathbb{C} \setminus K : \Re z \in \mathbb{Q}, \Im z \in \mathbb{Q} \setminus \{0\}\}.$$

The first p diagonal elements of A'_N are all equal to α , and thus Lemma 4.11 (i) and equation (5.2) of Proposition 5.1 (applied to $C_N = A'_N$ and $D_N = B'_N$) show that given $z \in \mathcal{D}$, the sequence $P_N(zI_N - X'_N)^{-1}P_N^*$ converges almost surely to $(1/(\omega_1(z) - \alpha))I_p$. Moreover, by [21], these functions are almost surely uniformly bounded on any compact subset of $\mathbb{C} \setminus K$. Uniform boundedness on some neighborhood of infinity in $\mathbb{C} \cup \{\infty\}$ is automatic. We deduce that, almost surely, this sequence converges uniformly on compact subsets of $\overline{\mathbb{C}} \setminus K$ to the function $(1/(\omega_1 - \alpha))I_p$. The proposition follows immediately from these facts and (5.1). \square

5.1.3. *Proofs of the main results for the additive model.*

PROOF OF THEOREM 2.1, PARTS (1) AND (2)—EIGENVALUE BEHAVIOR. We proceed in two steps.

Step 1. We investigate first the case in which $q = 0$, that is, $B_N = B'_N$ has no spikes. Equivalently, we prove our result for the simpler model $A_N + U_N B'_N U_N^*$. We can work on the almost sure event on which:

- there exists a random sequence $\{\delta_N\}_{N \in \mathbb{N}}$ (as introduced in Section 5.1.1) such that $\lim_{N \rightarrow \infty} \delta_N = 0$ and $\sigma(A'_N + U_N B'_N U_N^*) \subseteq K_{\delta_N}$ for all N , and

- the sequence $(F_N)_N$ defined in (5.1) converges to the function F defined by (5.4) uniformly on the compact subsets of $\overline{\mathbb{C}} \setminus K$ (guaranteed by Proposition 5.3).

We apply Lemma 4.5 on this event, with $\gamma = \mathbb{R}$. We first argue that the hypotheses of that lemma are satisfied. The values of F are clearly diagonal matrices and $F(\infty) = I_p$. We show that the zeros of $(F(z))_{ii}$ are simple. Indeed,

$$(F'(z))_{ii} = \frac{\omega'_1(z)(\theta_i - \alpha)}{(\omega_1(z) - \alpha)^2},$$

and the zeros of ω'_1 are simple by the Julia–Carathéodory theorem ([25], Chapter I, Exercises 6 and 7) because $\omega_1(\mathbb{C}^+) \subset \mathbb{C}^+$.

Hypotheses (1) and (3) of Lemma 4.5 follow from Proposition 5.3. To verify hypothesis (2) of Lemma 4.5, observe that if $F_N(z)$ is not invertible then z is an eigenvalue of the self-adjoint matrix $A_N + U_N B'_N U_N^*$, hence real. There are arbitrarily small numbers $\delta > 0$ such that the boundary points of K_δ are not zeros of $\det(F)$. When this condition is satisfied, Lemma 4.5 yields precisely the conclusion of Theorem 2.1(1)–(2), when $q = 0$. Indeed, as explained in Section 5.1.1, the eigenvalues of $A_N + U_N B'_N U_N^*$ in $\mathbb{C} \setminus K_\delta$ are exactly the zeros of $\det(F_N)$, and the set of points z such that $F(z)$ is not invertible is precisely $\bigcup_{i=1}^p \omega_1^{-1}(\{\theta_i\})$. This completes the first step.

Step 2. Suppose now that $q > 0$ and use Step 1 above to obtain the existence of a sequence of positive random variables $(\delta_N)_{N \in \mathbb{N}}$ such that $\lim_{N \rightarrow \infty} \delta_N = 0$ almost surely and $\sigma(A_N + U_N B'_N U_N^*) \subseteq K''_{\delta_N}$, where

$$K'' = K \cup \bigcup_{i=1}^p \omega_1^{-1}(\{\theta_i\}).$$

We proceed as in Step 1 (switching the roles of A_N and B_N) in order to conclude that the eigenvalues of X_N outside K''_{δ_N} are precisely the zeros of the function $\det(I_q - Q_N(zI_N - U_N A_N U_N^* - B'_N)^{-1} Q_N^* T)$ in that open set, where

$$T = \text{Diag}(\tau_1 - \beta, \dots, \tau_q - \beta),$$

and Q_N is the orthogonal projection $\mathbb{C}^N \rightarrow \mathbb{C}^q$. Lemma 4.5 is applied now to the functions

$$\tilde{F}_N(z) = I_q - Q_N(zI_N - U_N A_N U_N^* - B'_N)^{-1} Q_N^* T, \quad N \geq q,$$

$$\tilde{F}(z) = \text{Diag}\left(1 - \frac{\tau_1 - \beta}{\omega_2(z) - \beta}, \dots, 1 - \frac{\tau_q - \beta}{\omega_2(z) - \beta}\right),$$

and the compact set K'' . The convergence of $\{\tilde{F}_N\}_N$ to \tilde{F} follows by an adaptation of Proposition 5.3. This completes the proof of Parts (1) and (2) of Theorem 2.1 in the general case $q > 0$, provided that $k = 0$. By symmetry, we have also proved

these assertions in case $\ell = 0$. To prove (2) in case $k\ell \neq 0$, we use a perturbation argument. Fix $\rho \in \mathbb{R} \setminus K$ such that $\omega_1(\rho) = \theta_{i_0}$ for some $i_0 \in \{1, \dots, p\}$ and $\omega_2(\rho) = \tau_{j_0}$ for some $j_0 \in \{1, \dots, q\}$, and fix $\varepsilon > 0$ as in the statement of (2). Choose $\delta \in (0, \varepsilon/3)$ so small that $\omega_1((\rho - 3\delta, \rho + 3\delta))$ contains no spikes $\theta_i \neq \theta_{i_0}$ and $\omega_2((\rho - 3\delta, \rho + 3\delta))$ contains no spikes $\tau_j \neq \tau_{j_0}$. Since ω_1 is strictly increasing on $(\rho - 3\delta, \rho + 3\delta)$, we have $\omega_1(\rho + 2\delta) = \theta_{i_0} + \eta$, with $\eta > 0$. We use the already established Part 2 of the theorem to conclude that, almost surely for large N , the perturbed matrix

$$X'_N = X_N + \eta E_{A_N}(\{\theta_{i_0}\})$$

has ℓ eigenvalues in $(\rho - \delta, \rho + \delta)$ and another k eigenvalues in the disjoint interval $(\rho + \delta, \rho + 3\delta)$. An application of Lemma 4.10 and Part 1 of the theorem for sufficiently small δ shows that X_N has $k + \ell$ eigenvalues in $(\rho - \varepsilon, \rho + \varepsilon)$. \square

PROOF OF THEOREM 2.1, PARTS (3) AND (4)—EIGENSPACE BEHAVIOR. We borrow heavily from the techniques of [16]. There are again two steps.

Step A. We prove Theorem 2.1(3)–(4) under the additional assumption that $\theta_1 > \dots > \theta_p$, $\tau_1 > \dots > \tau_q$, $k = 1$, and $\ell = 0$. Thus $\omega_1(\rho) = \theta_{i_0}$ for some $i_0 \in \{1, \dots, p\}$, and $\omega_2(\rho) \notin \{\tau_1, \dots, \tau_q\}$. Assertion (3) of Theorem 2.1 follows if the equalities

$$\lim_{N \rightarrow \infty} \left\| E_{A_N}(\{\theta_i\}) E_{X_N}((\rho - \varepsilon, \rho + \varepsilon)) E_{A_N}(\{\theta_i\}) - \frac{\delta_{i_0 i}}{\omega'_1(\rho)} E_{A_N}(\{\theta_i\}) \right\| = 0,$$

and

$$\lim_{N \rightarrow \infty} \left\| E_{B_N}(\{\tau_j\}) E_{X_N}((\rho - \varepsilon, \rho + \varepsilon)) E_{B_N}(\{\tau_j\}) \right\| = 0,$$

are shown to hold almost surely for all $i = 1, \dots, p$, $j = 1, \dots, q$. The Hermitian matrices in these equations have rank one, so their norm is equal to the absolute value of their unnormalized trace. Thus, we need to show that

