Journal of Multivariate Analysis 192 (2022) 105051

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Edge statistics of large dimensional deformed rectangular matrices

Xiucai Ding [a,*,1], Fan Yang [b,1]

[a] Department of Statistics, University of California, Davis, United States of America
[b] Department of Statistics and Data Science, University of Pennsylvania, United States of America

A R T I C L E I N F O

Article history:
Received 27 October 2021
Received in revised form 24 May 2022
Accepted 24 May 2022
Available online 3 June 2022

AMS 2020 subject classifications:
primary 60B20
secondary 46L54
62H12

Keywords:
Interference-plus-noise matrix
Rectangular free convolution
Edge statistics

A B S T R A C T

We consider the edge statistics of large dimensional deformed rectangular matrices of the form $Y_t = Y + \sqrt{t}X$, where Y is a $p \times n$ deterministic signal matrix whose rank is comparable to n, X is a $p \times n$ random noise matrix with i.i.d. entries of mean zero and variance n^{-1}, and $t > 0$ gives the noise level. This model is referred to as the interference-plus-noise matrix in the study of massive multiple-input multiple-output (MIMO) system, which belongs to the category of the so-called signal-plus-noise model. For the case $t = 1$, the spectral statistics of this model have been studied to a certain extent in the literature (Dozier and Silverstein, 2007[17,18]; Vallet et al., 2012). In this paper, we study the singular value and singular vector statistics of Y_t around the right-most edge of the spectrum in the harder regime $n^{-1/3} \ll t \ll 1$. This regime is harder than the $t = 1$ case, because on the one hand, the edge behavior of the empirical spectral distribution (ESD) of $YY^\top$ has a strong effect on the edge statistics of $Y_tY_t^\top$ for a "small" $t \ll 1$, while on the other hand, the edge eigenvalue behavior of $Y_tY_t^\top$ is not merely a perturbation of that of $YY^\top$ for a "large" $t \gg n^{-1/3}$. Under certain regularity assumptions on Y, we prove the Tracy–Widom law for the edge eigenvalues, the eigenvalue rigidity, and eigenvector delocalization for the matrices $Y_tY_t^\top$ and $Y_t^\top Y_t$. These results can be used to estimate and infer the massive MIMO system. To prove the main results, we analyze the edge behavior of the asymptotic ESD of $Y_tY_t^\top$ and establish optimal local laws on its resolvent. These results are of independent interest, and can be used as important inputs for many other problems regarding the spectral statistics of Y_t.

1. Introduction

Large dimensional signal-plus-noise matrices are common objects in many scientific fields, such as signal processing [3,49], image denoising [44,48], wireless communications [51,53] and biology [27,56]. In these applications, researchers are interested in the estimation and inference of some deterministic matrix, known as the signal matrix, from its noisy observation. Specifically, we consider matrices of the form

$$Y_t = Y + \sqrt{t}X, \tag{1}$$

* Corresponding author.
E-mail address: xcading@ucdavis.edu (X. Ding).

1 The two authors, Xiucai Ding and Fan Yang, contribute equally to the paper in all aspects.

https://doi.org/10.1016/j.jmva.2022.105051

where Y is a $p \times n$ deterministic signal matrix, X is a white noise matrix whose entries x_{ij} are i.i.d. random variables of mean zero and variance n^{-1}, and $t > 0$ represents the noise level. In this paper, we consider the high dimensional setting where p is comparable to n.

There have been a lot of theoretical studies of this model in the literature by imposing various structural assumptions on Y. Among them, the most popular one is perhaps the low rank structure assumption [3–5,8,10,28,40,55]. Towards this direction, it is assumed that Y is a low-rank deterministic or random matrix, and admits a singular value decomposition (SVD)

$$Y = \sum_{i=1}^{r} \sqrt{d_i}\boldsymbol{u}_i\boldsymbol{v}_i^\top, \tag{2}$$

where $\sqrt{d_i}$, $1 \leqslant i \leqslant r$, are the singular values, and $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$, $1 \leqslant i \leqslant r$, are the left and right singular vectors, respectively. In the low-rank setting, r is a fixed integer that does not depend on n. This low-rank assumption is popular in many applications, including signal processing [3,49], imaging denoising [44,48] and statistical genetics [27,56]. Based on it, many statistical methods have been proposed to estimate Y from the noisy observation Y_t: the shrinkage estimation [28,40], the iterative thresholding procedure [10,55], and the regularization methods [30,47,57], to name a few.

Although the low-rank assumption is useful in many applications, it is not always feasible, especially in applications driven by wireless communications and massive MIMO systems [6,42,53], such as the subspace estimation [51] and direction of arrival (DOA) estimation [39]. In these applications, Y is a large rank interference matrix, where the rank r in (2) is comparable to n, and Y_t is called an interference-plus-noise matrix [53,58]; see Section 3.2 for a more detailed discussion. Moreover, in modern statistical learning theory, a large-rank matrix Y can provide deep insights into many optimization techniques. For example, it is necessary to take r to be proportional to n in order to obtain a minimax estimator on Y using nuclear norm penalization and singular value thresholding [15]. Furthermore, it is empirically observed that the mean square error of the minimax estimator of a large-rank Y is closely related to the phase transition phenomenon in matrix completion [16]. Motivated by the above applications, it is natural to extend the low-rank assumption and study the signal-plus-noise model (1) for a large-rank signal matrix Y.

From the perspective of random matrix theory, the signal-plus-noise model (1) falls into the category of the so-called deformed random matrix models [9], some of which have been studied in the literature, including the deformed Wigner matrices [31,32,35], deformed sample covariance matrices [33], and separable covariance matrices [13,54]. In this paper, we call Y_t a deformed rectangular matrix. Under the low-rank assumption that r is fixed, the empirical spectral distribution (ESD) of $Y_tY_t^\top$ is mostly determined by the noise matrix X, while the signal matrix Y will give rise to several outlier singular values (i.e., singular values that are detached from the bulk singular value spectrum) depending on the values of d_i's [5,10]. On the other hand, when Y is a large-rank matrix, the ESD of $Y_tY_t^\top$ will be governed by both the signal matrix Y and the noise matrix X. One purpose of this paper is to extend some of the known results in the low-rank setting to large-rank deformed rectangular matrices, which in turn will provide useful insights into the applications of the interference-plus-noise model (1).

Motivated by successful applications of edge statistics in high-dimensional statistics, we shall focus on the eigenvalue and eigenvector statistics of $\mathcal{Q}_t := Y_tY_t^\top$ and $\underline{\mathcal{Q}}_t := Y_t^\top Y_t$ (or equivalently, the singular value and singular vector statistics of Y_t) near the right-most edge of the spectrum. Furthermore, we are interested in the small noise regime $t \ll 1$. In fact, the case $t \sim 1$ has been studied to certain extent in [17,18,51]. Combining the results there with our arguments (in fact, most arguments will be greatly simplified in the $t \sim 1$ case), one can reproduce all the main results of this paper. We remark that there is an important difference between the $t \sim 1$ case and the $t = \mathrm{o}(1)$ case considered in this paper. In the case $t \sim 1$, due to the large noise component $\sqrt{t}X$, the asymptotic ESD of $\mathcal{Q}_t$ will have a regular square root behavior around the right-most edge regardless (to some extent) of the edge behavior of $YY^\top$. On the other hand, in the $t = \mathrm{o}(1)$ case, the edge behavior of $YY^\top$ will have a strong effect on the asymptotic ESD of $\mathcal{Q}_t$, since Y_t can be regarded as a perturbation of Y. Consequently, we need more assumptions on Y. In this paper, we shall follow the notion in [34] for deformed Wigner matrices, and impose an η_*-regular condition on Y (c.f. Definition 1) for some scale parameter $0 < \eta_* \ll 1$. This regularity assumption ensures a regular square root behavior of the ESD of $YY^\top$ on the scale η_* around the right-most edge. Moreover, we shall take t such that $\eta_* \ll t^2 \ll 1$. Intuitively, $t = \sqrt{\eta_*}$ is the threshold where the noise component $\sqrt{t}X$ starts to dominate the edge behavior of $\mathcal{Q}_t$ on the scale η_* as observed in [12].

We mention that in free probability theory [9], the asymptotic ESD of $\mathcal{Q}_t$ is called the rectangular free convolution with the Marchenko–Pastur (MP) law. In [17,18,51], it has been shown that the rectangular free convolution has a regular square root behavior near the right-most edge of the spectrum in the $t \sim 1$ case. However, the estimates there diverge as $t \to 0$, and hence are not strong enough for our setting with $t = \mathrm{o}(1)$. In this paper, we first establish some deterministic estimates regarding the rectangular free convolution based on a sophisticated analysis of the subordination function (cf. (49)). In particular, we will show that for $t \gg \sqrt{\eta_*}$, the rectangular free convolution still has a regular square root behavior near the right-most edge. Based on these estimates, we are able to prove sharp local laws on the resolvents of $\mathcal{Q}_t$ and $\underline{\mathcal{Q}}_t$. In the proof, we first establish the local laws for the so-called rectangular Gaussian divisible ensembles, which is a special case of Y_t with a Gaussian random noise component X. Together with a self-consistent comparison argument, we can extend the local laws to deformed rectangular matrices with generally distributed X, assuming only certain moment conditions. Once we have obtained the local laws, we can prove some important results regarding the eigenvalue and eigenvector statistics of $\mathcal{Q}_t$ and $\underline{\mathcal{Q}}_t$, including the edge universality, eigenvalue rigidity and eigenvector delocalization. We expect that these results also hold in the $t \sim 1$ case, and the relevant proofs should be much easier than those in the $t = o(1)$ case by using the estimates in [17,18,51] as inputs. Due to length constraint, we will not present the detailed proofs in this paper. We also notice a very recent work [59] that studies the edge statistics of Y_1 with $t = 1$ when the signal matrix Y is a large-rank diagonal matrix.

Finally, we remark that the results of this paper can be key inputs for many other problems regarding the spectral statistics of large-rank deformed rectangular matrices. For instance, in [12] we prove the Tracy–Widom fluctuation of edge eigenvalues for a general class of Gram type random matrices, using the estimates of the rectangular free convolution and the local laws proved in this paper. Furthermore, our results can be used to study the outlier eigenvalues and eigenvectors when Y_t is perturbed by another low-rank matrix, say Y_0. This kind of matrix model $Y_0 + Y + \sqrt{t}X$ will be useful in the estimation and inference of the massive MIMO system; see Section 3.2 for a more detailed discussion.

The rest of this paper is organized as follows. In Section 2, we introduce our model, the main assumptions and the rectangular free convolution. In Section 3, we state the main results and discuss some statistical applications. In Section 4, we state the key local laws on the resolvent of $\mathcal{Q}_t$, and use them to prove the main results. In Section 5, we analyze the rectangular free convolution and prove some useful deterministic estimates. We include additional technical proofs to an online supplementary file [14].

We now fix some notations that will be used frequently in the paper. The fundamental large parameter is n and we always assume that p is comparable to n. We use C to denote a generic large positive constant, whose value may change from one line to the next. Similarly, we use c, ε, τ, δ, etc. to denote generic small positive constants. If a constant depends on a quantity a, we use $C(a)$ or C_a to indicate this dependence. For two quantities a_n and b_n depending on n, the notation $a_n = O(b_n)$ means that $|a_n| \leqslant C|b_n|$ for a constant $C > 0$, and $a_n = o(b_n)$ means that $|a_n| \leqslant c_n|b_n|$ for a positive sequence $c_n \downarrow 0$ as $n \to \infty$. We also use the notations $a_n \lesssim b_n$ if $a_n = O(b_n)$, and $a_n \sim b_n$ if $a_n = O(b_n)$ and $b_n = O(a_n)$. For a matrix A, we use $\|A\| := \|A\|_{l^2 \to l^2}$ to denote the operator norm; for a vector $\mathbf{v} = (v_i)_{i=1}^n$, $\|\mathbf{v}\| \equiv \|\mathbf{v}\|_2$ stands for the Euclidean norm. For a matrix A and a positive number a, we write $A = O(a)$ if $\|A\| = O(a)$. In this paper, we often write an identity matrix of any dimension as I or 1 without causing any confusion.

2. The model and rectangular free convolution

We consider a class of deformed rectangular matrices of the form (1), where Y is a $p \times n$ deterministic signal matrix of large rank, X is a $p \times n$ random noise matrix whose entries x_{ij} are real independent random variables satisfying

$$\mathbb{E}x_{ij} = 0, \qquad \mathbb{E}|x_{ij}|^2 = n^{-1}, \quad 1 \leqslant i \leqslant p, \ 1 \leqslant j \leqslant n, \tag{3}$$

and $t > 0$ gives the noise level. Unlike [17,18,51], we have used t instead of σ^2 to denote the variance of the noise, because in this paper we will consider a full range of scales for the noise level. Hence, it is more instructive to take the noise variance to be a varying parameter. Following the random matrix literature (see e.g. [24]), we choose it to be the time parameter t. In particular, one can also consider the dynamics of Y_t as t changes. In this paper, we consider the high dimensional setting, where the aspect ratio $c_n := p/n$ converges to a finite positive constant. Without loss of generality, by switching the roles of Y_t and $Y_t^\top$ if necessary, we can assume that

$$\tau \leqslant c_n \leqslant 1, \tag{4}$$

for some small constant $0 < \tau < 1$. Let

$$Y = O_1 W O_2^\top, \tag{5}$$

be a singular value decomposition of Y, where W is a $p \times n$ rectangular diagonal matrix,

$$W = \big(D, 0\big), \quad \text{where} \quad D^2 = \operatorname{diag}(d_1, \ldots, d_p),$$

with $\sqrt{d_1} \geqslant \sqrt{d_2} \geqslant \cdots \geqslant \sqrt{d_p} \geqslant 0$ being the singular values of Y. We assume that the ESD of $YY^\top$ has a regular square root behavior near the right edge. Following [34], we state the regularity condition in terms of the Stieltjes transform of $V := WW^\top$,

$$m_V(z) := \frac{1}{p}\operatorname{Tr}(V - z)^{-1} = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{d_i - z}, \quad z \in \mathbb{C}_+ := \{z \in \mathbb{C} : \operatorname{Im} z > 0\}. \tag{6}$$

Definition 1 (η_**-regular*)**.** Let η_* be a parameter satisfying $\eta_* := n^{-\phi_*}$ for some constant $0 < \phi_* \leqslant 2/3$. We say V (or equivalently, Y, W or m_V) is η_*-regular around the largest eigenvalue d_1 if there exist constants $c_V, C_V > 0$ such that the following properties hold for $\lambda_+ := d_1$ (λ_+ is a standard notation for the right spectral edge in random matrix literature).