$$(5.5) \quad \lim_{N \rightarrow \infty} \text{Tr}_N [E_{A_N}(\{\theta_i\}) E_{X_N}((\rho - \varepsilon, \rho + \varepsilon))] = \frac{\delta_{i_0 i}}{\omega'_1(\rho)}, \quad i = 1, \dots, p,$$

and

$$(5.6) \quad \lim_{N \rightarrow \infty} \text{Tr}_N [E_{B_N}(\{\tau_j\}) E_{X_N}((\rho - \varepsilon, \rho + \varepsilon))] = 0, \quad j = 1, \dots, q,$$

almost surely. It is useful to write the random variable in (5.5) in terms of functional calculus with *continuous* rather than indicator functions. Choose $\delta > 0$ so small that each interval $[\theta_i - \delta, \theta_i + \delta]$ contains exactly one point of $\sigma(A_N)$ (namely, θ_i) for $i = 1, \dots, p$ and for large N . For each $i = 1, \dots, p$, choose a function $f_i \in \mathcal{D}(\mathbb{R})$ with support in $[\theta_i - \delta, \theta_i + \delta]$ such that $0 \leq f_i \leq 1$ and $f_i(\theta_i) = 1$. Also choose a function $h \in \mathcal{D}(\mathbb{R})$ with support in $[\rho - \varepsilon, \rho + \varepsilon]$ such

that $0 \leq h \leq 1$ and $h(x) = 1$ for $x \in [\rho - \varepsilon/2, \rho + \varepsilon/2]$. For sufficiently large N , we have $E_{A_N}(\{\theta_i\}) = f_i(A_N)$. Also, by the already established assertion (1) of the theorem, we have $E_{X_N}((\rho - \varepsilon, \rho + \varepsilon)) = h(X_N)$ almost surely for N sufficiently large. Thus, we see that, almost surely for sufficiently large N and for $i = 1, \dots, p$, we have

$$(5.7) \quad \text{Tr}_N[E_{A_N}(\{\theta_i\})E_{X_N}((\rho - \varepsilon, \rho + \varepsilon))] = \text{Tr}_N[h(X_N)f_i(A_N)].$$

To complete the proof of (5.5), we obtain as in Lemma 4.11 a concentration inequality for the right-hand side of (5.7) and then we estimate the expected value. In the following argument we use the fact that a Lipschitz function on \mathbb{R} remains Lipschitz, with the same constant, when considered as a function on the selfadjoint matrices endowed with the Hilbert–Schmidt norm. See [16], Lemma A.2, for a simple proof of this fact, first observed in [15].

LEMMA 5.4. *Fix $i \in \{1, \dots, p\}$, denote by γ the Lipschitz constant of the function h , and set $C = \sup_N \|B_N\|$. For N sufficiently large, the random variable $Z_N = \text{Tr}[h(X_N)f_i(A_N)]$ satisfies the concentration inequality*

$$\mathbb{P}(|Z_N - \mathbb{E}(Z_N)| > \eta) \leq 2 \exp\left(-\frac{\eta^2 N}{4C^2 \gamma^2}\right), \quad \eta > 0.$$

PROOF. The lemma follows from [2], Corollary 4.4.28, once we establish that the Lipschitz constant of the function

$$g(U) = \text{Tr}_N[h(A_N + UB_NU^*)f_n(A_N)], \quad U \in U(N),$$

is at most $2C\gamma$. For any U and V in $U(N)$, we have

$$\begin{aligned} |g(U) - g(V)| &= |\text{Tr}_N[f_i(A_N)(h(A_N + UB_NU^*) - h(A_N + VB_NV^*))]| \\ &\leq \|h(A_N + UB_NU^*) - h(A_N + VB_NV^*)\|_2 \\ &\leq \gamma \|U^*B_NU - V^*B_NV\|_2, \end{aligned}$$

where we used the Cauchy–Schwarz inequality for the Hilbert–Schmidt norm and the fact that $\|f_i(A_N)\|_2 \leq 1$. Since

$$\begin{aligned} \|UB_NU^* - VB_NV^*\|_2 &\leq \|UB_N(U^* - V^*)\|_2 + \|(U - V)B_NV^*\|_2 \\ &\leq 2\|B_N\| \|U - V\|_2, \end{aligned}$$

we conclude that $|g(U) - g(V)| \leq 2C\gamma \|U - V\|_2$, as desired. \square

The above result, combined with the Borel–Cantelli lemma, yields immediately

$$\lim_{N \rightarrow \infty} (\text{Tr}_N[h(X_N)f_i(A_N)] - \mathbb{E}[\text{Tr}_N[h(X_N)f_i(A_N)]]) = 0, \quad i = 1, \dots, p,$$

almost surely. We complete the proof of (5.5) by showing that

$$\lim_{N \rightarrow \infty} \mathbb{E}[\text{Tr}_N[h(X_N) f_i(A_N)]] = \frac{\delta_{i0i}}{\omega'_1(\rho)}, \quad i = 1, \dots, p.$$

Lemma 4.4 with $C_N = f_i(A_N)$ allows us to rewrite this as

$$\lim_{N \rightarrow \infty} \lim_{y \downarrow 0} \frac{1}{\pi} \Im \int_{\mathbb{R}} \mathbb{E}[\text{Tr}_N[R_N(t + iy) f_i(A_N)]] h(t) dt = -\frac{\delta_{i0i}}{\omega'_1(\rho)},$$

$i = 1, \dots, p,$

or more simply, because $f_i(A_N)$ is the projection of \mathbb{C}^N onto the i th coordinate,

$$(5.8) \quad \lim_{N \rightarrow \infty} \lim_{y \downarrow 0} \frac{1}{\pi} \Im \int_{\mathbb{R}} \mathbb{E}[[R_N(t + iy)]_{ii}] h(t) dt = -\frac{\delta_{i0i}}{\omega'_1(\rho)}, \quad i = 1, \dots, p.$$

Lemma 5.2 suggests writing

$$(5.9) \quad \mathbb{E}[R_N(z)_{ii}] = \frac{1}{\omega_1(z) - \theta_i} + \Delta_{i,N}(z), \quad i = 1, \dots, p, z \in \mathbb{C}^+.$$

We proceed to estimate the functions $\Delta_{i,N}$.

PROPOSITION 5.5. *There exist positive numbers $\{a_N\}_N$ such that*

$$|\Delta_{i,N}(z)| \leq a_N(1 + |z|)^4(1 + |\Im z|^{-1})^4, \quad z \in \mathbb{C} \setminus \mathbb{R}, i = 1, \dots, p,$$

and $\lim_{N \rightarrow \infty} a_N = 0$.

PROOF. Define analytic functions $\omega_{N,i}$ for $i = 1, \dots, p$ using (5.3) with A_N, B_N in place of C_N, D_N , respectively. These functions are analytic outside the interval $[-\|A_N\| - \|B_N\|, \|A_N\| + \|B_N\|]$, and the hypothesis that $\|A_N\|$ and $\|B_N\|$ are uniformly bounded implies their analyticity on $\mathbb{C} \setminus [-m, m]$ for some $m > 0$ independent of N . The matrix $\mathbb{E}[R_N(z)]$ belongs to $\{A_N\}''$ by Lemma 4.7, and is therefore a diagonal matrix, so

$$\omega_{N,i}(z) = (\mathbb{E}[R_N(z)]^{-1})_{ii} + \theta_i, \quad i = 1, \dots, p.$$

By Lemma 4.6 and the considerations preceding it [especially (4.5) and (4.6)],

$$\lim_{z \rightarrow \infty} \frac{\omega_{N,i}(z)}{z} = (I_N)_{ii} = 1,$$

$$\lim_{z \rightarrow \infty} \omega_{N,i}(z) - z = -(A_N + \mathbb{E}[U_N B_N U_N^*])_{ii} + \theta_i = -\text{tr}_N(B_N)$$

and

$$\begin{aligned} \lim_{z \rightarrow \infty} z(\omega_{N,i}(z) - z + \text{tr}_N(B_N)) &= -\mathbb{E}[(X_N - \mathbb{E}[(X_N)])^2]_{ii} \\ &= -(\text{tr}_N(B_N^2) - \text{tr}_N(B_N)^2), \end{aligned}$$

for $N \geq p$ and $i = 1, \dots, p$. It follows that we have, as in (3.3), Nevanlinna representations of the form

$$\omega_{N,i}(z) = z - \operatorname{tr}_N(B_N) - \int_{[-m,m]} \frac{1}{z-t} d\sigma_{N,i}(t), \quad z \in \mathbb{C} \setminus [-m, m],$$

where $\sigma_{N,i}$ is a positive measure on $[-m, m]$, with total mass $\operatorname{tr}_N(B_N^2) - \operatorname{tr}_N(B_N)^2$. Similarly, the subordination function ω_1 from (3.5) can be written as

$$\omega_1(z) = z - \int_{\mathbb{R}} t dv(t) - \int_{\mathbb{R}} \frac{1}{z-t} d\sigma(t), \quad z \in \mathbb{C}^+.$$