(i) For $z = E + \mathrm{i}\eta$ with $\lambda_+ - c_V \leqslant E \leqslant \lambda_+$ and $\eta_* + \sqrt{\eta_*|\lambda_+ - E|} \leqslant \eta \leqslant 10$, we have

$$\frac{1}{C_V}\sqrt{|\lambda_+ - E| + \eta} \leqslant \operatorname{Im} m_V(E + \mathrm{i}\eta) \leqslant C_V\sqrt{|\lambda_+ - E| + \eta}. \tag{7}$$

(ii) For $z = E + \mathrm{i}\eta$ with $\lambda_+ \leqslant E \leqslant \lambda_+ + c_V$ and $\eta_* \leqslant \eta \leqslant 10$, we have

$$\frac{1}{C_V}\frac{\eta}{\sqrt{|\lambda_+ - E| + \eta}} \leqslant \operatorname{Im} m_V(E + \mathrm{i}\eta) \leqslant C_V \frac{\eta}{\sqrt{|\lambda_+ - E| + \eta}}. \tag{8}$$

(iii) We have $2c_V \leqslant \lambda_+ \leqslant C_V/2$.

Remark 1. The motivation for conditions (7) and (8) is as follows: if $m(z)$ is the Stieltjes transform of a density ρ with square root behavior around λ_+, i.e., $\rho(x) \sim \sqrt{(\lambda_+ - x)_+}$, then (7) and (8) hold for $\operatorname{Im} m(z)$ with $\eta_* = 0$. For a general $\eta_* > 0$, (7) and (8) essentially mean that the empirical spectral density of V behaves like a square root function near λ_+ on any scale larger than η_*. The condition $\eta \leqslant 10$ in the assumption is purely for definiteness of presentation—one can replace 10 with any constant of order 1.

Consider a large rank matrix Y whose spectral density of singular values follows a continuous function, say ρ, on the scale η_*. Then the square root behavior of ρ appears naturally near the spectral edge, which is the point where the density becomes zero. The conditions (7) and (8) hold for many Gram type random matrix ensembles for certain $n^{-2/3} \ll \eta_* \ll 1$, such as sample covariance matrices [11,33,46], separable sample covariance matrices [13,54], random Gram matrices [1,2], sparse sample covariance matrices [29], and signal-plus-noise matrices with $t \sim 1$ [17,18,51,59].

Recall that $\mathcal{Q}_t = Y_t Y_t^\top$. Let $\rho_{w,t}$ be the asymptotic spectral density of $\mathcal{Q}_t$ as $n \to \infty$ and $m_{w,t}$ be the corresponding Stieltjes transform, i.e.,

$$m_{w,t}(z) := \int \frac{\rho_{w,t}(x)\mathrm{d}x}{x - z}. \tag{9}$$

Here, "w" in the subscript refers to the matrix W, and we have used it to remind ourselves that $\rho_{w,t}$ and $m_{w,t}$ only depend on the singular values of Y. Especially, when $t = 0$, we have the initial condition $m_{w,0}(z) = m_V(z)$. It is known that for any $t > 0$, $m_{w,t}$ is the unique solution to

$$m_{w,t} = \frac{1}{p}\sum_{i=1}^{p} \frac{1}{d_i(1 + c_n t m_{w,t})^{-1} - (1 + c_n t m_{w,t})z + t(1 - c_n)}, \tag{10}$$

such that $\operatorname{Im} m_{w,t} > 0$ for any $z \in \mathbb{C}_+$ [17,18,51]. Adopting notations from free probability theory [9], we shall call $\rho_{w,t}$ the rectangular free convolution of $\rho_{w,0}$ with the Marchenko–Pastur (MP) law at time t. Let $\lambda_{+,t}$ be the right-most edge of $\rho_{w,t}$. In [17,18,51], it has been shown that $\rho_{w,t}$ has a regular square root behavior near $\lambda_{+,t}$ when $t \sim 1$. In the following lemma, we extend this result to the $t = \mathrm{o}(1)$ case.

Lemma 1. *Suppose (4) holds, and V is η_*-regular in the sense of Definition 1. Moreover, we assume that t satisfies $n^{\varepsilon}\eta_* \leqslant t^2 \leqslant n^{-\varepsilon}$ for a small constant $\varepsilon > 0$. Then*

$$\rho_{w,t}(E) \sim \sqrt{(\lambda_{+,t} - E)_+} \quad \text{for} \quad \lambda_{+,t} - 3c_V/4 \leqslant E \leqslant \lambda_{+,t} + 3c_V/4, \tag{11}$$

and for $z = E + \mathrm{i}\eta \in \mathbb{C}_+$,

$$\operatorname{Im} m_{w,t}(z) \sim \begin{cases} \sqrt{|E - \lambda_{+,t}| + \eta}, & \text{if } \lambda_{+,t} - 3c_V/4 \leqslant E \leqslant \lambda_{+,t} \\ \dfrac{\eta}{\sqrt{|E - \lambda_{+,t}| + \eta}}, & \text{if } \lambda_{+,t} \leqslant E \leqslant \lambda_{+,t} + 3c_V/4 \end{cases}. \tag{12}$$

Proof. This lemma is an immediate consequence of Lemmas 17 and 18. □

Remark 2. We have required a lower bound $t^2 \gg \eta_*$ in the assumption due to the following reason. For very small t, the edge behavior of $\rho_{w,t}$ is only a perturbation of the edge behavior of $YY^\top$ near λ_+, and $t = \sqrt{\eta_*}$ is the threshold when the random matrix statistics of $\sqrt{t}X$ begin to dominate over the effect of the edge eigenvalues of $YY^\top$. Theoretically, if the entries of X are i.i.d. Gaussian, it has been shown in [12] that the edge statistics of $Y_t Y_t^\top$ already converge to local equilibrium when $t^2 \gg \eta_*$.

3. Main results and statistical applications

In this section, we state the main results and explain how they can be used to study the massive MIMO system. For simplicity, we introduce the following notion of stochastic domination, which was first introduced in [21] and subsequently used in many works on random matrix theory. It simplifies the presentation of the results and their proofs by systematizing statements of the form "ξ is bounded by ζ with high probability up to a small power of n".

Definition 2 (*Stochastic Domination*)**.** We define the following notion of stochastic domination.

(i) Let

$$\xi = \left(\xi^{(n)}(u) : n \in \mathbb{N}, u \in U^{(n)}\right), \quad \zeta = \left(\zeta^{(n)}(u) : n \in \mathbb{N}, u \in U^{(n)}\right),$$

be two families of nonnegative random variables, where $U^{(n)}$ is a possibly n-dependent parameter set. We say ξ is stochastically dominated by ζ, uniformly in u, if for any fixed (small) $\varepsilon > 0$ and (large) $D > 0$,

$$\sup_{u \in U^{(n)}} \mathbb{P}\left(\xi^{(n)}(u) > n^{\varepsilon}\zeta^{(n)}(u)\right) \leqslant n^{-D}$$

for large enough $n \geqslant n_0(\varepsilon, D)$, and we shall use the notation $\xi \prec \zeta$. If a complex family of random variables ξ satisfies $|\xi| \prec \zeta$, then we will also write $\xi \prec \zeta$ or $\xi = \mathrm{O}_{\prec}(\zeta)$.

(ii) We extend the definition of $\mathrm{O}_{\prec}(\cdot)$ to matrices in the operator norm sense as follows. Let A be a family of random matrices and ζ be a family of nonnegative random variables. Then $A = \mathrm{O}_{\prec}(\zeta)$ means that $\|A\| \prec \zeta$.

(iii) We say an event Ξ holds with high probability if for any constant $D > 0$, $\mathbb{P}(\Xi) \geqslant 1 - n^{-D}$ for large enough n.

3.1. Main results

Consider the singular value decomposition of Y_t,

$$Y_t = \sum_{k=1}^{p} \sqrt{\lambda_k}\boldsymbol{\xi}_k\boldsymbol{\zeta}_k^{\top},$$

where $\sqrt{\lambda_1} \geqslant \cdots \geqslant \sqrt{\lambda_p}$ are the singular values of Y_t, $\{\boldsymbol{\xi}_k\}_{k=1}^{p}$ are the left singular vectors, and $\{\boldsymbol{\zeta}_k\}_{k=1}^{n}$ are the right-singular vectors. For any fixed E, let $\eta_l(E)$ ("l" stands for "lower bound") be the solution of

$$n\eta_l(E)\left(t + \sqrt{|E - \lambda_{+,t}| + \eta_l(E)}\right) = 1. \tag{13}$$

Here, $\eta_l(E)$ can be understood as the lower bound of the spectral scale above which the local laws in Theorem 3 hold (see Eq. (37)). For t satisfying $t \gg n^{-1/3}$, it is easy to check that

$$\eta_l(E) \sim \frac{1}{n(t + \sqrt{|E - \lambda_{+,t}|})}. \tag{14}$$

We define the classical location γ_j for the jth eigenvalue of $\mathcal{Q}_t$ as

$$\gamma_j := \sup_x \left\{ \int_x^{+\infty} \rho_{w,t}(x)dx > \frac{j-1}{p} \right\}. \tag{15}$$

In particular, by the square root behavior of $\rho_{w,t}$ in (11), we have $\gamma_1 = \lambda_{+,t}$ and $|\gamma_j - \gamma_1| \sim j^{2/3}n^{-2/3}$ for $j \geqslant 2$ such that $\gamma_j \geqslant \lambda_{+,t} - 3c_V/4$. With the above notations, we now state our main theorem.

Theorem 1. *Suppose (4) holds, V is η_*-regular in the sense of Definition 1, and X is a $p \times n$ random matrix whose entries x_{ij} are real independent random variables satisfying (3) and*

$$\mathbb{E}x_{ij}^3 = 0, \quad 1 \leqslant i \leqslant p, \ 1 \leqslant j \leqslant n. \tag{16}$$

Moreover, assume that the entries of X have finite moments up to any order, that is, for any fixed $k \in \mathbb{N}$, there exists a constant $C_k > 0$ such that

$$\mathbb{E}|\sqrt{n}x_{ij}|^k \leqslant C_k. \tag{17}$$

Suppose t satisfies $n^{\varepsilon}\eta_* \leqslant t^2 \leqslant n^{-\varepsilon}$ for a small constant $\varepsilon > 0$. Then, we have:

(i) *(Eigenvalue rigidity) For any k such that $\lambda_{+,t} - c_V/2 < \gamma_k \leqslant \lambda_{+,t}$, we have*

$$|\lambda_k - \gamma_k| \prec n^{-2/3}k^{-1/3} + \eta_l(\gamma_k). \tag{18}$$

(ii) *(Edge universality) There exist constants $\varepsilon, \delta > 0$ such that for all $x \in \mathbb{R}$,*

$$\begin{aligned} \mathbb{P}^G\left(n^{2/3}(\lambda_1 - \lambda_{+,t}) \leqslant x - n^{-\varepsilon}\right) - n^{-\delta} &\leqslant \mathbb{P}\left(n^{2/3}(\lambda_1 - \lambda_{+,t}) \leqslant x\right) \\ &\leqslant \mathbb{P}^G\left(n^{2/3}(\lambda_1 - \lambda_{+,t}) \leqslant x + n^{-\varepsilon}\right) + n^{-\delta}, \end{aligned} \tag{19}$$

where $\mathbb{P}^G$ denotes the law for X with i.i.d. Gaussian entries satisfying (3).

(iii) *(Eigenvector delocalization) For any deterministic unit vectors $\mathbf{u} \in \mathbb{R}^p$, $\mathbf{v} \in \mathbb{R}^n$ and $k \in \mathbb{N}$ such that $\lambda_{+,t} - c_V/2 < \gamma_k \leqslant \lambda_{+,t}$, we have that*

$$\left|\mathbf{u}^{\top}\boldsymbol{\xi}_k\right|^2 + \left|\mathbf{v}^{\top}\boldsymbol{\zeta}_k\right|^2 \prec \frac{1}{nt(t + \sqrt{|\gamma_k - \lambda_{+,t}|})^2}. \tag{20}$$

Since $t \geqslant n^{\varepsilon/2}\sqrt{\eta_*} \geqslant n^{-1/3+\varepsilon/2}$, (20) gives that $|\mathbf{u}^\top \boldsymbol{\xi}_k|^2 + |\mathbf{v}^\top \boldsymbol{\zeta}_k|^2 \prec n^{-3\varepsilon/2}$. When k increases, the delocalization estimate (20) gives better bounds.

Remark 3. We make some remarks about the technical vanishing third moment condition (16). The proof of Theorem 1 relies on a key ingredient—the local laws in Theorem 3, whose proof uses a continuous self-consistent comparison argument developed in [33,54]. Roughly speaking, we first show that Theorems 1 and 3 hold in the Gaussian case, i.e., when the entries x_{ij} are i.i.d. Gaussian random variables, and then show that the non-Gaussian case is sufficiently close to the Gaussian case using the self-consistent comparison argument in [33,54]. For this comparison argument to work, we need to match the third moment of x_{ij} with that of a standard Gaussian random variable, which leads to the condition (16); we refer the reader to the argument between Eqs. (D.19) and (D.20) in the supplement [14] for more details. We believe that the condition (16) is mainly technical and can be removed with further theoretical development. But, since this is not the focus of the current paper, we leave it to future study.

In [12], we have proved the Tracy–Widom law of $n^{2/3}(\lambda_1 - \lambda_{+,t})$ for the Gaussian case. Together with the edge universality result (19), it immediately gives the following corollary.

Corollary 2. *Suppose the assumptions of Theorem 1 hold. There exists a positive parameter $\gamma_0 \equiv \gamma_0(Y, t)$ of order 1 such that*

$$\lim_{n\to\infty} \mathbb{P}(\gamma_0 n^{2/3}(\lambda_1 - \lambda_{+,t}) \leqslant s) = F_1(s) \quad \text{for all} \quad s \in \mathbb{R},$$

where $F_1(s)$ is the type-1 Tracy–Widom distribution.

Remark 4. As in [22,25,36], (19) can be generalized to the finite correlation function of the k largest eigenvalues for any fixed $k \in \mathbb{N}$:

$$\begin{aligned}\mathbb{P}^G\left(\left(n^{2/3}(\lambda_i - \lambda_{+,t}) \leqslant x_i - n^{-\varepsilon}\right)_{1\leqslant i\leqslant k}\right) - n^{-\delta} &\leqslant \mathbb{P}\left(\left(n^{2/3}(\lambda_i - \lambda_{+,t}) \leqslant x_i\right)_{1\leqslant i\leqslant k}\right)\\ &\leqslant \mathbb{P}^G\left(\left(n^{2/3}(\lambda_i - \lambda_{+,t}) \leqslant x_i + n^{-\varepsilon}\right)_{1\leqslant i\leqslant k}\right) + n^{-\delta},\end{aligned} \tag{21}$$

for all $x_i \in \mathbb{R}$, $1 \leqslant i \leqslant k$. Combining it with our results in [12], we obtain that $n^{2/3}(\lambda_i - \lambda_{+,t})$, $1 \leqslant i \leqslant k$, satisfy the same asymptotic joint distribution as Wishart matrices or the Gaussian orthogonal ensemble (GOE).