The hypothesis that the empirical eigenvalue distribution of B_N converges to ν implies in particular $\lim_{N \rightarrow \infty} \operatorname{tr}_N(B_N) = \int_{\mathbb{R}} t dv(t)$. In addition, the fact that $\lim_{N \rightarrow \infty} \omega_{N,i} = \omega_1$ uniformly on compact subsets of $\mathbb{C} \setminus [-m, m]$ implies that σ is supported in $[-m, m]$ and that $\lim_{N \rightarrow \infty} \sigma_{N,i} = \sigma$ in the weak*-topology. Lemma 4.1, applied to the sequence $\rho_{N,i} = \sigma_{N,i} - \sigma$ yields nonnegative numbers $\{v_{N,i}\}_{N \geq s}$ such that $\lim_{N \rightarrow \infty} v_{N,i} = 0$ and

$$|\omega_{N,i}(z) - \omega_1(z)| < \left(1 + \frac{1}{(\Im z)^2}\right) v_{N,i} + \left| \int_{\mathbb{R}} t dv(t) - \operatorname{tr}_N(B_N) \right|, \quad z \in \mathbb{C}^+.$$

We can now estimate

$$\begin{aligned} \left| \mathbb{E}[R_N(z)_{ii}] - \frac{1}{\omega_1(z) - \theta_i} \right| &= \left| \frac{1}{\omega_{N,i}(z) - \theta_i} - \frac{1}{\omega_1(z) - \theta_i} \right| \\ &= \frac{|\omega_{N,i}(z) - \omega_1(z)|}{|\omega_{N,i}(z) - \theta_i| |\omega_1(z) - \theta_i|} \\ &< \frac{|\omega_{N,i}(z) - \omega_1(z)|}{|\Im z|^2} \\ &\leq a_N (1 + |z|)^2 \cdot \frac{2}{|\Im z|^4}, \end{aligned}$$

where

$$a_N = \max \left\{ v_{N,1}, \dots, v_{N,p}, \left| \int_{\mathbb{R}} t dv(t) - \operatorname{tr}_N(B_N) \right| \right\}.$$

The proposition follows. \square

COROLLARY 5.6. *We have*

$$\lim_{N \rightarrow \infty} \limsup_{y \rightarrow 0} \left| \int_{\mathbb{R}} \Delta_{i,N}(t + iy) h(t) dt \right| = 0, \quad i = 1, \dots, p.$$

PROOF. The preceding proposition allows us to apply Lemma 4.3 to obtain a positive constant c such that

$$\limsup_{y \downarrow 0} \left| \int_{\mathbb{R}} \Delta_{i,N}(t + iy) h(t) dt \right| \leq ca_N$$

for $N \geq p$ and $i = 1, \dots, p$. The corollary follows. \square

The preceding result, combined with (5.9), shows that (5.8) is equivalent to

$$(5.10) \quad \lim_{y \downarrow 0} \frac{1}{\pi} \Im \int_{\rho-\varepsilon}^{\rho+\varepsilon} \frac{h(t)}{\omega_1(t+iy) - \theta_i} dt = -\frac{\delta_{i_0 i}}{\omega'_1(\rho)}, \quad i = 1, \dots, p.$$

This is easily verified. Indeed, denote by Ω_y , $y > 0$, the rectangle with vertices $\rho \pm \varepsilon/2 \pm iy$. Calculus of residues yields

$$\frac{1}{2\pi i} \int_{\partial\Omega_y} \frac{1}{\omega_1(z) - \theta_i} dz = \frac{\delta_{i_0 i}}{\omega'_1(\rho)}, \quad i = 1, \dots, p.$$

On the other hand,

$$\begin{aligned} & \frac{1}{\pi} \Im \int_{\rho-\varepsilon}^{\rho+\varepsilon} \frac{h(t)}{\omega_1(t+iy) - \theta_i} dt \\ &= \frac{1}{2\pi i} \int_{\rho-\varepsilon}^{\rho+\varepsilon} \left[\frac{h(t)}{\omega_1(t+iy) - \theta_i} - \frac{h(t)}{\omega_1(t-iy) - \theta_i} \right] dt. \end{aligned}$$

Now we use the fact that $h = 1$ on $(\rho - \varepsilon/2, \rho + \varepsilon/2)$ to conclude that

$$\frac{1}{\pi} \Im \int_{\rho-\varepsilon}^{\rho+\varepsilon} \frac{h(t)}{\omega_1(t+iy) - \theta_i} dt + \frac{1}{2\pi i} \int_{\partial\Omega_y} \frac{1}{\omega_1(z) - \theta_i} dz,$$

is a sum of the following four integrals:

$$\begin{aligned} & \frac{1}{\pi} \Im \int_{\rho-\varepsilon}^{\rho-\varepsilon/2} \frac{h(t)}{\omega_1(t+iy) - \theta_i} dt, & \frac{1}{\pi} \Im \int_{\rho+\varepsilon/2}^{\rho+\varepsilon} \frac{h(t)}{\omega_1(t+iy) - \theta_i} dt, \\ & \frac{1}{2\pi i} \int_{\rho-\varepsilon/2+iy}^{\rho-\varepsilon/2-iy} \frac{1}{\omega_1(z) - \theta_i} dz, & \frac{1}{2\pi i} \int_{\rho+\varepsilon/2-iy}^{\rho+\varepsilon/2+iy} \frac{1}{\omega_1(z) - \theta_i} dz, \end{aligned}$$

all of which are easily seen to tend to zero as $y \downarrow 0$. This completes the proof of (5.10) and therefore of (5.5). We observe now that the proof of (5.5) for $i \neq i_0$ uses only the fact that $\omega_1(\rho) \neq \theta_i$. Therefore, switching the roles of A_N and B_N in this argument yields a proof of (5.6) and completes the proof of Part (3). of Theorem 2.1 in this case if $k = 1$ and $\ell = 0$. The case $\ell = 1$ and $k = 0$ follows by symmetry.

Assertion (4) of the theorem follows from (3) simply because $E_{X_N}((\rho - \varepsilon, \rho + \varepsilon))$ is a projection of rank one. Indeed, denote by $\{e_i\}_{i=1}^N$ the canonical basis in \mathbb{C}^N , so $A_N e_i = \theta_i e_i$. Let ξ be a unit vector in the range of $E_{X_N}((\rho - \varepsilon, \rho + \varepsilon))$, so $E_{X_N}((\rho - \varepsilon, \rho + \varepsilon))h = \langle h, \xi \rangle \xi$ for every $h \in \mathbb{C}^N$. Direct calculation shows that

$$P_N E_{X_N}((\rho - \varepsilon, \rho + \varepsilon)) P_N e_{i_0} = \sum_{i=1}^p \langle e_{i_0}, \xi \rangle \langle \xi, e_i \rangle e_i$$

and thus, almost surely, for all $\iota > 0$, there exists N_0 such that if $N \geq N_0$ and $\xi \in E_{X_N}((\rho - \varepsilon, \rho + \varepsilon))$, then

$$\left\| \sum_{i=1}^p \langle e_{i_0}, \xi \rangle \langle \xi, e_i \rangle e_i - \frac{1}{\omega'_1(\rho)} e_{i_0} \right\| < \iota.$$

In particular, we obtain $|\langle e_{i_0}, \xi \rangle|^2 - 1/\omega'_1(\rho)| < \iota$, which is precisely the first relation in (4). The case $k = 0, \ell = 1$ follows by symmetry.