Remark 5. We make a few remarks about the main results. Although in this paper we focus on the $t = o(1)$ setting, all the above results also apply to the $t \sim 1$ setting, and our proofs can be readily extended to that setting after minor modifications (in fact, all proofs will become much easier because we do not need to keep track of the t factors). First, notice that $\eta_l(\gamma_k) \sim n^{-1}$ by (14) when $t \sim 1$, and hence the eigenvalue rigidity estimate (18) becomes $|\lambda_k - \gamma_k| \prec n^{-2/3}k^{-1/3}$. Second, the delocalization estimate (20) becomes $|\mathbf{u}^\top \boldsymbol{\xi}_k|^2 + |\mathbf{v}^\top \boldsymbol{\zeta}_k|^2 \prec n^{-1}$, which is a sharp estimate. Finally, the edge universality result in Theorem 1 still holds, which, together with a recent result in [59], implies the Tracy–Widom asymptotics for edge eigenvalues as given by Corollary 2.

Our results also apply to the complex setting where Y can be a complex matrix and X contains independent complex entries satisfying (3) and $\mathbb{E}x_{ij}^2 = 0$, except that the type-1 Tracy–Widom law in Corollary 2 should be replaced by the type-2 Tracy–Widom law. All our current proofs still work with some minor changes in notations. However, for definiteness and due to length constraint, we focus on the real setting in this paper.

3.2. Statistical applications

In this subsection, we discuss some potential applications of our results in high dimensional statistics. Specifically, we shall consider the model used in multicell multiuser MIMO system [53] as an example, which is a massive MIMO system [6,42,52]. The massive MIMO system is a promising technique to deal with large wireless communication systems, such as the design of 5G [42]. In contrast to the standard MIMO system that assumes a low-rank structure of Y [50], the massive MIMO system usually requires a large-rank signal matrix.

We first introduce the model for the multicell multiuser MIMO system [53]. For ease of discussion, in what follows, we focus on the real setting, but similar arguments also hold for the complex setting. Suppose there exist a single target cell and r nearby interfering cells. Each cell contains a single base station equipped with p antennas and K single-antenna users. Consider the uplink (reverse link) transmission where the target base station receives signals from all users in all cells. Then, we observe n i.i.d. samples $\boldsymbol{y}_i$, $i \in \{1, \ldots, n\}$, each of which can be modeled as

$$\boldsymbol{y}_i = \mathbf{H}\boldsymbol{z}_i + \sum_{k=1}^{r} \mathbf{W}_k \mathbf{z}_i^k + \boldsymbol{w}_i. \tag{22}$$

Here, the transmitted data $\boldsymbol{z}_i \in \mathbb{R}^K$ is a centered random vector with covariance matrix Γ; $\mathbf{H} \in \mathbb{R}^{p\times K}$ is the channel matrix between the base station and the K users; $\mathbf{z}_i^k \in \mathbb{R}^K$ is the interfering data in the kth interfering cell with i.i.d. centered entries of unit variance; $\mathbf{W}_k \in \mathbb{R}^{p\times K}$ is the channel matrix between the base station and the users in cell k; $\boldsymbol{w}_i \in \mathbb{R}^p$ is the additive noise with i.i.d. centered entries of variance $\sqrt{t}$. We assume that all random vectors $\boldsymbol{z}_i$, $\mathbf{z}_i^k$ and $\boldsymbol{w}_i$ are independent of each other. Suppose that the number of users in each cell is fixed, i.e., K is a fixed integer, and that the number of the neighboring interfering cells is large, i.e., r is large. Denote

$$\widetilde{\mathbf{z}}_i = ((\mathbf{z}_i^1)^\top, \ldots, (\mathbf{z}_i^r)^\top)^\top, \quad W := (\mathbf{W}_1, \ldots, \mathbf{W}_r).$$

Then, we obtain the following matrix model by concatenating the n observed samples:

$$\widetilde{Y}_t := (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) = \mathbf{H}\Gamma^{1/2}Z + WZ_I + \sqrt{t}X, \tag{23}$$

where $Z \in \mathbb{R}^{K\times n}$, $Z_I := (\widetilde{\mathbf{z}}_1, \ldots, \widetilde{\mathbf{z}}_n) \in \mathbb{R}^{(rK)\times n}$ and $X := t^{-1/2}(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n) \in \mathbb{R}^{p\times n}$ are independent random matrices with i.i.d. centered entries of unit variance. Denoting $Y_0 := \mathbf{H}\Gamma^{1/2}Z$ and $Y := WZ_I$, we can rewrite (23) as

$$\widetilde{Y}_t = Y_0 + Y + \sqrt{t}X, \tag{24}$$

where Y_0 is a low-rank matrix representing the transmitted signals of the home cell, Y is a large-rank matrix representing the signals of the interfering cells, and $\sqrt{t}X$ is the additive noise. Note that $Y_t := Y + \sqrt{t}X$ is the deformed rectangular matrix model, which, as mentioned in [53, Section 3], is called the interference-plus-noise matrix in the above application. We are interested in the estimation and inference of the model (24). For definiteness, in the following discussion, we assume that Y_t satisfies the assumptions of Theorem 1 for $t = \mathrm{o}(1)$ or $t \sim 1$ as discussed in Remark 5. We point out that when $t \sim 1$, the signals and multi-cell interference are both comparable with the noise and when $t = \mathrm{o}(1)$, the signal-to-noise ratio will diverge. Both settings are important in statistical applications.

In the first application, we are interested in detecting the signals, that is, the existence of Y_0. Let r_* the number of non-zero singular values of Y_0. Formally, we consider the following hypothesis testing problem

$$\mathbf{H}_0: \ \mathrm{r}_* = 0 \quad \text{vs} \quad \mathbf{H}_a: \ \mathrm{r}_* > 0. \tag{25}$$

Under the null hypothesis of $\mathbf{H}_0$, by (21) we know that the joint distribution of the largest few eigenvalues of $\widetilde{Y}_t\widetilde{Y}_t^\top$ is universal regardless of the distributions of the entries of X. Since we do not have a priori information on the interference matrix Y and the noise level t, we shall use the following pivotal statistic [45]

$$\mathbb{T}_1 = \frac{\mu_1 - \mu_2}{\mu_2 - \mu_3}, \tag{26}$$

where $\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_p$ are the eigenvalues of $\widetilde{Y}_t\widetilde{Y}_t^\top$. The statistic $\mathbb{T}_1$ will be powerful if some singular values of Y_0 are above the threshold of BBP transition such that they give rise to some outliers, that is, eigenvalues that are detached from the bulk eigenvalue spectrum. This kind of assumption appears commonly in the literature of signal detection, see e.g. [3,40,41,43]. Furthermore, by Remark 4, we know that under the null hypothesis $\mathbf{H}_0$, $\mathbb{T}_1$ actually satisfies an explicit distribution that can be derived from the Tracy–Widom law.

For the second application, we consider the estimation of the number of signals once we reject the null hypothesis of (25). For simplicity, for now we assume that the eigenvalues $d_1 > d_2 > \cdots > d_{\mathrm{r}_*}$ of $Y_0Y_0^\top$ are reasonably large such that they gives rise to r_* outliers $\mu_1 > \mu_2 > \cdots > \mu_{\mathrm{r}_*}$. Following the discussions in Sections 3 and 4 of [9], one can show that with probability $1 - \mathrm{o}(1)$,

$$\mu_i = \zeta_t^{-1}(d_i) + \mathrm{o}(1), \quad 1 \leqslant i \leqslant \mathrm{r}_*,$$

where $\zeta_t^{-1}(\cdot)$ is the inverse function of the subordination function defined in (49) below. With the estimates proved in Section 5.2, we can show that $\mu_i > \lambda_{+,t}$ if $d_i > \zeta_t(\lambda_{+,t})$, where $\zeta_t(\lambda_{+,t})$ gives the threshold of the BBP transition. On the other hand, by (18) and Cauchy interlacing theorem, we have that

$$\mu_{j+\mathrm{r}_*} = \lambda_{+,t} + \mathrm{O}_\prec(n^{-2/3+\varepsilon}), \quad \text{for any fixed } \ j \geqslant 1.$$

In light of the above observations, we propose the following statistic,

$$\widehat{\mathrm{r}}_* := \underset{1\leqslant i\leqslant \ell}{\operatorname{argmin}} \left\{ \frac{\mu_{i+1}}{\mu_{i+2}} - 1 \leqslant \omega \right\}.$$

Here, ℓ is a pre-given large constant and ω is a small number that can be chosen using a calibration procedure. We refer the readers to [13, Section 4.1] for more details. Using Theorem 1, it is not hard to show that $\widehat{\mathrm{r}}_*$ is a consistent estimator of r_*. We also remark that the local law, Theorem 3, combined with the strategy in [4,7,10,13] can give optimal convergent rates and exact asymptotic distributions for the outlier eigenvalues μ_i, $1 \leqslant i \leqslant \mathrm{r}_*$. However, this requires a lot more dedicated efforts and is beyond the scope of the current paper. We will pursue this direction somewhere else. We also remark that in general, it may happen that only a subset of the eigenvalues of $Y_0Y_0^\top$ are above the BBP transition threshold, say $\mu_1 > \cdots > d_{r_+} > \zeta_t(\lambda_{+,t}) > d_{r_++1} > \cdots > d_{\mathrm{r}_*}$ for a fixed $0 < r_+ < \mathrm{r}_*$. In this case, the estimator $\widehat{\mathrm{r}}_*$ will consistently give the value r_+, and it is well-known that eigenvalues $d_{r_++1}, \ldots, d_{\mathrm{r}_*}$ cannot be detected reliably using the singular values of $\widetilde{Y}_t$ only.

It is worth pointing out that several other results of this paper can also be applied to some statistical problems involving large-rank deformed rectangular matrices. For instance, in Section 5, we will conduct a thorough analysis of

$m_{w,t}$ and Eq. (10). Based on these results, we can propose a convex optimization based methodology to estimate the large-rank matrix Y by utilizing (10) and the strategy in [20].

As a last example of applications, the model (24) also appears naturally in financial economics. For example, in the factor model [26,37,45], Y_0 represents the excess return matrix, Y is the cross-section part (i.e., the common factors), and $\sqrt{t}X$ is the idiosyncratic component. In existing literature, Y is commonly assumed to be sparse or low-rank. Based on the results of this paper, we can study factor models beyond these assumptions. It will be an interesting topic for future works.

4. Local laws and proofs of the main results

Our local laws can be formulated in a simple and unified fashion using the following $(p+n)\times(p+n)$ symmetric block matrix

$$H_t \equiv H_t(Y_t) := \begin{pmatrix} 0 & Y_t \\ Y_t^\top & 0 \end{pmatrix},$$

which we shall refer to as the linearization of the matrices $\mathcal{Q}_t$ and $\underline{\mathcal{Q}}_t$.

Definition 3 (*Resolvents*)**.** We define the resolvent of H_t as

$$G(z) \equiv G(Y_t, z) := (z^{1/2}H_t - z)^{-1}, \quad z \in \mathbb{C}_+. \tag{27}$$

For $\mathcal{Q}_t = Y_t Y_t^\top$ and $\underline{\mathcal{Q}}_t = Y_t^\top Y_t$, we define the resolvents

$$\mathcal{G} \equiv \mathcal{G}(Y_t, z) := (\mathcal{Q}_t - z)^{-1}, \quad \underline{\mathcal{G}} \equiv \underline{\mathcal{G}}(Y_t, z) := \left(\underline{\mathcal{Q}}_t - z\right)^{-1}. \tag{28}$$

We denote the empirical spectral density ρ of $\mathcal{Q}_t$ and its Stieltjes transform as

$$\rho \equiv \rho(Y_t, z) := \frac{1}{p}\sum_{i=1}^{p} \delta_{\lambda_i(\mathcal{Q}_t)}, \quad m(z) \equiv m(Y_t, z) := \int \frac{1}{x-z}\rho(\mathrm{d}x) = \frac{1}{p}\mathrm{Tr}\,\mathcal{G}(z), \tag{29}$$

where $\lambda_i(\mathcal{Q}_t)$, $1 \leqslant i \leqslant p$, denote the eigenvalues of $\mathcal{Q}_t$ in descending order. Similarly, we denote the empirical spectral density $\underline{\rho}$ of $\underline{\mathcal{Q}}_t$ and its Stieltjes transform as

$$\underline{\rho} \equiv \underline{\rho}(Y_t, z) := n^{-1}\sum_{i=1}^{n} \delta_{\lambda_i(\underline{\mathcal{Q}}_t)}, \quad \underline{m}(z) \equiv \underline{m}(Y_t, z) := \int \frac{1}{x-z}\underline{\rho}(\mathrm{d}x) = n^{-1}\mathrm{Tr}\,\underline{\mathcal{G}}(z). \tag{30}$$

Using the Schur complement formula, one can check that

$$G = \begin{pmatrix} \mathcal{G} & z^{-1/2}\mathcal{G}Y_t \\ z^{-1/2}Y_t^\top \mathcal{G} & \underline{\mathcal{G}} \end{pmatrix} = \begin{pmatrix} \mathcal{G} & z^{-1/2}Y_t\underline{\mathcal{G}} \\ z^{-1/2}\underline{\mathcal{G}}Y_t^\top & \underline{\mathcal{G}} \end{pmatrix}. \tag{31}$$

Thus, a control of $G(z)$ yields directly a control of the resolvents $\mathcal{G}$ and $\underline{\mathcal{G}}$. Since $\underline{\mathcal{Q}}_t$ share the same nonzero eigenvalues with $\mathcal{Q}_t$ and has $n-p$ more zero eigenvalues due to (4), we have

$$\underline{m} = \frac{p}{n}m - \frac{n-p}{nz} = c_n m - \frac{1-c_n}{z}. \tag{32}$$

We will see that $(m, \underline{m})$ is asymptotically equal to $(m_{w,t}, \underline{m}_{w,t})$ as $n \to \infty$, where

$$\underline{m}_{w,t} := c_n m_{w,t} - \frac{1-c_n}{z}. \tag{33}$$

Furthermore, we define the asymptotic matrix limit of $G(z)$ as

$$\Pi(z) := \begin{bmatrix} \dfrac{-(1+c_n t m_{w,t})}{z(1+c_n t m_{w,t})(1+t\underline{m}_{w,t}) - YY^\top} & \dfrac{-z^{-1/2}}{z(1+c_n t m_{w,t})(1+t\underline{m}_{w,t}) - YY^\top}Y \\ Y^\top \dfrac{-z^{-1/2}}{z(1+c_n t m_{w,t})(1+t\underline{m}_{w,t}) - YY^\top} & \dfrac{-(1+t\underline{m}_{w,t})}{z(1+c_n t m_{w,t})(1+t\underline{m}_{w,t}) - Y^\top Y} \end{bmatrix}. \tag{34}$$

It is easy to check that the inverse of Π is

$$\Pi^{-1}(z) := \begin{pmatrix} -z(1+t\underline{m}_{w,t})I_p & z^{1/2}Y \\ z^{1/2}Y^\top & -z(1+c_n t m_{w,t})I_n \end{pmatrix}. \tag{35}$$

For $z = E + \mathrm{i}\eta \in \mathbb{C}_+$, we introduce the notation

$$\kappa \equiv \kappa_E := |E - \lambda_{+,t}|. \tag{36}$$

Then, for any constant $\vartheta > 0$, we define the spectral domains

$$\mathcal{D}_\vartheta := \left\{ z = E + \mathrm{i}\eta : \lambda_{+,t} - \frac{3}{4}c_V \leqslant E \leqslant \lambda_{+,t} + \vartheta^{-1}t^2, t + \sqrt{\kappa+\eta} \geqslant \frac{n^\vartheta}{n\eta}, \eta \leqslant 10 \right\} \cup \mathcal{D}_\vartheta^{out}, \tag{37}$$

where

$$\mathcal{D}_\vartheta^{out} := \left\{ z = E + \mathrm{i}\eta : \lambda_{+,t} \leqslant E \leqslant \lambda_{+,t} + \frac{3}{4}c_V, n\eta\sqrt{\kappa+\eta} \geqslant n^\vartheta, \eta \leqslant 10 \right\}. \tag{38}$$

For simplicity, we introduce the notation

$$\zeta_t(z) := z(1 + c_n t m_{w,t})(1 + t\underline{m}_{w,t}). \tag{39}$$

From Lemmas 14 and 15, we can obtain a lower bound on the smallest singular value of $\zeta_t(z) - YY^\top$, which, together with Lemma 1, implies that for $z \in \mathcal{D}_\vartheta$,

$$\|\Pi(z)\| \lesssim \varpi^{-1}(z), \quad \text{where} \quad \varpi(z) := \begin{cases} t^2 + \eta + t\sqrt{\kappa+\eta}, & E \leqslant \lambda_{+,t} \\ t^2 + \kappa + \eta, & E \geqslant \lambda_{+,t}. \end{cases} \tag{40}$$

Now, we are ready to state our local laws on the resolvent $G(z)$. For simplicity of notations, we introduce the following deterministic control parameters

$$\Psi(z) := \sqrt{\frac{\operatorname{Im} m_{w,t}(z)}{n\eta}} + \frac{1}{n\eta}, \quad \Phi(z) := \frac{1}{\varpi(z)}\left(t\Psi(z) + \frac{t^{1/2}}{n^{1/2}} \right). \tag{41}$$

Using the definition of $\mathcal{D}_\vartheta$ and Lemma 1, it is easy to check that

$$\Phi(z) \lesssim n^{-\vartheta/2}, \quad z \in \mathcal{D}_\vartheta. \tag{42}$$

For any vector $\mathbf{u} \in \mathbb{R}^{p+n}$, we introduce the notation

$$\|\mathbf{u}\|_\Pi(z) := \|\Pi(z)\mathbf{u}_1\| + \|\Pi(z)\mathbf{u}_2\|, \tag{43}$$

where $\mathbf{u}_1 \in \mathbb{R}^p$ and $\mathbf{u}_2 \in \mathbb{R}^n$ are the subvectors such that $\mathbf{u} = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix}$. We now state the anisotropic and averaged local laws and defer its proof to the supplement [14].

Theorem 3. *Suppose the assumptions of Theorem 1 hold. Then, for any constant $\vartheta > 0$, the following estimates hold uniformly in $z \in \mathcal{D}_\vartheta$:*

(i) *(Anisotropic local law) For any deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{p+n}$,*

$$\left| \mathbf{u}^\top [G(z) - \Pi(z)] \mathbf{v} \right| \prec \Phi(z) \|\mathbf{u}\|_\Pi^{1/2} \|\mathbf{v}\|_\Pi^{1/2}; \tag{44}$$

(ii) *(Averaged local law) For $z \in \mathcal{D}_\vartheta$, we have*

$$|m(z) - m_{w,t}(z)| \prec (n\eta)^{-1}; \tag{45}$$

for $z \in \mathcal{D}_\vartheta^{out}$, we have

$$|m(z) - m_{w,t}(z)| \prec \frac{1}{n(\kappa+\eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa+\eta}} + \frac{\Phi(z)}{n\eta}. \tag{46}$$

Remark 6. As mentioned in Remark 3, the proof of Theorem 3 employs the two-step strategy in [33,54]: we first prove the local laws for the Gaussian case and then use a self-consistent comparison argument to show the closeness between the non-Gaussian case and the Gaussian case. In the first step, the Gaussian assumption is only used to diagonalize the matrix Y (while keeping the distribution of Y_t) as discussed in Eq. (B.1) of the supplement [14]. If Y is already diagonalized, then the comparison argument (and hence the vanishing third moment condition (16)) is not necessary anymore and we can prove the anisotropic and averaged local laws directly. On the other hand, for a general non-diagonal Y, the reduction to the diagonal setting is critical for proving the local laws; we refer the reader to equation (B.1) of the supplement [14] for more details.

With Theorem 3, we can prove Theorem 1 with some standard arguments. We will not write down all details, but refer the reader to some known arguments in the literature instead.

Proof of Theorem 1. The estimate (18) follows from the averaged local laws (45) and (46) combined with a standard argument using Helffer-Sjöstrand calculus. The reader can refer to e.g. [23, Theorem 2.13], [25, Theorem 2.2] and [46, Theorem 3.3] for more details.

The estimate (19) can be proved with a resolvent comparison argument as in the proof of [25, Theorem 2.4]. We have collected all necessary inputs for this argument, including the rigidity of eigenvalues (18), the averaged local laws (45) and (46), and the anisotropic local law (44). We remark that in order for these arguments to work, we need to know that the local laws hold at $z = E + \mathrm{i}n^{-2/3-\delta}$ for E around $\lambda_{+,t}$ and for some small constant $\delta > 0$. Our domain $\mathcal{D}_\vartheta$ covers such a spectral parameter regime by the assumption $t \geqslant n^{\varepsilon/2}\sqrt{\eta_*} \geqslant n^{-1/3+\varepsilon/2}$ as long as ϑ is taken sufficiently small.

To prove (20), we use the following spectral decomposition of G: for $1 \leqslant i, j \leqslant p$ and $p+1 \leqslant \mu, \nu \leqslant p+n$,

$$G_{ij} = \sum_{k=1}^{p} \frac{\boldsymbol{\xi}_k(i)\boldsymbol{\xi}_k^\top(j)}{\lambda_k - z}, \qquad G_{\mu\nu} = \sum_{k=1}^{n} \frac{\boldsymbol{\zeta}_k(\mu)\boldsymbol{\zeta}_k^\top(\nu)}{\lambda_k - z}. \tag{47}$$

Let $z_k := \lambda_k + \mathrm{i}n^\delta \eta_l(\lambda_k)$ for a small constant $\delta > 0$. By (13), (18) and (37), we see that $z_k \in \mathcal{D}_\vartheta$. Then, with (44) and (47), we obtain that

$$\frac{|\mathbf{u}^\top \boldsymbol{\xi}_k|^2}{n^\delta \eta_l(\lambda_k)} \leqslant \sum_{j=1}^{p} \frac{n^\delta \eta_l(\lambda_k)|\mathbf{u}^\top \boldsymbol{\xi}_k|^2}{(\lambda_j - \lambda_k)^2 + n^{2\delta}\eta_l^2(\lambda_k)} = \operatorname{Im} \mathbf{u}^\top G(z_k)\mathbf{u} = \operatorname{Im} \mathbf{u}^\top \Pi(z_k)\mathbf{u} + \mathrm{O}_\prec(\Phi(z_k) \cdot \|\mathbf{u}\|_\Pi(z_k)).$$

Together with (40), this estimate immediately gives that

$$|\mathbf{u}^\top \boldsymbol{\xi}_k|^2 \leqslant \frac{n^\delta \eta_l(\lambda_k)}{\varpi(z_k)} \lesssim \frac{n^\delta}{nt(t + \sqrt{|\gamma_k - \lambda_{+,t}|})^2}.$$

where in the second step we have used (14) and applied the rigidity estimate (18) to replace λ_k with γ_k. Since δ is arbitrary, we conclude the estimate (20) for $\left|\mathbf{u}^\top \boldsymbol{\xi}_k\right|^2$. The estimate for $\left|\mathbf{v}^\top \boldsymbol{\zeta}_k\right|^2$ can be proved in the same way. □

Proof of Corollary 2. In [12, Theorem 2.3], the Tracy–Widom law has been established when X is a Gaussian random matrix. Then the corollary follows immediately from (19). □

Remark 7. We now briefly explain how the local laws and their proofs can be readily generalized to the setting with $t \sim 1$ or complex X and Y entries.

First, Theorem 3 still holds when $t \sim 1$. Similar to the current one, the proof of Theorem 3 in the $t \sim 1$ case also relies on the analysis of rectangular free convolution in Section 5, which does not change at all when we relax the upper bound on t. In fact, some relevant estimates have already been proved in [17,18,51] in the $t \sim 1$ setting, and hence the arguments in Section 5 will be simplified greatly. Moreover, the proof of Theorem 3 in the supplement [14] will also be much simpler, because we do not need to track the t factors in all estimates. As we have seen, with the local laws, we can readily complete the proof of Theorem 1 using some standard arguments in the literature. Regarding the proof of Corollary 2, we mention that in a recent work [59], the authors proved the Tracy–Widom law in the setting with $t \sim 1$ and Y being a diagonal matrix. Together with the edge universality result in Theorem 1, we can remove the diagonal assumption on Y and prove Corollary 2 in the $t \sim 1$ case.

Second, from the real setting to the complex one, all results and proofs in Section 5 are unchanged because the defining Eq. (10) of $m_{w,t}$ is still the same. On the other hand, in the complex setting, the proof of Theorem 3 in the supplement [14] is only subject to some minor notational changes if we assume in addition that $\mathbb{E}x_{ij}^2 = 0$. Moreover, generalizing the proof in [12] to the complex setting shows that the edge eigenvalue statistics match those of a complex sample covariance matrix asymptotically, which satisfy the type-2 Tracy–Widom law as shown in [19]. Hence, we should replace F_1 by the type-2 Tracy–Widom distribution F_2 in Corollary 2. We remark that it has been shown in many random matrix theory papers that whether the entries are real or complex is not critical to the proofs of local laws and that the complex case is generally easier than the real case; see e.g., [7,8,11,24,29,31–33,35,36,46,54]. Therefore, for definiteness and due to length constraint, we follow their conventions and focus on the real setting in this paper.

5. Analysis of rectangular free convolution

The proof of the local laws in Theorem 3 depends crucially on a good understanding of the rectangular free convolution $\rho_{w,t}$ and its Stieltjes transform $m_{w,t}$. In this section, we prove some deterministic estimates on them given that $m_{w,0} \equiv m_V$ is η_*-regular as in Definition 1. In particular, we will show that $\rho_{w,t}$ has a regular square root behavior around the right edge as given by Lemma 1. When $t \sim 1$, some of the estimates have been proved in [17,18,51]. In this section we extend them to the case $n^\varepsilon \eta_* \leqslant t^2 \leqslant n^{-\varepsilon}$, which requires much more careful analysis of Eq. (10). We expect the results of this section to be of independent interest in the statistical estimation of large-rank deformed rectangular matrices.

5.1. Preliminaries

Following [17], we denote $b_t(z) := 1 + c_n t m_{w,t}(z)$. It is easy to see from (10) that b_t satisfies the following equation

$$b_t = 1 + \frac{tc_n}{p} \sum_{i=1}^{p} \frac{1}{b_t^{-1}d_i - b_t z + t(1 - c_n)}. \tag{48}$$

Using the notation b_t and Eq. (33), we can rewrite (39) as

$$\zeta_t(z) := b_t^2 z - b_t t(1-c_n), \tag{49}$$

where $\zeta_t(z)$ is actually the subordination function of the rectangular free convolution [9]. Then, Eq. (48) can be rewritten as

$$\frac{1}{c_n t}\left(1-\frac{1}{b_t}\right) = m_{w,0}(\zeta_t). \tag{50}$$

We remark that $m_{w,0}(\zeta_t)$ is well-defined because $\operatorname{Im}\zeta_t > 0$ whenever $\operatorname{Im} z > 0$; see Lemma 2.

As will be shown later, our main analysis will boil down to the study of the analytic functions $m_{w,t}$, ζ_t and b_t on $\mathbb{C}_+$. We first summarize some basic properties of them, which have been proved in previous works [17,18,51].

Lemma 2 (*Existence and Uniqueness of Asymptotic Density*)**.** *For any $t > 0$, the following properties hold.*

(i) *There exists a unique solution $m_{w,t}$ to (10) satisfying that $\operatorname{Im} m_{w,t}(z) > 0$ and $\operatorname{Im} zm_{w,t}(z) > 0$ for $z \in \mathbb{C}_+$.*
(ii) *For all $x \in \mathbb{R}\setminus\{0\}$, $\lim_{\eta\downarrow 0} m_{w,t}(x+\mathrm{i}\eta)$ exists, and we denote it as $m_{w,t}(x)$. The function $m_{w,t}$ is continuous on $\mathbb{R}\setminus\{0\}$, and $\rho_{w,t}(x) := \pi^{-1}\operatorname{Im} m_{w,t}(x)$ is a continuous probability density function on $\mathbb{R}_+ := \{x \in \mathbb{R} : x > 0\}$. Moreover, $m_{w,t}(z)$ is the Stieltjes transform of $\rho_{w,t}$. Finally, $m_{w,t}(x)$ is a solution to (10) for $z = x$.*
(iii) *For all $x \in \mathbb{R}\setminus\{0\}$, $\lim_{\eta\downarrow 0}\zeta_t(x+\mathrm{i}\eta)$ exists, and we denote it as $\zeta_t(x)$. Moreover, we have $\operatorname{Im}\zeta_t(z) > 0$ for $z \in \mathbb{C}_+$.*
(iv) *We have $\operatorname{Re} b_t(z) > 0$ for $z \in \mathbb{C}_+$, and*

$$|m_{w,t}(z)| \leqslant (c_n t|z|)^{-1/2}. \tag{51}$$

Proof. (i) follows from [18, Theorem 4.1], (ii) and (iii) follow from [17, Theorem 2.1] and [51, Proposition 1], and (iv) follows from [17, Lemma 2.1]. □

Denote the support of $\rho_{w,t}$ as $S_{w,t}$. It has been shown in [17,51] that the support and edges of $S_{w,t}$ can be completely characterized by $m_{w,t}$.

Lemma 3. *The interior $Int(S_{w,t})$ of $S_{w,t}$ is given by*

$$Int(S_{w,t}) = \{x > 0 : \operatorname{Im} m_{w,t}(x) > 0\} = \{x > 0 : \operatorname{Im}\zeta_t(x) > 0\},$$

which is a subset of $\mathbb{R}_+$. Moreover, $\zeta_t(x) \notin \{d_1, \ldots, d_p\}$ when $x \notin \partial S_{w,t}$.

Proof. This result is contained in [51, Propositions 1 and 2]. □

From Eq. (50), we can solve that

$$m_{w,t}(z) = m_{w,0}(\zeta_t)/(1 - c_n t m_{w,0}(\zeta_t)). \tag{52}$$

Plugging it into (49), we get

$$\Phi_t(\zeta_t(z)) = z, \tag{53}$$

where Φ_t is an analytic function on $\mathbb{C}_+$ defined as

$$\Phi_t(\zeta) = \zeta(1 - c_n t m_{w,0}(\zeta))^2 + (1-c_n)t(1 - c_n t m_{w,0}(\zeta)), \quad \zeta \in \mathbb{C}_+. \tag{54}$$

In [51], the authors characterize the support of $\rho_{\omega,t}$ and its edges using the local extrema of Φ_t on $\mathbb{R}$.