Step B. In this step, we prove (3) and (4) in the general case of spikes with higher multiplicities and arbitrary values for k and ℓ . We use an idea from [16] to reduce the problem to the case considered in Step A. Given $N \geq p + q$, let η and δ be positive numbers such that the matrices

$$(5.11) \quad A_{N,\eta} = A_N + \text{Diag}(p\eta, (p - 1)\eta, \dots, \eta, \underbrace{0, \dots, 0}_{N-p}),$$

$$(5.12) \quad B_{N,\delta} = B_N + \text{Diag}(q\delta, (q - 1)\delta, \dots, \delta, \underbrace{0, \dots, 0}_{N-q})$$

have distinct spikes $\theta_i(\eta) = \theta_i + (p - i + 1)\eta$ and $\tau_j(\delta) = \tau_j + (q - j + 1)\delta$, respectively. The fact that ω_1 is increasing and continuous at ρ implies that, for sufficiently small η , there exist exactly k indices i_1, \dots, i_k such that the equations $\omega_1(t) = \theta_{i_n}(\eta)$ each have a solution $\rho_n = \rho_n(\eta) \in (\rho - \varepsilon, \rho + \varepsilon)$, $n = 1, 2, \dots, k$. Similarly, for sufficiently small δ there exist ℓ indices j_1, \dots, j_ℓ and ℓ values $\rho_{k+n} = \rho_{k+n}(\delta) \in (\rho - \varepsilon, \rho + \varepsilon)$ such that $\omega_2(\rho_{k+n}(\delta)) = \tau_{j_n}(\delta)$, $n = 1, \dots, \ell$. The numbers η and δ can be chosen such that the intervals $(\rho_n - 2\eta, \rho_n + 2\eta)$, $n = 1, 2, \dots, k + \ell$, are pairwise disjoint and contained in $(\rho - \varepsilon, \rho + \varepsilon)$. We conclude that the arguments of Step A hold with $X_{N,\eta,\delta} = A_{N,\eta} + U_N B_{N,\delta} U_N^*$, ρ_n , and η in place of X_N , ρ and ε , respectively. Thus,

$$\lim_{N \rightarrow \infty} \left\| P_N E_{X_{N,\eta,\delta}}((\rho_n - \eta, \rho_n + \eta)) P_N - \frac{1}{\omega'_1(\rho_n)} E_{A_{N,\eta}}(\{\omega_1(\rho_n)\}) \right\| = 0$$

almost surely for $n = 1, \dots, k + \ell$. We have

$$\sum_{n=1}^{k+\ell} E_{X_{N,\eta,\delta}}((\rho_n - \eta, \rho_n + \eta)) = E_{X_{N,\eta,\delta}}((\rho - \varepsilon, \rho + \varepsilon))$$

and also, noting that $E_{A_{N,\eta}}(\{\omega_1(\rho_n)\}) = 0$ for $n = k + 1, \dots, k + \ell$,

$$\sum_{n=1}^{k+\ell} E_{A_{N,\eta}}(\{\omega_1(\rho_n)\}) = E_{A_N}(\{\omega_1(\rho)\})$$

for small η . In addition, $1/\omega'_1(\rho_n)$ can be made arbitrarily close to $1/\omega'_1(\rho)$ by making η sufficiently small. We conclude that for any $\gamma > 0$, if $\eta, \delta > 0$ are sufficiently small, almost surely for all large N

$$\left\| P_N E_{X_{N,\eta,\delta}}((\rho - \varepsilon, \rho + \varepsilon)) P_N - \frac{1}{\omega'_1(\rho)} E_{A_N}(\{\omega_1(\rho)\}) \right\| < \gamma.$$

Clearly,

$$\|X_N - X_{N,\eta,\delta}\| \leq p\eta + q\delta.$$

An application of Lemma 4.10 shows that almost surely, there exists an N_0 depending only on ε such that if $N > N_0$, for any $\gamma > 0$, if $\eta, \delta > 0$ are sufficiently small (depending only on γ and not on N), then

$$(5.13) \quad \|E_{X_{N,\eta,\delta}}((\rho - \varepsilon, \rho + \varepsilon)) - E_{X_N}((\rho - \varepsilon, \rho + \varepsilon))\| < \iota.$$

The first inequality in (3) follows at once, and the second one is proved similarly.

We now verify assertion (4) when $\ell = 0$. Let $\xi^{(N)}$ be a unit vector in the range of $E_{X_N}((\rho - \varepsilon, \rho + \varepsilon))$. Since the quantity in (5.13) is small, we can find, almost surely for large N , unit vectors $\xi_{\eta,\delta}^{(N)} \in E_{X_{N,\eta,\delta}}((\rho - \varepsilon, \rho + \varepsilon))\mathbb{C}^N$ such that $\lim_{\delta+\eta \rightarrow 0} \|\xi^{(N)} - \xi_{\eta,\delta}^{(N)}\| = 0$, uniformly in N . It suffices therefore to prove that

$$\limsup_{N \rightarrow \infty} \left| \|E_{A_N}(\{\omega_1(\rho)\})\xi_{\eta,\delta}^{(N)}\|^2 - \frac{1}{\omega'_1(\rho)} \right|$$

can be made arbitrarily small for appropriate choices of η and δ . Write $\xi_{\eta,\delta}^{(N)} = \xi_1^{(N)} + \dots + \xi_k^{(N)}$ with $\xi_n^{(N)}$ in the range of $E_{X_{N,\eta,\delta}}((\rho_n - \eta, \rho_n + \eta))$, $n = 1, \dots, k$. The case of assertion (4) proved in Step A shows that for any $\eta, \delta > 0$ sufficiently small,

$$\lim_{N \rightarrow \infty} \|E_{A_{N,\eta}}(\{\omega_1(\rho_n)\})\xi_n^{(N)}\|^2 - \frac{\|\xi_n^{(N)}\|^2}{\omega'_1(\rho_n)} = 0, \quad n = 1, \dots, k.$$

We also have $\lim_{N \rightarrow \infty} \|E_{A_{N,\eta}}(\{\theta_i(\eta)\})\xi_n^{(N)}\| = 0$ for $i \notin \{i_1, \dots, i_k\}$. Since

$$E_{A_N}(\{\omega_1(\rho)\}) = E_{A_{N,\eta}}(\{\omega_1(\rho_1), \dots, \omega_1(\rho_k)\}),$$

the relation

$$\begin{aligned} & \left| \|E_{A_N}(\{\omega_1(\rho)\})\xi_{\eta,\delta}^{(N)}\|^2 - \frac{1}{\omega'_1(\rho)} \right| \\ & \leq \left| \sum_{n=1}^k \|E_{A_{N,\eta}}(\{\omega_1(\rho_n)\})\xi_n^{(N)}\|^2 - \frac{\|\xi_n^{(N)}\|^2}{\omega'_1(\rho_n)} \right| \\ & \quad + \sum_{n=1}^k \|\xi_n^{(N)}\|^2 \left| \frac{1}{\omega'_1(\rho_n)} - \frac{1}{\omega'_1(\rho)} \right| \\ & \quad + \sum_{m \neq n} \|E_{A_{N,\eta}}(\{\omega_1(\rho_m)\})\xi_n^{(N)}\|^2 \end{aligned}$$

implies

$$\limsup_{N \rightarrow \infty} \left| \|E_{A_N}(\{\omega_1(\rho)\})\xi_{\eta,\delta}^{(N)}\|^2 - \frac{1}{\omega'_1(\rho)} \right| \leq \max_{1 \leq n \leq k} \left| \frac{1}{\omega'_1(\rho_n)} - \frac{1}{\omega'_1(\rho)} \right|.$$

The desired conclusion follows by noting that $\omega'_1(\rho_n) = \omega'_1(\rho_n(\eta))$ can be made arbitrarily close to $\omega'_1(\rho)$. The second part of assertion (4) follows by symmetry. \square

The proofs of the versions of Theorem 2.5 for the positive line and for the unit circle follow the same outline. We avoid excessive repetition and only indicate the differences in the tools used throughout the proof.