Lemma 4. *Fix any $t > 0$. The function $\Phi_t(x)$ on $\mathbb{R}\setminus\{0\}$ admits $2q$ positive local extrema counting multiplicities for some $q \in \mathbb{N}$. The preimages of these extrema are denoted by $\zeta_{1,-}(t) < 0 < \zeta_{1,+}(t) \leqslant \zeta_{2,-}(t) \leqslant \zeta_{2,+}(t) \leqslant \cdots \leqslant \zeta_{q,-}(t) \leqslant \zeta_{q,+}(t)$, and they belong to the set $\{\zeta \in \mathbb{R} : 1 - c_n t m_{w,0}(\zeta_t) > 0\}$. Moreover, the rightmost edge of $S_{w,t}$ is given by $\lambda_{+,t} = \Phi_t(\zeta_{q,+}(t))$, and Φ_t is increasing on the intervals $(-\infty, \zeta_{1,-}(t)]$, $[\zeta_{1,+}(t), \zeta_{2,-}(t)]$, $\ldots$, $[\zeta_{q-1,+}(t), \zeta_{q,-}(t)]$, $[\zeta_{q,+}(t), \infty)$. Finally, for $k \in \{1, \ldots, q\}$, each interval $(\zeta_{k,-}(t), \zeta_{k,+}(t))$ contains at least one element in $\{d_1, \ldots, d_p, 0\}$. In particular, we have $\zeta_{q,-}(t) < d_1 < \zeta_{q,+}(t)$.*

Proof. See [51, Proposition 3] and the discussion below [51, Theorem 2], or see [38, Lemma 1]. □

For our purpose, in some cases Eq. (50) is more convenient to use than (53). Now, we rewrite (50) into a equation of ζ_t and z only. We focus on $z \in \mathbb{C}_+$ with $\operatorname{Re} z > 0$. Then, we can solve from (49) that

$$b_t = \frac{t(1-c_n) + \sqrt{t^2(1-c_n)^2 + 4\zeta_t z}}{2z}, \tag{55}$$

where we have chosen the branch of the solution such that Lemma 2 (iv) holds. Together with (50), we find that (z, b_t) solves (50) if and only if (z, ζ_t) is a solution to

$$F_t(z,\zeta_t)=0, \quad \text{where} \quad F_t(z,\zeta_t):=1+\frac{t(1-c_n)-\sqrt{t^2(1-c_n)^2+4\zeta_t z}}{2\zeta_t}-c_n t m_{w,0}(\zeta_t). \tag{56}$$

Since $\Phi_t(\zeta_t(x))=x$ and $F_t(x,\zeta_t)=0$ are the same equation, from Lemma 4, we can derive the following characterization of the edges of $S_{w,t}$.

Lemma 5. *Denote $a_{k,\pm}(t):=\Phi_t(\zeta_{k,\pm}(t))$, $1\leqslant k\leqslant q$. Then, $(a_{k,\pm}(t),\zeta_{k,\pm}(t))$ are real solutions to the equations*

$$F_t(z,\zeta)=0, \quad \text{and} \quad \frac{\partial F_t}{\partial \zeta}(z,\zeta)=0. \tag{57}$$

Proof. By chain rule, if we regard z as a function of ζ, then we have

$$0=\frac{\mathrm{d}F_t}{\mathrm{d}\zeta}=\frac{\partial F_t}{\partial \zeta}+\frac{\partial F_t}{\partial z}z'(\zeta). \tag{58}$$

By Lemma 4, we have $\Phi_t'(\zeta_{k,\pm})=0$ since $a_{k,\pm}$ are local extrema of Φ_t. Then, from Eq. (53), we can derive

$$z'(\zeta_{k,\pm})=\Phi_t'(\zeta_{k,\pm})=0,$$

Plugging it into (58) and using $z(\zeta_{k,\pm}(t))=a_{k,\pm}(t)$ by definition, we get

$$\frac{\partial F_t}{\partial \zeta}(a_{k,\pm}(t),\zeta_{k,\pm}(t))=0,$$

which concludes the proof. □

Our proof will use extensively the estimates in Lemma 6, which are consequences of the regularity assumption in Definition 1. Define the spectral domains

$$\begin{aligned}\mathcal{D}:=&\{z=E+\mathrm{i}\eta:\lambda_+\leqslant E\leqslant\lambda_++3c_V/4,\ 2\eta_*\leqslant\eta\leqslant 10\}\\ &\cup\{z=E+\mathrm{i}\eta:\lambda_+-3c_V/4\leqslant E\leqslant\lambda_+,\ \eta_*+\sqrt{\eta_*(\lambda_+-E)}\leqslant\eta\leqslant 10\},\\ &\cup\{z=E+\mathrm{i}\eta:\lambda_++2\eta_*\leqslant E\leqslant\lambda_++3c_V/4,\ 0\leqslant\eta\leqslant 10\}.\end{aligned} \tag{59}$$

Lemma 6 (*Lemma C.1 of [34]*)**.** *Suppose V is η_*-regular in the sense of Definition 1. Recall that $m_{w,0}=m_V$ and let $\mu_{w,0}$ be the measure associated with $m_{w,0}$. For any fixed $a\geqslant 2$, the following estimates hold for $z=E+\mathrm{i}\eta\in\mathcal{D}$: if $E\leqslant\lambda_+$, we have*

$$\int\frac{\mathrm{d}\mu_{w,0}(x)}{|x-E-\mathrm{i}\eta|^a}\sim\frac{\sqrt{|E-\lambda_+|+\eta}}{\eta^{a-1}}; \tag{60}$$

if $E>\lambda_+$, we have

$$\int\frac{\mathrm{d}\mu_{w,0}(x)}{|x-E-\mathrm{i}\eta|^a}\sim\frac{1}{(|E-\lambda_+|+\eta)^{a-3/2}}. \tag{61}$$

5.2. Behavior of the contour $\zeta_t(E)$

In this subsection, we study the behaviors of $\zeta_t(E)$ for $E\in\mathbb{R}$ around the right edge λ_+ of V. Throughout the rest of this section, we assume that V is η_*-regular, and

$$t:=n^{-1/3+\omega}, \qquad 1/3-\phi_*/2+\varepsilon/2\leqslant\omega\leqslant 1/3-\varepsilon/2, \tag{62}$$

such that $n^{\varepsilon}\eta_*\ll t^2\leqslant n^{-\varepsilon}$. We will not repeat them in the assumptions of our results.

For simplicity of notations, we shall abbreviate b_t and ζ_t as b and ζ, respectively. Moreover, we centralize ζ at the right-most edge λ_+ of V as

$$\xi(z)\equiv\xi_t(z):=\zeta_t(z)-\lambda_+, \qquad \xi_+\equiv\xi_+(t):=\zeta_+(t)-\lambda_+, \tag{63}$$

where $\zeta_+(t):=\zeta_t(\lambda_{+,t})$. The following lemma gives a basic estimate on ξ_+.

Lemma 7. *For $\xi_+(t)$ defined in (63), we have $\xi_+(t)\geqslant 0$ and*

$$\xi_+(t)\sim t^2. \tag{64}$$

Proof. The statement $\xi_+(t) \geqslant 0$ follows directly from Lemma 4 because $\zeta_+(t) \equiv \zeta_{q,+}(t) \geqslant d_1 = \lambda_+$. For the estimate (64), by Lemma 4, we know that $\Phi_t(\zeta_+(t))$ is the only local extrema of $\Phi_t(\zeta)$ on the interval $(d_1, +\infty)$. Hence, we have $\Phi_t'(\zeta_+(t)) = 0$, which gives the equation

$$(1 - c_n t m_{w,0}(\zeta_+))^2 - 2c_n t m_{w,0}'(\zeta_+) \cdot \zeta_+ \left(1 - c_n t m_{w,0}(\zeta_+)\right) - c_n(1-c_n)t^2 m_{w,0}'(\zeta_+) = 0. \tag{65}$$

From this equation, we can get that

$$c_n t m_{w,0}'(\zeta_+) = \frac{(1 - c_n t m_{w,0}(\zeta_+))^2}{2\zeta_+ \left(1 - c_n t m_{w,0}(\zeta_+)\right) + (1-c_n)t}. \tag{66}$$

By (51), we have the bound

$$b = 1 + \mathrm{O}(t^{1/2}), \tag{67}$$

which gives $c_n t m_{w,0}(\zeta_+) = \mathrm{O}(t^{1/2})$ by (50). Plugging it into (66), we obtain that

$$m_{\omega,0}'(\zeta_+(t)) \sim t^{-1}. \tag{68}$$

Together with (61), it implies that $\sqrt{\xi_+(t)} \sim t$. $\square$

For $E \leqslant \lambda_{+,t}$, $\xi(E)$ has a nonzero imaginary part by Lemma 3, and we denote

$$\xi(E) := \alpha(E) + \mathrm{i}\beta(E). \tag{69}$$

We now establish an equation satisfied by α and β. We remark that this equation corresponds to equation (7.12) of [34], which takes a much simpler form than our Eq. (70) due to the simple form of additive free convolution of symmetric random matrices.

Lemma 8. *For any $E \in \mathbb{R}$, $\alpha \equiv \alpha(E)$ and $\beta \equiv \beta(E)$ satisfy the following equation:*

$$\begin{aligned} 1 - 2c_n t \int \frac{x\,\mathrm{d}\mu_{w,0}(x)}{(x-\alpha-\lambda_+)^2+\beta^2} + c_n^2 t^2 \Bigg[\left(\int \frac{x\,\mathrm{d}\mu_{w,0}(x)}{(x-\alpha-\lambda_+)^2+\beta^2} \right)^2 \\ - \frac{1-c_n}{c_n} \int \frac{\mathrm{d}\mu_{w,0}(x)}{(x-\alpha-\lambda_+)^2+\beta^2} - \left((\alpha+\lambda_+)^2+\beta^2\right) \left(\int \frac{\mathrm{d}\mu_{w,0}(x)}{(x-\alpha-\lambda_+)^2+\beta^2} \right)^2 \Bigg] = 0, \end{aligned} \tag{70}$$

where we recall $\mu_{w,0} \equiv \mu_V := p^{-1}\sum_{i=1}^p \delta_{d_i}$ is the ESD associated with V.

Proof. By (53), we have that $\Phi_t(\zeta(E)) = E$. Taking the imaginary parts of both sides of this equation and using (69), we obtain that

$$\begin{aligned} &\beta \left(1 - 2c_n t \,\mathrm{Re}\, m_{w,0}(\alpha+\mathrm{i}\beta) + c_n^2 t^2 \left[\left(\mathrm{Re}\, m_{w,0}(\alpha+\mathrm{i}\beta)\right)^2 - \left(\mathrm{Im}\, m_{w,0}(\alpha+\mathrm{i}\beta)\right)^2 \right] \right) \\ &\quad + (\alpha+\lambda_+)\left(-2c_n t \,\mathrm{Im}\, m_{w,0}(\alpha+\mathrm{i}\beta) + 2c_n^2 t^2 \,\mathrm{Im}\, m_{w,0}(\alpha+\mathrm{i}\beta) \cdot \mathrm{Re}\, m_{w,0}(\alpha+\mathrm{i}\beta)\right) \\ &\quad - c_n(1-c_n)t^2 \,\mathrm{Im}\, m_{w,0}(\alpha+\mathrm{i}\beta) = 0. \end{aligned} \tag{71}$$

After a straightforward calculation using (6), we can conclude (70). $\square$

Now, with Lemma 8, we study the behaviors of α and β for E around λ_+. The next lemma corresponds to [34, Lemma 7.1], but our proof is slightly different from the proof strategy there.

Lemma 9. *For $-3c_V/4 \leqslant \alpha(E) \leqslant \xi_+(t)$ and $c_V/8 \leqslant E \leqslant \lambda_{+,t}$, we have that*

$$\beta(E) \sim t|\alpha(E) - \xi_+(t)|^{1/2}. \tag{72}$$

Proof. Under the given condition, we have $c_V < \mathrm{Re}\,\zeta = \alpha + \lambda_+ \leqslant \lambda_+ + \mathrm{O}(t^2)$. Moreover, by (50) and (51), we have that

$$|c_n t m_{w,0}(\zeta_t)| = \left| \frac{c_n t m_{w,t}(\zeta)}{1 + c_n t m_{w,t}(\zeta)} \right| \lesssim t^{1/2}. \tag{73}$$

In the following proof, for simplicity of notations, we regard function $\Phi_t(\zeta)$ as a function of $\xi = \zeta - \lambda_+$. We consider the following two cases for ξ.

Case 1: $|\xi - \xi_+(t)| \leqslant \tau t^2$ for some small constant $\tau > 0$. By (64), for small enough $\tau > 0$, we have that $\alpha \gtrsim t^2 \gg \eta_*$. Then, using Lemma 6, we can get that for any fixed $k \geqslant 1$,

$$|c_n t m_{w,0}^{(k)}(\zeta)| \lesssim t^{-(2k-2)}, \quad \text{for } |\xi - \xi_+| \leqslant \tau t^2. \tag{74}$$

Moreover, by (64) and Lemma 6, we have that

$$|c_n t m^{(k)}_{w,0}(\zeta_+)| \sim t^{-(2k-2)}. \tag{75}$$

Note that $m^{(k)}_{w,0}(\xi_+)$ are real numbers for any k.