5.2. *The multiplicative model* $X_N = A_N^{1/2} U_N B_N U_N^* A_N^{1/2}$. We use the notation from Section 2.2. As in the previous section, we assume that both A_N and B_N are diagonal matrices:

$$A_N = \text{Diag}(\theta_1, \dots, \theta_p, \alpha_1^{(N)}, \dots, \alpha_{N-p}^{(N)}),$$

$$B_N = \text{Diag}(\tau_1, \dots, \tau_q, \beta_1^{(N)}, \dots, \beta_{N-q}^{(N)}).$$

Since $\mu, \nu \neq \delta_0$, fix $\alpha \in \text{supp}(\mu) \setminus \{0\}$ and $\beta \in \text{supp}(\nu) \setminus \{0\}$. We use the following multiplicative decompositions:

$$A_N = A'_N A''_N = A''_N A'_N,$$

$$A'_N = \text{Diag}(\alpha, \dots, \alpha, \alpha_1^{(N)}, \dots, \alpha_{N-p}^{(N)}),$$

$$A''_N = \text{Diag}(\theta_1/\alpha, \dots, \theta_p/\alpha, 1, \dots, 1),$$

$$B_N = B'_N B''_N = B''_N B'_N,$$

$$B'_N = \text{Diag}(\beta, \dots, \beta, \beta_1^{(N)}, \dots, \beta_{N-q}^{(N)}),$$

$$B''_N = \text{Diag}(\tau_1/\beta, \dots, \tau_q/\beta, 1, \dots, 1).$$

As before, we write $A''_N = P_N^* \Theta P_N + I_N - P_N^* P_N$, where P_N is defined as in Section 5.1 and

$$\Theta = \text{Diag}(\theta_1/\alpha, \dots, \theta_p/\alpha).$$

Similarly, $B''_N = Q_N^* T Q_N + I_N - Q_N^* Q_N$, where Q_N is defined as in Section 5.1 and

$$T = \text{Diag}(\tau_1/\beta, \dots, \tau_q/\beta).$$

We discuss first the behavior of the model $A_N^{1/2} U_N B'_N U_N^* A_N^{1/2}$ with spikes only on A_N . The essential step is a reduction to a convergence problem for a sequence of matrices of fixed size.

5.2.1. *Reduction to the almost sure convergence of a $p \times p$ matrix.* Recall that $K = \text{supp}(\mu \boxtimes \nu)$. Corollary 2.2 of [21] yields the existence of positive random variables $\{\delta_N\}_{N \in \mathbb{N}}$ such that

$$\sigma((A'_N)^{1/2} U_N B'_N U_N^* (A'_N)^{1/2}) \subseteq K_{\delta_N}$$

and $\lim_{N \rightarrow \infty} \delta_N = 0$ almost surely.

We argue first that, in case $0 \notin \text{supp}(\mu \boxtimes \nu)$, it follows that X_N is almost surely invertible for large N . Indeed, in this case, $0 \notin \text{supp}(\mu) \cup \text{supp}(\nu)$. Therefore, A_N and B_N are invertible for large N , and thus so is X_N . This observation allows us to restrict the analysis to nonzero eigenvalues of X_N .

Denote $X'_N = A_N^{1/2} U_N B'_N U_N^* A_N^{1/2}$. Fix $z \in \mathbb{C} \setminus (K_{\delta_N} \cup \{0\})$ such that the matrix $zI_N - (A'_N)^{1/2} U_N B'_N U_N^* (A'_N)^{1/2}$ is invertible. Using Sylvester's identity $\det(I - XY) = \det(I - YX)$, we obtain for large N

$$\begin{aligned} \det(zI_N - X'_N) &= z^N \det(I_N - z^{-1} A''_N (A'_N)^{1/2} U_N^* B'_N U_N (A'_N)^{1/2}) \\ &= z^N \det(I_N - A''_N + A''_N (I_N - z^{-1} (A'_N)^{1/2} U_N B'_N U_N^* (A'_N)^{1/2})) \\ &= \det((I_N - A''_N)(I_N - z^{-1} (A'_N)^{1/2} U_N B'_N U_N^* (A'_N)^{1/2})^{-1} + A''_N) \\ &\quad \times \det(zI_N - (A'_N)^{1/2} U_N B'_N U_N^* (A'_N)^{1/2}). \end{aligned}$$

The matrix $(I_N - A''_N)(I_N - \frac{1}{z} (A'_N)^{1/2} U_N B'_N U_N^* (A'_N)^{1/2})^{-1} + A''_N$ is of the form

$$\begin{bmatrix} F_N(z) & * \\ 0 & I_{N-p} \end{bmatrix},$$

where F_N is the analytic function with values in $M_p(\mathbb{C})$ defined on $\mathbb{C} \setminus (K_{\delta_N} \cup \{0\})$ by

$$(5.14) \quad F_N(z) := (I_p - \Theta) P_N \left(I_N - \frac{1}{z} (A'_N)^{1/2} U_N B'_N U_N^* (A'_N)^{1/2} \right)^{-1} P_N^* + \Theta,$$

and Θ is the diagonal $p \times p$ matrix defined earlier. Thus, for large N , the nonzero eigenvalues of X'_N outside K_{δ_N} are precisely the zeros of $\det(F_N)$ in that open set. As in Section 5.1, the random matrix functions sequence $\{F_N\}_N$ converges a.s. to a diagonal deterministic $p \times p$ matrix function:

5.2.2. *Convergence of F_N .* We start with the analogue of Proposition 5.1.

PROPOSITION 5.7. *Fix a positive integer p , and let C_N and D_N be deterministic nonnegative diagonal $N \times N$ matrices with uniformly bounded norms such that, for all $i = 1, \dots, p$, $(C_N)_{ii} \neq 0$ and the limits*

$$\eta_i = \lim_{N \rightarrow \infty} (C_N)_{ii}$$

exist. Suppose that the empirical eigenvalue distributions of C_N and D_N converge weakly to μ and ν , respectively. Then the resolvent

$$R_N(z) = (zI_N - C_N^{1/2} U_N D_N U_N^* C_N^{1/2})^{-1}, \quad z \in \mathbb{C} \setminus \mathbb{R}$$

satisfies

$$\lim_{N \rightarrow \infty} P_N \mathbb{E}[zR_N(z)] P_N^* = \text{Diag} \left(\frac{1}{1 - \eta_1 \omega_1(z^{-1})}, \dots, \frac{1}{1 - \eta_p \omega_1(z^{-1})} \right).$$

PROOF. We consider without loss of generality elements $z \in \mathbb{C}^+$. If C_N is invertible, then Lemma 4.7(2) applied to $b = zC_N^{-1}$ implies that $\mathbb{E}[R_N(z)]$ is diagonal. If C_N is not invertible, then

$$\lim_{\varepsilon \downarrow 0} \mathbb{E}[(zI_N - (C_N + \varepsilon I_N)^{1/2} U_N D_N U_N^* (C_N + \varepsilon I_N)^{1/2})^{-1}] = \mathbb{E}[R_N(z)].$$

The limit of diagonal matrices is diagonal, so $\mathbb{E}[R_N(z)]$ is diagonal. Define

$$(5.15) \quad \omega_{N,i}(z) := \frac{1}{(C_N)_{ii}} \left(1 - \frac{z}{\mathbb{E}[R_N(z^{-1})]_{ii}} \right), \quad 1 \leq i \leq p.$$

We prove the uniform convergence on compact subsets of $\mathbb{C} \setminus \mathbb{R}^+$ of the sequences $\{\omega_{N,i}\}_{N \geq p}$ of analytic functions to ω_1 . The multiplicative counterpart of Lemma 5.2 is as follows.

LEMMA 5.8. *Assume that $C_N \geq \varepsilon I_N$ for some $\varepsilon > 0$. We have*

$$\lim_{N \rightarrow \infty} \|z\mathbb{E}[R_N(z)] - (I_N - \omega_{N,i}(z^{-1})C_N)^{-1}\| = 0, \quad z \in \mathbb{C} \setminus \mathbb{R}, i \in \{1, \dots, p\}.$$

PROOF. For $z \in \mathbb{C}^+$, define

$$\Omega_N(z) = (C_N)^{-1} \mathbb{E}[R_N(z^{-1})]^{-1} = \mathbb{E}[(zC_N)^{-1} - U_N D_N U_N^*]^{-1}.$$

This function is well defined by Lemma 4.6, and the second equality is justified by Lemma 4.7(2). We apply Lemma 4.7(1) with $b = (zC_N)^{-1}$ to obtain

$$\begin{aligned} & Y(\Omega_N(z) - (zC_N)^{-1}) - (\Omega_N(z) - (zC_N)^{-1})Y \\ &= \Omega_N(z) \mathbb{E}[\left((zC_N)^{-1} - U_N D_N U_N^* - \Omega_N(z)^{-1} \right) \\ & \quad \times (Y(zC_N)^{-1} - (zC_N)^{-1}Y) \\ & \quad \times \left((zC_N)^{-1} - U_N D_N U_N^* - \Omega_N(z)^{-1} \right) \Omega_N(z). \end{aligned}$$

Consider arbitrary norm one vectors $h, k \in \mathbb{C}^N$ and an Y of rank one to conclude the existence of rank one projections p_1, p_2 and rank 2 projections q_1, q_2 such that

$$\begin{aligned} & |k^*(Y(\Omega_N(z) - (zC_N)^{-1}) - (\Omega_N(z) - (zC_N)^{-1})Y)h| \\ & \leq \|\Omega_N(z)\|^2 \|Y(zC_N)^{-1} - (zC_N)^{-1}Y\| \\ & \quad \times \mathbb{E}[\|p_1((zC_N)^{-1} - U_N D_N U_N^* - \Omega_N(z)^{-1})q_1\|^2]^{1/2} \\ & \quad \times \mathbb{E}[\|q_2((zC_N)^{-1} - U_N D_N U_N^* - \Omega_N(z)^{-1})p_2\|^2]^{1/2}. \end{aligned}$$

Lemma 4.6 yields $\|\Omega_N(z)\| < \|(zC_N)^{-1}\| + \|D_N\| + 4c\|(1/\Im(1/z))C_N\|$, with $c \in (0, +\infty)$. Remark 4.13 provides estimates for the last two factors. The estimate $\|Y(zC_N)^{-1} - (zC_N)^{-1}Y\| < 2\|Y\||z^{-1}|\varepsilon^{-1}$ is obvious. Thus,

$$|k^*(Y(\Omega_N(z) - (zC_N)^{-1}) - (\Omega_N(z) - (zC_N)^{-1})Y)h| \leq \frac{C(z, \varepsilon)}{N} \|Y\|$$

for some constant $C(z, \varepsilon)$ independent of N . The (i, i) entry of the matrix $\Omega_N(z) - (zC_N)^{-1}$ is precisely $e_i^*(\Omega_N(z) - (zC_N)^{-1})e_i$, which belongs to the numerical range of $\Omega_N(z) - (zC_N)^{-1}$. Lemma 4.9 yields

$$\|\Omega_N(z) - (zC_N)^{-1} - (e_i^*(\Omega_N(z) - (zC_N)^{-1})e_i)I_N\| < 2\frac{C(z, \varepsilon)}{N}.$$

Since C_N is diagonal, the lemma follows by letting $N \rightarrow \infty$. \square

The proof of Proposition 5.7 when C_N is bounded from below by a positive multiple of I_N is now completed by an application of the above lemma. Indeed, using Biane’s subordination formula (2.4) and the asymptotic freeness result of Voiculescu [39], we obtain

$$\lim_{N \rightarrow \infty} \operatorname{tr}_N(z\mathbb{E}[R_N(z)]) = 1 + \psi_{\mu \boxtimes \nu}(1/z) = 1 + \psi_{\mu}(\omega_1(1/z)).$$

Clearly, $\operatorname{tr}_N((I_N - \omega_{N,i}(1/z)C_N)^{-1}) \rightarrow 1 + \psi_{\mu}(\lim_{N \rightarrow \infty} \omega_{N,i}(1/z))$. The result follows by analytic continuation. The general case follows by replacing a noninvertible C_N by $C_N + \varepsilon I_N$. The approximation is uniform in N , so a normal family argument yields the desired result as $\varepsilon \rightarrow 0$. \square

Observe that

$$F_N(z) = (I_p - \Theta)P_N(zR'_N(z))P_N^* + \Theta,$$

where R'_N denotes the resolvent of $(A'_N)^{1/2}U_N B'_N U_N^* (A'_N)^{1/2}$. An application of Proposition 5.7 to $C_N = A'_N$ and $D_N = B'_N$, Remark 4.13, as well as of Lemma 3.2, yield the following result. We leave the details, similar to the ones in the proof of Proposition 5.3, to the reader.

PROPOSITION 5.9. *Almost surely, the sequence $\{F_N\}_N$ converges uniformly on the compact subsets of $\overline{\mathbb{C}} \setminus K$ to the analytic function F defined on $\overline{\mathbb{C}} \setminus K$ by*

$$F(z) = \operatorname{diag}\left(\left(1 - \frac{\theta_j}{\alpha}\right)\frac{1}{1 - \alpha\omega_1(z^{-1})} + \frac{\theta_j}{\alpha}\right)_{j=1}^p.$$

5.2.3. *Proofs of the main results for the positive multiplicative model.*

PROOF OF THEOREM 2.5(1)–(2)—EIGENVALUE BEHAVIOR. *Step 1.* We prove our result for the model $A_N^{1/2}U_N B'_N U_N^* A_N^{1/2}$ in which only A_N has spikes. We consider the almost sure event, whose existence is guaranteed by Proposition 5.9, on which there exist a sequence $\{\delta_N\}_N \subset (0, +\infty)$ converging to zero such that:

- $\sigma(A_N^{1/2}U_N B'_N U_N^* A_N^{1/2}) \subseteq K_{\delta_N}$, and

- the sequence $\{F_N(z)\}_{N \geq p}$ converges to

$$F(z) = \text{diag} \left(\left(1 - \frac{\theta_j}{\alpha} \right) \frac{1}{1 - \alpha \omega_1(z^{-1})} + \frac{\theta_j}{\alpha} \right)_{j=1}^p$$

uniformly on the compact subsets of $\overline{\mathbb{C}} \setminus K$.

On this event, we apply Lemma 4.5 with $\gamma = \mathbb{R}$, the sequence $\{F_N\}_{N \geq p}$ and its uniform on compacts limit F . We argue first that the function $F_N(z)$ given by equation (5.14) is invertible for $z \notin \mathbb{R}$. Indeed, the relations preceding (5.14) imply that, if $F_N(z)$ is not invertible, then z is an eigenvalue of the selfadjoint matrix X'_N , and hence it is a real number. This verifies hypothesis 2. Hypotheses 1 and 3 follow from Proposition 5.9. Finally, $F(\infty) = I_p$,

$$(F'(z))_{jj} = \frac{\omega'_1(1/z)(\theta_j - \alpha)}{z^2(1 - \alpha \omega_1(1/z))^2},$$

and the zeros of ω'_1 are simple by the Julia–Carathéodory theorem. Thus, Lemma 4.5 applies to F_N and F .

For almost every $\delta > 0$, the boundary points of K_δ are not zeros of $\det(F)$. When this condition is satisfied, Lemma 4.5 yields precisely the conclusion of Theorem 2.5(1)–(2), when $q = 0$. Indeed, as noted above, the nonzero eigenvalues of X'_N in $\mathbb{C} \setminus K_\delta$ are exactly the zeros of $\det(F_N)$, and the set of points z such that $F(z)$ is not invertible is precisely $\bigcup_{i=1}^p v_i^{-1}(\{1/\theta_i\})$. This completes the first step.

Step 2. This is completely analogous to the reasoning from the second step of the proof of Theorem 2.1(1)–(2). We omit the details. \square

PROOF OF THEOREM 2.5, PARTS 3 AND 4—EIGENSPACE BEHAVIOR. *Step A.* We assume first that $\theta_1 > \dots > \theta_p > 0$, $\tau_1 > \dots > \tau_q > 0$, $\ell = 0$ and $k = 1$. Step A of the proof of Theorem 2.1 is modified as follows: X_N is replaced by $A_N^{1/2} U_N B_N U_N^* A_N^{1/2}$, the analogue of Lemma 5.4 holds with the constant C replaced by $\sup_N \|A_N\| \|B_N\|$, and Proposition 5.5 is replaced by the following statement.