Now, for Eq. (53), we expand $\Phi_t(\xi)$ around ξ_+ and get that

$$E - \lambda_{+,t} = \Phi_t(\xi) - \Phi_t(\xi_+) = \frac{\Phi_t''(\xi_+)}{2}(\xi-\xi_+)^2 + \frac{\Phi_t^{(3)}(\xi_+)}{6}(\xi-\xi_+)^3 + \mathrm{O}\left(t^{-6}|\xi-\xi_+|^4\right). \tag{76}$$

Using (74), it is not hard to check that

$$|\Phi_t^{(k)}(\xi)| \lesssim t^{-(2k-2)}. \tag{77}$$

Moreover, we can calculate directly that

$$\begin{aligned}\Phi_t''(\xi) = &- 2c_n t m''_{w,0}(\zeta)\cdot\zeta(1-c_n t m_{w,0}(\zeta_t)) - 4c_n t m'_{w,0}(\zeta)\cdot(1-c_n t m_{w,0}(\zeta_t)) \\ &+ 2\zeta[c_n t m'_{w,0}(\zeta)]^2 - c_n(1-c_n)t^2 m''_{w,0}(\zeta) = -2c_n t m''_{w,0}(\zeta)\cdot\zeta + \mathrm{O}(t^{-3/2}),\end{aligned} \tag{78}$$

where we used (73) and (75) in the second step. Since

$$m''_{w,0}(\zeta_+) = \int \frac{\mathrm{d}\mu_{w,0}(x)}{(x-\zeta_+)^3} < 0,$$

we get that $\Phi_t''(\xi_+) > 0$ and

$$\Phi_t''(\xi_+) \sim t^{-2}. \tag{79}$$

Now, inverting Eq. (76) and using (77)–(79), we obtain that

$$\xi - \xi_+ = \sqrt{\frac{2(E-\lambda_{+,t})}{\Phi_t''(\xi_+)}}\left(1 - \frac{\Phi_t^{(3)}(\xi_+)}{3\Phi_t''(\xi_+)}(\xi-\xi_+) + \mathrm{O}\left(t^{-4}|\xi-\xi_+|^2\right)\right).$$

Back-substituting this equation once more gives that

$$\xi - \xi_+ = \sqrt{\frac{2(E-\lambda_{+,t})}{\Phi_t''(\xi_+)}}\left(1 - \frac{\Phi_t^{(3)}(\xi_+)}{3\Phi_t''(\xi_+)}\sqrt{\frac{2(E-\lambda_{+,t})}{\Phi_t''(\xi_+)}} + \mathrm{O}\left(t^{-4}|\xi-\xi_+|^2\right)\right). \tag{80}$$

Taking the real and imaginary parts of the above equation and using (77) and (79), we obtain that

$$|\alpha - \xi_+| \sim |E-\lambda_{+,t}|, \quad \beta \sim t|E-\lambda_{+,t}|^{1/2} \sim t|\alpha-\xi_+|^{1/2}, \tag{81}$$

for $|\xi - \xi_+| \leqslant \tau t^2$.

Case 2: $-3c_V/4 < \alpha \leqslant \xi_+ - \tau_1 t^2$ for a small constant $\tau_1 > 0$. In this case, we have $t|\alpha-\xi_+|^{1/2} \gtrsim t^2$ by (81) as long as τ_1 is small enough. First, suppose that $\beta \gg t|\alpha-\xi_+|^{1/2}$. Then, using Lemma 6 and $|\alpha| \lesssim |\alpha-\xi_+|$, we get that

$$\int \frac{x\mathrm{d}\mu_{w,0}(x)}{(x-\alpha-\lambda_+)^2+\beta^2} \lesssim \int \frac{\mathrm{d}\mu_{w,0}(x)}{(x-\alpha-\lambda_+)^2+\beta^2} = \mathrm{o}(t^{-1}).$$

This contradicts (70), so we must have $\beta \lesssim t|\alpha-\xi_+|^{1/2}$.

On the other hand, suppose $\beta \ll t|\alpha-\xi_+|^{1/2}$. For any small constant $\delta > 0$, we take $\beta_0 := \delta t|\alpha-\xi_+|^{1/2}$ and $\zeta_0 := (\alpha+\lambda_+) + \mathrm{i}\beta_0$. Then, we can check that

$$\frac{\operatorname{Im} m_{w,0}(\zeta)}{\beta} = \int \frac{\mathrm{d}\mu_{w,0}(x)}{(x-\alpha-\lambda_+)^2+\beta^2} \geqslant \int \frac{\mathrm{d}\mu_{w,0}(x)}{(x-\alpha-\lambda_+)^2+\beta_0^2} = \frac{\operatorname{Im} m_{w,0}(\zeta_0)}{\beta_0}.$$

Using Lemma 6, we can bound:

$$\frac{\operatorname{Im} m_{w,0}(\zeta_0)}{\beta_0} \geqslant \frac{c_1}{\sqrt{\beta_0}} \geqslant \frac{c_2}{t\sqrt{\delta}}, \quad -t^2 \leqslant \alpha \leqslant \xi_+ - \tau_1 t^2; \quad \frac{\operatorname{Im} m_{w,0}(\zeta_0)}{\beta_0} \geqslant \frac{c_1\sqrt{|\alpha|+\beta_0}}{\beta_0} \geqslant \frac{c_2}{t\delta}, \quad \alpha < -t^2.$$

Here, $c_1, c_2 > 0$ are constants that do not depend on δ. Now, taking the imaginary part of Eq. (56), we obtain that

$$\frac{1}{\beta}\operatorname{Im}\frac{t(1-c_n) - \sqrt{t^2(1-c_n)^2+4\zeta E}}{2\zeta} = \frac{c_n t \operatorname{Im} m_{w,0}(\zeta_t)}{\beta} \geqslant \frac{c_n c_2}{\sqrt{\delta}}. \tag{82}$$

On the other hand, using $|\zeta| \sim 1$ and $t = \mathrm{o}(1)$, we can upper bound the left-hand side of (82) by a constant $C > 0$ that does not depend on δ. This gives a contradiction if δ is taken sufficiently small. Hence, we must also have $\beta \gtrsim t|\alpha-\xi_+|^{1/2}$. $\square$

Based on Lemma 9, we are able to prove the following result.

Lemma 10. *For $-3c_V/4 \leqslant \alpha \leqslant \xi_+(t)$, we have*

$$|\alpha(E) - \xi_+(t)| \sim |E - \lambda_{+,t}|. \tag{83}$$

Proof. Note that by (81), (83) holds true when $|\alpha - \xi_+(t)| \leqslant \tau t^2$. To conclude the proof, it suffices to show that $\mathrm{d}\alpha/\mathrm{d}E \geqslant 0$ and

$$\frac{\mathrm{d}\alpha}{\mathrm{d}E} \sim 1, \quad \text{for} \quad |\alpha - \xi_+| \geqslant \tau t^2. \tag{84}$$

From Eq. (53), we obtain that

$$\frac{\mathrm{d}\alpha}{\mathrm{d}E} = \mathrm{Re}\, \frac{1}{\Phi_t'(\zeta)} = \frac{\mathrm{Re}\, \Phi_t'(\zeta)}{|\Phi_t'(\zeta)|^2}. \tag{85}$$

After some calculations,

$$\begin{aligned}
\mathrm{Re}\, \Phi_t'(\zeta) =& 1 + 2c_n t \left[\int \frac{-x}{(x-\alpha-\lambda_+)^2 + \beta^2} \mathrm{d}\mu_{w,0}(x) + \int \frac{2\beta^2 x}{\left[(x-\alpha-\lambda_+)^2 + \beta^2\right]^2} \mathrm{d}\mu_{w,0}(x) \right] \\
&+ c_n^2 t^2 \left\{ \left(\int \frac{x-\alpha-\lambda_+}{(x-\alpha-\lambda_+)^2 + \beta^2} \mathrm{d}\mu_{w,0}(x) \right)^2 - \left(\int \frac{\beta}{(x-\alpha-\lambda_+)^2 + \beta^2} \mathrm{d}\mu_{w,0}(x) \right)^2 \right. \\
&+ 2(\alpha+\lambda_+) \int \frac{x-\alpha-\lambda_+}{(x-\alpha-\lambda_+)^2 + \beta^2} \mathrm{d}\mu_{w,0}(x) \cdot \int \frac{(x-\alpha-\lambda_+)^2 - \beta^2}{[(x-\alpha-\lambda_+)^2 + \beta^2]^2} \mathrm{d}\mu_{w,0}(x) \\
&- 4(\alpha+\lambda_+) \int \frac{\beta^2}{(x-\alpha-\lambda_+)^2 + \beta^2} \mathrm{d}\mu_{w,0}(x) \cdot \int \frac{x-\alpha-\lambda_+}{[(x-\alpha-\lambda_+)^2 + \beta^2]^2} \mathrm{d}\mu_{w,0}(x) \\
&- 4 \int \frac{\beta^2 (x-\alpha-\lambda_+)}{[(x-\alpha-\lambda_+)^2 + \beta^2]^2} \mathrm{d}\mu_{w,0}(x) \cdot \int \frac{x-\alpha-\lambda_+}{(x-\alpha-\lambda_+)^2 + \beta^2} \mathrm{d}\mu_{w,0}(x) \\
&- 2 \int \frac{\beta^2}{(x-\alpha-\lambda_+)^2 + \beta^2} \mathrm{d}\mu_{w,0}(x) \cdot \int \frac{(x-\alpha-\lambda_+)^2 - \beta^2}{[(x-\alpha-\lambda_+)^2 + \beta^2]^2} \mathrm{d}\mu_{w,0}(x) \\
&\left. - \frac{1-c_n}{c_n} \int \frac{(x-\alpha-\lambda_+)^2 - \beta^2}{[(x-\alpha-\lambda_+)^2 + \beta^2]^2} \mathrm{d}\mu_{w,0}(x) \right\}.
\end{aligned} \tag{86}$$

Then, using (70), we can rewrite the above equation as

$$\mathrm{Re}\, \Phi_t'(\zeta) = \int \frac{4c_n t \beta^2 x}{\left[(x-\alpha-\lambda_+)^2 + \beta^2\right]^2} \mathrm{d}\mu_{w,0}(x) - \mathrm{R}, \tag{87}$$

where R is defined as

$$\begin{aligned}
\mathrm{R} :=& 4c_n^2 t^2 (\alpha+\lambda_+) \int \frac{\beta^2}{(x-\alpha-\lambda_+)^2 + \beta^2} \mathrm{d}\mu_{w,0}(x) \cdot \int \frac{x-\alpha-\lambda_+}{[(x-\alpha-\lambda_+)^2 + \beta^2]^2} \mathrm{d}\mu_{w,0}(x) \\
&+ 4c_n^2 t^2 \int \frac{\beta^2 x}{[(x-\alpha-\lambda_+)^2 + \beta^2]^2} \mathrm{d}\mu_{w,0}(x) \cdot \int \frac{x-\alpha-\lambda_+}{(x-\alpha-\lambda_+)^2 + \beta^2} \mathrm{d}\mu_{w,0}(x) \\
&+ 2c_n^2 t^2 \int \frac{\beta^2}{(x-\alpha-\lambda_+)^2 + \beta^2} \mathrm{d}\mu_{w,0}(x) \cdot \int \frac{(x-\alpha-\lambda_+)^2 - \beta^2}{[(x-\alpha-\lambda_+)^2 + \beta^2]^2} \mathrm{d}\mu_{w,0}(x) \\
&+ 2c_n(1-c_n) t^2 \int \frac{(x-\alpha-\lambda_+)^2}{[(x-\alpha-\lambda_+)^2 + \beta^2]^2} \mathrm{d}\mu_{w,0}(x) =: \mathrm{R}_1 + \mathrm{R}_2 + \mathrm{R}_3 + \mathrm{R}_4.
\end{aligned} \tag{88}$$

Now, we estimate $\mathrm{Re}\, \Phi_t'(\zeta)$ using Lemma 6. We first consider the case $\alpha \leqslant \tau_1 t^2$ for a small enough constant $\tau_1 > 0$. In this case, by (64) and (72), we have that $\zeta = (\alpha + \lambda_+) + \mathrm{i}\beta \in \mathcal{D}$ defined in (59). Then, using Lemma 6 and (72), we obtain that

$$\int \frac{\mathrm{d}\mu_{w,0}(x)}{(x-\alpha-\lambda_+)^2 + \beta^2} \sim \frac{\sqrt{|\alpha| + \beta}}{\beta} \sim \frac{\sqrt{|\alpha| + t|\alpha - \xi_+(t)|^{1/2}}}{t|\alpha - \xi_+(t)|^{1/2}} \sim t^{-1}, \tag{89}$$

where we used $|\alpha| + \beta \sim \beta$ for $\alpha \geqslant 0$ in the first step, (72) in the second step, and $\alpha \leqslant \xi_+(t) - \tau t^2$ by (64) in the last step. Similarly, we have

$$\int \frac{\mathrm{d}\mu_{w,0}(x)}{\left[(x-\alpha-\lambda_+)^2 + \beta^2\right]^2} \sim \frac{\sqrt{|\alpha| + t|\alpha - \xi_+(t)|^{1/2}}}{\beta^3} \sim t^{-1}\beta^{-2}, \tag{90}$$

$$\int \frac{(x-\alpha-\lambda_+)^2}{\left[(x-\alpha-\lambda_+)^2+\beta^2\right]^2} \mathrm{d}\mu_{w,0}(x) \leqslant \int \frac{\mathrm{d}\mu_{w,0}(x)}{(x-\alpha-\lambda_+)^2+\beta^2} \lesssim t^{-1}. \tag{91}$$

Using (89)–(91), we can bound each term in R as:

$$|\mathrm{R}_1| \lesssim t^2 \frac{\beta^2}{t} \left[\int \frac{(x-\alpha-\lambda_+)^2}{[(x-\alpha-\lambda_+)^2+\beta^2]^2} \mathrm{d}\mu_{w,0}(x)\right]^{1/2} \left[\int \frac{\mathrm{d}\mu_{w,0}(x)}{[(x-\alpha-\lambda_+)^2+\beta^2]^2}\right]^{1/2} \lesssim \beta \lesssim t,$$

$$|\mathrm{R}_2| \lesssim t^2 \frac{\beta^2}{t\beta^2} \left[\int \frac{(x-\alpha-\lambda_+)^2}{[(x-\alpha-\lambda_+)^2+\beta^2]^2} \mathrm{d}\mu_{w,0}(x)\right]^{1/2} \lesssim t^{1/2}, \quad |\mathrm{R}_3| \lesssim t^2 \cdot \frac{\beta^2}{t} \cdot t^{-1} \lesssim t^2, \quad |\mathrm{R}_4| \lesssim t^2 \cdot t^{-1} \leqslant t.$$

Summing them, we get that

$$|\mathrm{R}| \leqslant |\mathrm{R}_1| + |\mathrm{R}_2| + |\mathrm{R}_3| + |\mathrm{R}_4| \lesssim t^{1/2}.$$

On the other hand, by (90), we have

$$\int \frac{4c_n t\beta^2 x}{\left[(x-\alpha-\lambda_+)^2+\beta^2\right]^2} \mathrm{d}\mu_{w,0}(x) \sim 1.$$

Hence, (87) gives that $\operatorname{Re} \Phi_t'(\zeta) \sim 1$, which also gives a lower bound $|\Phi_t'(\zeta)| \geqslant \operatorname{Re} \Phi_t'(\zeta) \gtrsim 1$. For an upper bound of $|\Phi_t'(\zeta)|$, we get

$$|\Phi_t'(\zeta)| = \left|(1-c_n t m_{w,0}(\zeta_t))^2 - 2c_n t m_{w,0}'(\zeta)\cdot\zeta\left(1-c_n t m_{w,0}(\zeta_t)\right) - c_n(1-c_n)t^2 m_{w,0}'(\zeta)\right| \lesssim 1,$$

using (73) and

$$c_n t |m_{w,0}'(\zeta)| \leqslant c_n t \int \frac{\mathrm{d}\mu_{w,0}(x)}{(x-\alpha-\lambda_+)^2+\beta^2} \sim 1$$

by (89). This concludes (84) when $\alpha \leqslant \tau_1 t^2$.

For the case $\tau_1 t^2 \leqslant \alpha \leqslant \xi_+(t) - \tau t^2$, the proof is similar and the only difference is that we shall use (61) to estimate each term. We omit the details. □

5.3. Behavior of $\zeta_t(z)$ on general domain

In this subsection, we first extend the result of Lemma 9 to $\xi(z) = \alpha(z) + \mathrm{i}\beta(z)$ for complex $z = E + \mathrm{i}\eta$ around the right edge λ_+. In the proof, we will regard α and β as functions of both E and η. First, we claim the following simple estimate.