PROPOSITION 5.10. *There is a polynomial P with nonnegative coefficients, a sequence $\{a_N\}_N$ of nonnegative real numbers converging to zero when N goes to infinity and some nonnegative integer number t , such that for every $i = 1, \dots, p$ and $z \in \mathbb{C} \setminus \mathbb{R}$,*

$$(5.16) \quad \mathbb{E}[R_N(z)_{ii}] = \frac{1}{z} \frac{1}{1 - \theta_i \omega_1(1/z)} + \Delta_{i,N}(z),$$

with

$$|\Delta_{i,N}(z)| \leq (1 + |z|)^t P(|\Im z|^{-1}) a_N.$$

PROOF. We set

$$\omega_{N,i}(z) = \frac{1}{\theta_i} \left(1 - \frac{z}{\mathbb{E}[R_N(\frac{1}{z})]_{ii}} \right), \quad z \in \mathbb{C} \setminus [0, +\infty).$$

As established in Proposition 5.7, $\lim_{N \rightarrow \infty} \mathbb{E}[zR_N(z)]_{ii} = (1 - \theta_i \omega_1(1/z))^{-1}$. It follows that $\omega_{N,i}$ converges to ω_1 uniformly on compacts of $\mathbb{C} \setminus [0, +\infty)$. Clearly, $\omega_{N,i}$ is also defined on a neighborhood of zero. Note that

$$\lim_{y \rightarrow +\infty} \omega_{N,i}(-1/iy) = \frac{1}{\theta_i} \left(1 - \frac{1}{\lim_{y \rightarrow +\infty} \mathbb{E}[-iyR_N(-iy)]_{ii}} \right) = \frac{1}{\theta_i} \left(1 - \frac{1}{1} \right) = 0$$

and

$$\lim_{y \rightarrow +\infty} iy\omega_{N,i}\left(\frac{-1}{iy}\right) = -\frac{1}{\theta_i} \lim_{y \rightarrow +\infty} \frac{\mathbb{E}[iyX_N(iy + X_N)^{-1}]_{ii}}{\mathbb{E}[iy(iy + X_N)^{-1}]_{ii}} = -\frac{1}{\theta_i} \mathbb{E}[X_N]_{ii}.$$

In addition, since $\|X_N\| \leq \|A_N\| \|B_N\|$ which is uniformly bounded, the map $z \mapsto \omega_{N,i}(-1/z)$ is analytic and real on the complement of an interval $[-m, 0]$, with $m = \sup_N \|A_N\| \|B_N\|$. Thus, the maps $z \mapsto \omega_{N,i}(-1/z)$ and $z \mapsto \omega_1(-1/z)$ are Nevanlinna maps (3.3), and hence can be represented as

$$\omega_{N,i}\left(\frac{-1}{z}\right) = \int_{[-m,0]} \frac{1}{t-z} d\Phi_{N,i}(t), \quad z \in \mathbb{C}^+,$$

and

$$\omega_1\left(\frac{-1}{z}\right) = \int_{[-m,0]} \frac{1}{t-z} d\Phi(t), \quad z \in \mathbb{C}^+.$$

Here, $\Phi_{N,i}, \Phi$ are positive measures on $[-m, 0]$, $\Phi_{N,i}([-m, 0]) = \frac{1}{\theta_i} \mathbb{E}[X_N]_{ii}$, and $\Phi([-m, 0]) = \frac{\int_{\mathbb{R}} t d(\mu \boxtimes \nu)(t)}{\int_{\mathbb{R}} t d\mu(t)} = \int_{\mathbb{R}} t d\nu(t)$. Thus, Lemma 4.1 applies to $\rho_{N,i} = \Phi_{N,i} - \Phi$ to allow the estimate

$$\left| \omega_{N,i}\left(\frac{-1}{z}\right) - \omega_1\left(\frac{-1}{z}\right) \right| < v_{N,i} \left(1 + \frac{1}{(\Im z)^2} \right).$$

We have

$$\begin{aligned} & \left| \mathbb{E}[R_N(z)]_{ii} - \frac{1}{z} \cdot \frac{1}{1 - \theta_i \omega_1(\frac{1}{z})} \right| \\ &= \left| \frac{1}{z} \cdot \frac{1}{1 - \theta_i \omega_{N,i}(\frac{1}{z})} - \frac{1}{z} \cdot \frac{1}{1 - \theta_i \omega_1(\frac{1}{z})} \right| \\ &= \frac{\theta_i}{|z|} \frac{|\omega_{N,i}(\frac{1}{z}) - \omega_1(\frac{1}{z})|}{|(1 - \theta_i \omega_1(\frac{1}{z}))(1 - \theta_i \omega_{N,i}(\frac{1}{z}))|} \\ &< \frac{1}{|\Im z|^3} \left(1 + \frac{1}{(\Im z)^2} \right) \frac{(|z| + m)^4}{\theta_i \Phi_{N,i}([-m, 0]) \Phi([-m, 0])} v_{N,i}. \end{aligned}$$

The proposition follows. \square

To complete the argument of Step A, it suffices now to observe that the residue of the function $1/(z(1 - \theta_i \omega_1(z^{-1})))$ at ρ is equal to

$$\delta_{\omega_1(1/\rho), 1/\theta_i} \frac{\omega_1(1/\rho)\rho}{\omega'_1(1/\rho)}, \quad i = 1, \dots, p.$$

Step B: We use the same perturbation argument as in Step B of the proof of Theorem 2.1. We reduce the problem to the case of a spike with multiplicity one, to which we apply Step A. The only change from the argument in Step B of Theorem 2.1 comes from the form of the subordination functions. We use perturbations (5.11) and (5.12) and define $X_{N,\delta,\eta} = A_{N,\delta}^{1/2} U_N B_{N,\eta} U_N^* A_{N,\delta}^{1/2}$. The quantity $\|X_{N,\delta,\eta} - X_N\|$ tends to zero uniformly in N as $\delta + \eta \rightarrow 0$. The details are omitted. \square

5.3. *The unitary multiplicative model* $X_N = A_N U_N B_N U_N^*$. We use the notation from Section 2.3. The tools used are identical to the ones used in the analysis of the positive model $X_N = A_N^{1/2} U_N B_N U_N^* A_N^{1/2}$. However, the domains of definition of the analytic transforms involved are different. We indicate the relevant differences. Choose $\alpha, \beta \in \mathbb{T}$ such that $1/\alpha \in \text{supp}(\mu)$ and $1/\beta \in \text{supp}(\nu)$. The reduction to the almost sure convergence of a $p \times p$ matrix is performed the same way, and the same concentration inequality holds [this time with Lipschitz constant $\frac{2}{(1-|z|)^2}$] in Lemma 4.11. The counterparts of Propositions 5.7 and 5.9 hold, but in Proposition 5.9 we must consider $z \in \mathbb{C} \setminus \mathbb{T}$. The resolvent R_N is defined by $R_N(z) = (zI_N - A_N U_N B_N U_N^*)^{-1}$. The function $\omega_{N,i}$ defined by

$$\omega_{N,i}(z) = \frac{1}{(A_N)_{ii}} \left(1 - \frac{z}{\mathbb{E}[R_N(z^{-1})]_{ii}} \right), \quad z \in \mathbb{D},$$

is easily seen to map \mathbb{D} into itself and fix the origin. Indeed, $|(A_N)_{ii}| = |(A'_N)_{ii}| = 1$. In the unitary version of Lemma 5.8, no supplementary condition on A'_N is required, and for Ω_N defined as in the proof of Lemma 5.8, the estimate becomes $\|\Omega_N(z)\| < 2/|z|$ if $|z| < 1$. The estimates for the corresponding resolvents are provided by Lemma 4.6.

5.3.1. *Proofs of the main results for the unitary multiplicative model.*

PROOF OF THEOREM 2.5(1)–(2)—EIGENVALUE BEHAVIOR. We must now apply Lemma 4.5 with $\gamma = \mathbb{T}$. It will be applied to $\gamma = \mathbb{T}$, the sequence $\{F_N(z)\}_N$ defined by

$$F_N(z) = z(I_p - \Theta) P_N (zI_N - A'_N U_N B'_N U_N^*)^{-1} P_N^* + \Theta, \quad z \in \mathbb{C} \setminus \mathbb{T},$$

and the limit F provided by Proposition 5.9. Observe that $F_N(z)$ is invertible for $z \notin \mathbb{T}$. Indeed, it is easy to see that, if $F_N(z)$ is not invertible, z belongs to the

spectrum of the unitary operator $A_N U_N B'_N U_N^*$. The convergence of F_N to F follows from the appropriate version of Proposition 5.9. Clearly, $F(z)$ is diagonal and, again by the Julia–Carathéodory theorem, this time applied to the disk, its diagonal entries have only simple zeros. The remainder of the argument requires no further adjustments. \square