Lemma 11. *Suppose $-3c_V/4 \leqslant \alpha \leqslant \xi_+(t) - \tau t^2$ for a constant $\tau > 0$, $|E-\lambda_+| \leqslant c_V/2$ and $0 \leqslant \eta \leqslant 10$. Then, we have*

$$\beta(E,\eta) \geqslant c_\tau t|\alpha(E,\eta)-\xi_+|^{1/2} \tag{92}$$

for some constant $c_\tau > 0$.

Proof. First, taking the imaginary parts of both sides of (49), we get

$$\beta = \eta \operatorname{Re}\left(1+c_n t m_{w,t}(z)\right)^2 + E\left[2c_n t \operatorname{Im} m_{w,t}(z) + c_n^2 t^2 \operatorname{Im} m_{w,t}(z)\cdot \operatorname{Re} m_{w,t}(z)\right] - c_n(1-c_n)t^2 \operatorname{Im} m_{w,t}(z). \tag{93}$$

On the other hand, taking the imaginary part of Eq. (56) and using (52), we get that

$$t \operatorname{Im} m_{w,t}(z) \lesssim t \operatorname{Im} m_{w,0}(\zeta) \lesssim \beta + \eta.$$

Plugging it into (93) and using (51), we get that for some constant $C > 0$,

$$\beta \geqslant \eta - Ct^{1/2}(\beta+\eta) \Rightarrow \beta \geqslant \frac{1}{2}\eta. \tag{94}$$

The rest of the proof is similar to Case 2 in the proof of Lemma 9. Suppose $\beta \ll t|\alpha-\xi_+|^{1/2}$. For any small constant $\delta > 0$, we take $\beta_0 := \delta t|\alpha-\xi_+|^{1/2}$ and $\zeta_0 := (\alpha+\lambda_+) + \mathrm{i}\beta_0$. Then, using Lemma 6, we can get that

$$\frac{\operatorname{Im} m_{w,0}(\zeta_t)}{\beta} \geqslant \frac{\operatorname{Im} m_{w,0}(\zeta_0)}{\beta_0} \geqslant \frac{c_2}{t\sqrt{\delta}},$$

for some constant $c_2 > 0$ that does not depend on δ. On the other hand, taking the imaginary part of (56), we get that for some constant $C > 0$ independent of δ,

$$\frac{c_2}{\sqrt{\delta}} \leqslant \frac{t \operatorname{Im} m_{w,0}(\zeta_t)}{\beta} = \frac{1}{c_n\beta} \operatorname{Im} \frac{t(1-c_n) - \sqrt{t^2(1-c_n)^2+4\zeta z}}{2\zeta} \leqslant \frac{C}{c_n}\frac{\beta+\eta}{\beta} \leqslant \frac{3C}{c_n},$$

where we used (94) in the last step. This gives a contradiction if δ is taken sufficiently small. Hence, we must have $\beta \gtrsim t|\alpha-\xi_+|^{1/2}$. □

Then, we prove the following estimate.

Lemma 12. *For $|E-\lambda_+| \leqslant c_V/2$ and $0 \leqslant \eta \leqslant 10$, we have that*

$$|\Phi_t'(\zeta)| \sim \min\left\{1, \frac{|\alpha-\xi_+(t)|+\beta}{t^2}\right\}. \tag{95}$$

Proof. We first assume that $|\alpha-\xi_+(t)|+\beta \leqslant c_1 t^2$ for some small constant $c_1>0$. In this case, applying the mean value theorem to $\Phi_t'(\xi)$, we get

$$\Phi_t'(\xi) = \Phi_t''(\xi_+(t))(\xi-\xi_+(t))\cdot\left[1+\mathrm{O}(t^{-2}|\xi-\xi_+(t)|)\right], \tag{96}$$

where we used (77), (79) and $\Phi_t'(\xi_+)=0$. Hence, for a small enough c_1, we have

$$|\Phi_t'(\xi)| \sim |\Phi_t''(\xi_+(t))||\xi-\xi_+(t)| \sim t^{-2}|\xi-\xi_+(t)|.$$

It remains to prove that $|\Phi_t'(\xi)|\sim 1$ when $|\alpha-\xi_+(t)|+\beta > c_1t^2$. The proof is based on a careful analysis of $\mathrm{Re}\,\Phi_t'(\xi)$. First, we observe that (86) still holds for general $z=E+\mathrm{i}\eta$. On the other hand, the right-hand side of Eq. (71) is now η, and hence (87) becomes

$$\mathrm{Re}\,\Phi_t'(\zeta) = \frac{\eta}{\beta} + \int \frac{4c_n t\beta^2 x}{\left[(x-\alpha-\lambda_+)^2+\beta^2\right]^2}\mathrm{d}\mu_{w,0}(x) - \mathrm{R}. \tag{97}$$

We now estimate $\mathrm{Re}\,\Phi_t'(\zeta)$ using (86) and (97).

Case 1: We first consider the case where $\alpha \leqslant \xi_+(t)-\tau t^2$ for a small constant $\tau>0$. By (92), we have that $\zeta=(\alpha+\lambda_+)+\mathrm{i}\beta\in\mathcal{D}$. Then, with the same arguments as in the proof of Lemma 10, we can derive two estimates from (86) and (97):

$$\mathrm{Re}\,\Phi_t'(\zeta) = 1+\mathrm{o}(1)+2c_n t\left[\int\frac{-x}{(x-\alpha-\lambda_+)^2+\beta^2}\mathrm{d}\mu_{w,0}(x)+\int\frac{2\beta^2x}{\left[(x-\alpha-\lambda_+)^2+\beta^2\right]^2}\mathrm{d}\mu_{w,0}(x)\right], \tag{98}$$

$$\mathrm{Re}\,\Phi_t'(\zeta) = \frac{\eta}{\beta}+\int\frac{4c_nt\beta^2x}{\left[(x-\alpha-\lambda_+)^2+\beta^2\right]^2}\mathrm{d}\mu_{w,0}(x)+\mathrm{o}(1). \tag{99}$$

Denote $Q := t\sqrt{|\alpha|+\beta}/\beta$. Using Lemma 6, we get that

$$2c_nt\int\frac{x}{(x-\alpha-\lambda_+)^2+\beta^2}\mathrm{d}\mu_{w,0}(x) \sim \mathbf{1}_{\alpha\leqslant 0}\frac{t\sqrt{|\alpha|+\beta}}{\beta}+\mathbf{1}_{\alpha>0}\frac{t}{\sqrt{|\alpha|+\beta}} \sim Q,$$

where we used $\beta\gtrsim\alpha+\beta$ for $0<\alpha\leqslant\xi_+(t)-\tau t^2$ by (92). Similarly, we have

$$4c_nt\int\frac{\beta^2x}{\left[(x-\alpha-\lambda_+)^2+\beta^2\right]^2}\mathrm{d}\mu_{w,0}(x)\sim t\beta^2\frac{\sqrt{|\alpha|+\beta}}{\beta^3}=Q.$$

Inserting these two estimates into (98) and (99), we obtain that for some constants $c, C>0$,

$$\mathrm{Re}\,\Phi_t'(\zeta) \geqslant \max\{1-CQ, cQ\} \gtrsim 1,$$

which gives a lower bound for $|\Phi_t'(\zeta)|$. Finally, using Lemma 6 and (92), it is easy to check that $|\Phi_t'(\zeta)|\lesssim 1$. Hence, we obtain the estimates $\mathrm{Re}\,\Phi_t'(\zeta)\sim|\Phi_t'(\zeta)|\sim 1$.

Case 2: Second, we assume that $|\alpha-\xi_+(t)|+\beta\geqslant C_1t^2$ for some large constant $C_1>0$ and $\alpha\geqslant\xi_+(t)-\tau t^2$. Then, $\zeta=(\alpha+\lambda_+)+\mathrm{i}\beta\in\mathcal{D}$ with $\alpha\gtrsim t^2$ by (64). Hence, using (61), we get that

$$t\int\frac{2c_nx}{(x-\alpha-\lambda_+)^2+\beta^2}\mathrm{d}\mu_{w,0}(x)\geqslant c\frac{t}{\sqrt{\alpha+\beta}},$$

for some constant $c>0$ that does not depend on C_1. On the other hand, we have

$$0\leqslant t\int\frac{4c_n\beta^2x}{\left[(x-\alpha-\lambda_+)^2+\beta^2\right]^2}\mathrm{d}\mu_{w,0}(x)\leqslant C\frac{t\beta^2}{(\alpha+\beta)^{5/2}}\leqslant C\frac{t}{\sqrt{\alpha+\beta}},$$

for some constant $C>0$ that does not depend on C_1. Therefore, we conclude that as long as C_1 is chosen large enough, then

$$\frac{1}{2}\leqslant 1+t\left[\int\frac{-2c_nx}{(x-\alpha-\lambda_+)^2+\beta^2}\mathrm{d}\mu_{w,0}(x)+\int\frac{4c_n\beta^2x}{\left[(x-\alpha-\lambda_+)^2+\beta^2\right]^2}\mathrm{d}\mu_{w,0}(x)\right]\leqslant\frac{3}{2}.$$

Moreover, using (61) we can readily show that the rest of the terms on the right-hand side of (86) are all of order $o(1)$, and that $|\Phi_t'(\zeta)| \lesssim 1$. We omit the details since they are similar to the arguments in the proof of Lemma 10. In sum, we obtain that $\mathrm{Re}\,\Phi_t'(\zeta) \sim |\Phi_t'(\zeta)| \sim 1$ for Case 2.

Case 3: It remains to consider the case $c_1 t^2 \leqslant |\alpha - \xi_+(t)| + \beta \leqslant C_1 t^2$ and $\alpha \geqslant \xi_+(t) - \tau t^2$. If $|\alpha - \xi_+(t)| \leqslant c_1 t^2/2$, we have $\beta \geqslant c_1 t^2/2$. Then, using Lemma 6, we can check that $|\Phi_t'(\zeta)| = O(1)$, (99) still holds, and

$$4c_n t \int \frac{\beta^2 x}{\left[(x-\alpha-\lambda_+)^2+\beta^2\right]^2} \mathrm{d}\mu_{w,0}(x) \sim t\frac{\sqrt{|\alpha|+\beta}}{\beta} \sim 1.$$

Thus, we get $\mathrm{Re}\,\Phi_t'(\zeta) \sim |\Phi_t'(\zeta)| \sim 1$.

In the above proof, we can take the constants such that $\tau \leqslant c_1/2$. Then, we are only left with the regime $\alpha \geqslant \xi_+(t) + c_1 t^2/2$ and $c_1 t^2 \leqslant |\alpha - \xi_+(t)| + \beta \leqslant C_1 t^2$. In this regime, we have $\zeta = (\alpha + \lambda_+) + \mathrm{i}\beta \in \mathcal{D}$ with $\alpha \gtrsim t^2$ by (64). Then, using (61) and the same arguments as in the proof of Lemma 10, we can check that $|\Phi_t'(\zeta)| = O(1)$, (99) still holds, and that

$$4c_n t \int \frac{\beta^2 x}{\left[(x-\alpha-\lambda_+)^2+\beta^2\right]^2} \mathrm{d}\mu_{w,0}(x) \sim t\frac{\beta^2}{(|\alpha|+\beta)^{5/2}} \sim 1.$$

Thus, we get $\mathrm{Re}\,\Phi_t'(\zeta) \sim |\Phi_t'(\zeta)| \sim 1$. This completes the proof. □

Armed with Lemma 12, we can prove the following estimates. Recall the notation (36).

Lemma 13. *If $\kappa + \eta \leqslant \tau_1 t^2$ for a sufficiently small constant $\tau_1 > 0$, then we have that*

$$t\sqrt{\kappa+\eta} \sim |\xi - \xi_+(t)|, \tag{100}$$

which also implies that

$$|\Phi_t'(\zeta)| \sim \min\left\{1, \frac{\sqrt{\kappa+\eta}}{t}\right\}. \tag{101}$$

In the region $|\kappa + \eta| \geqslant \tau t^2$ for any constant $\tau > 0$, we have

$$\frac{\partial\alpha}{\partial E} = \frac{\partial\beta}{\partial\eta} \sim 1, \quad \left|\frac{\partial\alpha}{\partial\eta}\right| = \left|\frac{\partial\beta}{\partial E}\right| \lesssim 1. \tag{102}$$

The above two estimates imply that

$$|\alpha| + |\alpha - \xi_+(t)| + \beta \lesssim t^2 + \eta + \kappa. \tag{103}$$

Proof. If $|\xi - \xi_+(t)| \leqslant c_1 t^2$ for some small constant $c_1 > 0$, then with the same Taylor expansion argument as in the proof of Lemma 9, we can obtain that (recall (80))

$$\xi - \xi_+ = \sqrt{\frac{2(z-\lambda_{+,t})}{\Phi_t''(\xi_+)}}\left(1 - \frac{\Phi_t^{(3)}(\xi_+)}{3\Phi_t''(\xi_+)}\sqrt{\frac{2(z-\lambda_{+,t})}{\Phi_t''(\xi_+)}} + O\left(t^{-4}|\xi-\xi_+|^2\right)\right). \tag{104}$$

As long as c_1 is small enough, we have

$$|\xi - \xi_+| \sim \sqrt{\frac{|z-\lambda_{+,t}|}{\Phi_t''(\xi_+)}} \sim t|z-\lambda_{+,t}|^{1/2} \sim t\sqrt{\kappa+\eta}, \tag{105}$$

where we used the estimates (77) and (79). Moreover, with (104) we observe that there exists a constant $\tau_1 > 0$ such that $|\Phi_t^{-1}(z) - \xi_+(t)| \leqslant c_1 t^2$ for all z with $\kappa + \eta \leqslant \tau_1 t^2$, where Φ_t^{-1} denotes the inverse function of Φ_t. This concludes (100) together with (105). Moreover, inserting (100) into (95) we get (101).

Now, we consider the region $|\kappa + \eta| \geqslant \tau t^2$ for a constant $\tau > 0$. In the proof of Lemma 12, we have shown that $\mathrm{Re}\,\Phi_t'(\zeta) \sim |\Phi_t'(\zeta)| \sim 1$. Together with (85), we get the first estimate in (102), where the first equality comes from Cauchy–Riemann equation. Similarly, we have

$$\left|\frac{\partial\alpha}{\partial\eta}\right| = \left|\frac{\partial\beta}{\partial E}\right| \leqslant \frac{1}{|\Phi_t'(\zeta)|} \lesssim 1,$$

which gives the second estimate in (102). Finally, (103) is an easy consequence of (102). □

Remark 8. Besides (100), we will also use the expansion (104) in the following proof, which gives more precise behavior of ξ near ξ_+. In the following proof, whenever we refer to Lemma 13, it also includes (104).

Next, we collect some useful estimates that are needed in the proof of local laws. They are established on different spectral domains. Recall the definition of $\mathcal{D}_\vartheta$ in (37).