PROOF OF THEOREM 2.5(3)–(4)—EIGENSPACE BEHAVIOR. The relevant changes for this part of the proof occur in Proposition 5.10, where $(1 - |z|)^{-1}$ must be used instead of $|\Im z|^{-1}$ and an application of Lemma 4.2 in place of Lemma 4.1. Also, the perturbations (5.11) and (5.12) are applied to the arguments of θ_i and τ_j , respectively. \square

Acknowledgments. The work on this paper was initiated while Serban T. Belinschi was visiting the Institute of Mathematics of Toulouse as “Maître de conférence invité” and largely completed while all four authors visited the Fields Institute during the 2013 special programme on noncommutative distributions. We gratefully acknowledge the hospitality of the institutions involved. We also thank Charles Bordenave for providing us the simulations from Examples 2.3 and 2.4

REFERENCES

- [1] AKHIEZER, N. I. (1965). *The Classical Moment Problem and Some Related Questions in Analysis*. Hafner Publishing, New York. MR0184042
- [2] ANDERSON, G. W., GUIONNET, A. and ZEITOUNI, O. (2010). *An Introduction to Random Matrices. Cambridge Studies in Advanced Mathematics 118*. Cambridge Univ. Press, Cambridge. MR2760897
- [3] BAI, Z. and DING, X. (2012). Estimation of spiked eigenvalues in spiked models. *Random Matrices Theory Appl.* **1** 1150011, 21. MR2934717
- [4] BAI, Z. and YAO, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal.* **106** 167–177. MR2887686
- [5] 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. MR2165575
- [6] BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680
- [7] BELINSCHI, S. T. (2006). A note on regularity for free convolutions. *Ann. Inst. Henri Poincaré Probab. Stat.* **42** 635–648.
- [8] BELINSCHI, S. T. (2008). The Lebesgue decomposition of the free additive convolution of two probability distributions. *Probab. Theory Related Fields* **142** 125–150.
- [9] BELINSCHI, S. T. and BERCOVICI, H. (2005). Partially defined semigroups relative to multiplicative free convolution. *Int. Math. Res. Not.* **2** 65–101. MR2128863
- [10] BELINSCHI, S. T. and BERCOVICI, H. (2007). A new approach to subordination results in free probability. *J. Anal. Math.* **101** 357–365.
- [11] BELINSCHI, S. T., POPA, M. and VINNIKOV, V. (2012). Infinite divisibility and a non-commutative Boolean-to-free Bercovici–Pata bijection. *J. Funct. Anal.* **262** 94–123. MR2852257
- [12] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521.

- [13] BERCOVICI, H. and VOICULESCU, D. (1993). Free convolution of measures with unbounded support. *Indiana Univ. Math. J.* **42** 733–773. [MR1254116](#)
- [14] BIANE, P. (1998). Processes with free increments. *Math. Z.* **227** 143–174. [MR1605393](#)
- [15] BIRMAN, M. S. and SOLOMYAK, M. Z. (1973). Double Stieltjes operator integrals. III. *Prob. Math. Phys., Leningrad Univ.* **6** 27–53.
- [16] CAPITAINE, M. (2013). Additive/multiplicative free subordination property and limiting eigenvectors of spiked additive deformations of Wigner matrices and spiked sample covariance matrices. *J. Theoret. Probab.* **26** 595–648. [MR3090543](#)
- [17] CAPITAINE, M. (2014). Exact separation phenomenon for the eigenvalues of large information-plus-noise type matrices, and an application to spiked models. *Indiana Univ. Math. J.* **63** 1875–1910. [MR3298725](#)
- [18] CAPITAINE, M. and DONATI-MARTIN, C. (2007). Strong asymptotic freeness for Wigner and Wishart matrices. *Indiana Univ. Math. J.* **56** 767–803. [MR2317545](#)
- [19] CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.* **37** 1–47. [MR2489158](#)
- [20] CAPITAINE, M., DONATI-MARTIN, C., FÉRAL, D. and FÉVRIER, M. (2011). Free convolution with a semi-circular distribution and eigenvalues of spiked deformations of Wigner matrices. *Electron. J. Probab.* **16** 1750–1792. [MR1488333](#)
- [21] COLLINS, B. and MALE, C. (2014). The strong asymptotic freeness of Haar and deterministic matrices. *Ann. Sci. Éc. Norm. Supér. (4)* **47** 147–163. [MR3205602](#)
- [22] FÉRAL, D. and PÉCHÉ, S. (2007). The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.* **272** 185–228. [MR2291807](#)
- [23] FULTON, W. (1997). Eigenvalues of sums of Hermitian matrices (after A. Klyachko). *Astérisque* **252** 255–269.
- [24] FÜREDI, Z. and KOMLÓS, J. (1981). The eigenvalues of random symmetric matrices. *Combinatorica* **1** 233–241. [MR0637828](#)
- [25] GARNETT, J. B. (1981). *Bounded Analytic Functions. Pure and Applied Mathematics* **96**. Academic Press, New York. [MR0628971](#)
- [26] GESZTESY, F. and TSEKANOVSKII, E. (2000). On matrix-valued Herglotz functions. *Math. Nachr.* **218** 61–138. [MR1784638](#)
- [27] GUSTAFSON, K. E. and RAO, K. M. D. (1997). *Numerical Range*. Springer, New York. [MR1417493](#)
- [28] HAAGERUP, U. and THORBJØRNSSEN, S. (2005). A new application of random matrices: $\text{Ext}(C_{\text{red}}^*(F_2))$ is not a group. *Ann. of Math. (2)* **162** 711–775. [MR2183281](#)
- [29] JOHNSTONE, I. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. [MR1863961](#)
- [30] KARGIN, V. (2011). Subordination of the resolvent for a sum of random matrices. Preprint. Available at [arXiv:1205.0993\[math.PR\]](#).
- [31] LOUBATON, P. and VALLET, P. (2011). Almost sure localization of the eigenvalues in a Gaussian information plus noise model—Application to the spiked models. *Electron. J. Probab.* **16** 1934–1959. [MR2851051](#)
- [32] PÉCHÉ, S. (2006). The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields* **134** 127–173. [MR2221787](#)
- [33] PIZZO, A., RENFREW, D. and SOSHNIKOV, A. (2013). On finite rank deformations of Wigner matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 64–94.
- [34] RAO, N. R. and SILVERSTEIN, J. W. (2010). Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples. *IEEE J. Sel. Top. Signal Process.* **4** 468–480.
- [35] REMMERT, R. (1984). *Funktionentheorie. I*. Springer, Berlin. [MR0753290](#)

- [36] SPEICHER, R. (1993). Free convolution and the random sum of matrices. *Publ. Res. Inst. Math. Sci.* **29** 731–744. [MR1245015](#)
- [37] VOICULESCU, D. (1986). Addition of certain noncommuting random variables. *J. Funct. Anal.* **66** 323–346. [MR0839105](#)
- [38] VOICULESCU, D. (1987). Multiplication of certain noncommuting random variables. *J. Operator Theory* **18** 223–235.
- [39] VOICULESCU, D. (1991). Limit laws for random matrices and free products. *Invent. Math.* **104** 201–220.
- [40] VOICULESCU, D. (1993). The analogues of entropy and of Fisher’s information measure in free probability theory. I. *Comm. Math. Phys.* **155** 71–92.
- [41] VOICULESCU, D. (2000). The coalgebra of the free difference quotient and free probability. *Internat. Math. Res. Notices* 79–106.
- [42] VOICULESCU, D. V., DYKEMA, K. J. and NICA, A. (1992). *Free Random Variables. CRM Monograph Series 1*. Amer. Math. Soc., Providence, RI. [MR1217253](#)

S. T. BELINSCHI
M. CAPITAINE
INSTITUT DE MATHÉMATIQUES DE TOULOUSE
118 RTE DE NARBONNE
F-31062 TOULOUSE CEDEX 09
FRANCE
E-MAIL: serban.belinschi@math.univ-toulouse.fr
mireille.capitaine@math.univ-toulouse.fr

H. BERCOVICI
DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
BLOOMINGTON, INDIANA 47405
USA
E-MAIL: bercovic@indiana.edu

M. FÉVRIER
LABORATOIRE DE MATHÉMATIQUES D’ORSAY
UNIVERSITÉ PARIS-SUD, CNRS
UNIVERSITÉ PARIS-SACLAY
91405 ORSAY
FRANCE
E-MAIL: maxime.fevrier@math.u-psud.fr