Lemma 14. *Fix any constant $C_1 > 0$. For $z \in \mathcal{D}_\vartheta$ with $E \leqslant \lambda_{+,t} + C_1 t^2$, the following estimates hold for $a = O(1)$:*

$$\min_{1\leqslant i\leqslant p} |d_i - \zeta| \gtrsim t^2 + \eta + t\,\mathrm{Im}\, m_{w,t}; \tag{106}$$

$$\int \frac{\mathrm{d}\mu_{w,0}(x)}{|x-\zeta|^a} \lesssim \frac{t+\sqrt{\kappa+\eta}}{(t^2+\mathrm{Im}\,\zeta)^{a-1}}, \quad a \geqslant 2; \tag{107}$$

$$t + \sqrt{\kappa+\eta} \lesssim t + \mathrm{Im}\, m_{w,t}. \tag{108}$$

Proof. Using (49) and (51), we can obtain that

$$\beta = \mathrm{Im}\left[(1+c_n t m_{w,t}(z))^2 z - t(1-c_n)(1+c_n t m_{w,t}(z))\right] \sim \eta + t\,\mathrm{Im}\, m_{w,t}. \tag{109}$$

Since d_i's are real values, we get that $|d_i - \zeta| \geqslant \mathrm{Im}\,\zeta \gtrsim \eta + t\,\mathrm{Im}\, m_{w,t}$. Thus, to show (106), it remains to show that

$$\min_{1\leqslant i\leqslant p} |d_i - \zeta| \geqslant \tau t^2, \tag{110}$$

for a constant $\tau > 0$. If $E - \lambda_{+,t} \geqslant c_1 t^2$ for a small constant $c_1 > 0$, then by Lemma 13, we have $\alpha(z) - \xi_+(t) \gtrsim t^2$. Hence, by (64) and $d_i \leqslant \lambda_+$,

$$\min_{1\leqslant i\leqslant p} |d_i - \zeta| \geqslant \min_{1\leqslant i\leqslant p} |d_i - \alpha| \gtrsim t^2.$$

If $\mathrm{Im}\,\zeta \geqslant c_1 t^2$, then

$$\min_{1\leqslant i\leqslant p} |d_i - \zeta| \geqslant \mathrm{Im}\,\zeta \geqslant c_1 t^2.$$

Finally, we are left with the case $\mathrm{Im}\,\zeta \leqslant c_1 t^2$ and $E \leqslant \lambda_{+,t} + c_1 t^2$ for a small enough constant $c_1 > 0$. In this case, we claim that

$$\kappa \leqslant C c_1 t^2, \tag{111}$$

for some constant $C > 0$ that does not depend on c_1. If (111) does not hold, then by Lemma 13 we must have $\mathrm{Im}\,\zeta \geqslant C_2\sqrt{Cc_1}t^2$ for some constant C_2 that does not depend on C and c_1, which gives a contradiction for large enough C. Now, given (111), by (100) we can choose $c_1 > 0$ small enough such that $|\xi - \xi_t| \leqslant \tau_1 t^2$ for a small constant $\tau_1 > 0$. Together with (64) and $d_i^2 \leqslant \lambda_+$, we conclude that

$$|d_i^2 - \zeta| = |d_i^2 - (\lambda_+ + \xi_{+,t}) - (\xi - \xi_{+,t})| \geqslant |\xi_+(t)| - |\xi - \xi_{+,t}| \geqslant \tau t^2.$$

For (107), we first suppose that $\mathrm{Im}\,\zeta \leqslant \tau t^2$ for a small constant $\tau > 0$. As shown in the above proof of (106), we must have that $\alpha \geqslant \tau t^2$ as long as τ is sufficiently small. Thus, using (61), we get

$$\int \frac{\mathrm{d}\mu_{w,0}(x)}{|x-\zeta|^a} \lesssim \frac{1}{(t^2+\mathrm{Im}\,\xi)^{a-3/2}} \lesssim \frac{t}{(t^2+\mathrm{Im}\,\zeta)^{a-1}}.$$

Then, we consider the case $\mathrm{Im}\,\zeta \geqslant \tau t^2$. In this case, (60) always serves as an upper bound regardless of the sign of α. Hence,

$$\int \frac{\mathrm{d}\mu_{w,0}(x)}{|x-\zeta|^a} \lesssim \frac{\sqrt{|\alpha|+\beta}}{\beta^{a-1}} \lesssim \frac{\sqrt{|\alpha|+\beta}}{(t^2+\mathrm{Im}\,\zeta)^{a-1}}.$$

Then, we conclude (107) using (103).

Finally, the estimate (108) follows from (119) below. □

Lemma 15. *Fix any constant $C_1 > 0$. For $z \in \mathcal{D}_\vartheta$ with $E \geqslant \lambda_{+,t} + C_1 t^2$, the following estimates hold for $a = O(1)$:*

$$\min_{1\leqslant i\leqslant p} |d_i - \zeta| \gtrsim t^2 + \kappa + \eta; \tag{112}$$

$$\int \frac{\mathrm{d}\mu_{w,0}(x)}{|x-\zeta|^a} \lesssim \frac{1}{(\kappa+\eta+t^2)^{a-3/2}}, \quad a \geqslant 2. \tag{113}$$

Proof. From the proof of Lemma 14, we have seen that $|d_i - \zeta| \gtrsim t^2$. Moreover, since $\mathrm{Im}\,\zeta \gtrsim \eta$, we have $|d_i - \zeta| \gtrsim \eta$. Finally, if $E \geqslant \lambda_{+,t} + C(\eta + t^2)$ for some large constant $C > 0$, then by Lemma 13, we have $\alpha \geqslant c_1\kappa$ for a constant $c_1 > 0$. Hence, using $d_i \leqslant \lambda_+$, we get

$$\min_{1\leqslant i\leqslant p} |d_i - \zeta| \geqslant \alpha \gtrsim \kappa.$$

This concludes (112). For (113), if $E \geqslant \lambda_{+,t} + C(\eta + t^2)$, then by (61) we have

$$\int \frac{\mathrm{d}\mu_{w,0}(x)}{|x-\zeta|^a} \lesssim \frac{1}{\kappa^{a-3/2}} \lesssim \frac{1}{(\kappa+\eta+t^2)^{a-3/2}}.$$

On the other hand, suppose $C_1 t^2 \leqslant E - \lambda_{+,t} \leqslant C(\eta + t^2)$. Then, we have $\alpha \gtrsim t^2$ by Lemma 13 and $\beta \gtrsim \eta$ by (109). Thus, using (61), we get

$$\int \frac{\mathrm{d}\mu_{w,0}(x)}{|x-\zeta|^a} \lesssim \frac{1}{(t^2+\eta)^{a-3/2}} \lesssim \frac{1}{(\kappa+\eta+t^2)^{a-3/2}},$$

where we used $\kappa \lesssim \eta + t^2$ in the second step. This concludes (113). □

Lemma 16. *If $\kappa + \eta \leqslant \tau_1 t^2$ for a sufficiently small constant $\tau_1 > 0$, then we have*

$$|m''_{w,0}(\zeta)| \sim t^{-3}. \tag{114}$$

Proof. If $\kappa + \eta \leqslant \tau_1 t^2$, then by (100) we have $\zeta \in \mathcal{D}$ with $\alpha \gtrsim t^2$ and $|\xi| \sim t^2$. Thus, using (61), we get

$$\int \frac{\mathrm{d}\mu_{w,0}(x)}{|x-\zeta|^3} \sim t^{-3}.$$

Furthermore, by (64) and (100), we have $\alpha \geqslant ct^2$ and $\operatorname{Im}\zeta \leqslant C\sqrt{\tau_1}t^2$ for some constants $c, C > 0$ that do not depend on τ_1. As long as τ_1 is taken sufficiently small, we have $-\operatorname{Re}(x-\zeta)^{-3} \sim |x-\zeta|^{-3}$ for all $x \in \operatorname{supp}(\mu_{w,0})$. Thus, we get

$$|m''_{w,0}(\zeta)| = 2\left|\int \frac{\mathrm{d}\mu_{w,0}(x)}{(x-\zeta)^3}\right| \sim \int \frac{\mathrm{d}\mu_{w,0}(x)}{|x-\zeta|^3} \sim t^{-3},$$

which concludes the proof. □

5.4. Qualitative properties of $\rho_{w,t}$ and $m_{w,t}$

The following lemma describes the square root behavior of $\rho_{w,t}$ around the edge $\lambda_{+,t}$.

Lemma 17. *For $|E - \lambda_{+,t}| \leqslant 3c_V/4$, the asymptotic density $\rho_{w,t}$ satisfies*

$$\rho_{w,t}(E) \sim \sqrt{(\lambda_{+,t} - E)_+}. \tag{115}$$

Moreover, if $-\tau t^2 \leqslant E - \lambda_{+,t} \leqslant 0$ for a sufficiently small constant $\tau > 0$, we have

$$\rho_{w,t}(E) = \frac{1}{\pi}\sqrt{\frac{2(\lambda_{+,t} - E)}{[4\lambda_{+,t}\xi_+(t) + (1-c_n)^2 t^2]c_n^2 t^2 \Phi''(\xi_+(t))}}\left[1 + \mathrm{O}\left(\frac{|E-\lambda_{+,t}|}{t^2}\right)\right]. \tag{116}$$

Recall that by (79), we have $t^2\Phi''(\xi_+(t)) \sim 1$.

Proof. By (109), we have

$$\beta(E) = \operatorname{Im}\zeta(E) \sim t \operatorname{Im} m_{w,t}(E) = t\pi\rho_{w,t}(E). \tag{117}$$

Then, (115) follows from (72) and (83). For (116), we use (55) and $b(z) = 1 + c_n t m_{w,t}(z)$ to get that

$$m_{w,t}(E) = \frac{t(1-c_n) + \sqrt{t^2(1-c_n)^2 + 4(\alpha(E) + \mathrm{i}\beta(E) + \lambda_+)E}}{2Ec_n t} - \frac{1}{c_n t}. \tag{118}$$

Taking the imaginary part of (118) and using (80), we can conclude (116). □

Lemma 17 immediately implies the following estimates on $\operatorname{Im} m_{w,t}$.

Lemma 18. *We have the following estimates for $z = E + \mathrm{i}\eta$ with $\lambda_{+,t} - 3c_V/4 \leqslant E \leqslant \lambda_{+,t} + 3c_V/4$ and $0 \leqslant \eta \leqslant 10$:*

$$|m_{w,t}(z)| \lesssim 1, \quad \operatorname{Im} m_{w,t}(z) \sim \begin{cases} \sqrt{\kappa+\eta}, & \lambda_{+,t} - 3c_V/4 \leqslant E \leqslant \lambda_{+,t} \\ \frac{\eta}{\sqrt{\kappa+\eta}}, & \lambda_{+,t} \leqslant E \leqslant \lambda_{+,t} + 3c_V/4 \end{cases}. \tag{119}$$

Proof. (119) can be derived easily from (9) combined with the square root behavior of $\rho_{w,t}$ in (115). □

We also need to control the derivative $\partial_z m_{w,t}(z)$ in our proof. First, by (9), we have the trivial estimate

$$\left|\partial_z m_{w,t}(z)\right| = \left|\int \frac{\rho_{w,t}(x)\mathrm{d}x}{(x-z)^2}\right| \leqslant \int \frac{\rho_{w,t}(x)\mathrm{d}x}{|x-z|^2} = \frac{\operatorname{Im} m_{w,t}}{\eta}. \tag{120}$$

Moreover, we claim the following estimates.

Lemma 19. *For $\kappa+\eta \leqslant t^2$, we have*

$$\left|\partial_z m_{w,t}(z)\right| \lesssim (\kappa+\eta)^{-1/2}. \tag{121}$$

Moreover, if $\kappa+\eta \geqslant t^2$, we have that for $E \geqslant \lambda_{+,t}$,

$$|\partial_z m_{w,t}(z)| \lesssim (\kappa+\eta)^{-1/2}, \tag{122}$$

and for $E \leqslant \lambda_{+,t}$,

$$|\partial_z m_{w,t}(z)| \lesssim \frac{\sqrt{\kappa+\eta}}{t\sqrt{\kappa+\eta}+\eta}. \tag{123}$$

Proof. By Eq. (53), we have $\partial_z \zeta = [\Phi_t'(\zeta)]^{-1}$. Then, using the definition of ζ in (49), we can solve that

$$\partial_z m_{w,t}(z) = \frac{[\Phi_t'(\zeta)]^{-1} - b^2}{[2bz-(1-c_n)t]\, c_n t}.$$

Using (101), we get that

$$\left|\partial_z m_{w,t}(z)\right| \lesssim \max\left\{\frac{1}{\sqrt{\kappa+\eta}}, \frac{1}{t}\right\}, \tag{124}$$

which concludes (121) for $\kappa+\eta \leqslant t^2$. The bound (122) follows directly from (119) and (120). Similarly, with (119) and (120), we can get (123) when $t\sqrt{\kappa+\eta} \leqslant \eta$. If $t\sqrt{\kappa+\eta} \geqslant \eta$, we use (124) to get

$$\left|\partial_z m_{w,t}(z)\right| \lesssim \max\left\{\frac{1}{\sqrt{\kappa+\eta}}, \frac{1}{t}\right\} \lesssim \frac{\sqrt{\kappa+\eta}}{t\sqrt{\kappa+\eta}+\eta}.$$

This concludes (123). □

Acknowledgments

The authors would like to thank the editor, the associated editor and two anonymous references for their insightful comments and suggestions, which have improved the paper significantly. The first author is partially supported by NSF, USA DMS-2113489. The second author is partially supported by the Wharton Dean's Fund for Postdoctoral Research, USA.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2022.105051.

References

[1] J. Alt, Singularities of the density of states of random Gram matrices, Electron. Commun. Probab. 22 (2017) 1–13.
[2] J. Alt, L. Erdős, T. Krüger, Local law for random Gram matrices, Electron. J. Probab. 22 (2017) 1–41.
[3] N. Asendorf, R.R. Nadakuditi, Improved detection of correlated signals in low-rank-plus-noise type data sets using informative canonical correlation analysis (ICCA), IEEE Trans. Inform. Theory 63 (6) (2017) 3451–3467.
[4] Z. Bao, X. Ding, K. Wang, Singular vector and singular subspace distribution for the matrix denoising model, Ann. Statist. 49 (1) (2021) 370–392.
[5] F. Benaych-Georges, R.R. Nadakuditi, The singular values and vectors of low rank perturbations of large rectangular random matrices, J. Multivariate Anal. 111 (2012) 120–135.
[6] E. Björnson, J. Hoydis, M. Kountouris, M. Debbah, Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits, IEEE Trans. Inf. Theory 60 (11) (2014) 7112–7139.
[7] A. Bloemendal, A. Knowles, H.-T. Yau, J. Yin, On the principal components of sample covariance matrices, Probab. Theory Related Fields 164 (1) (2016) 459–552.
[8] M. Capitaine, Limiting eigenvectors of outliers for spiked information-plus-noise type matrices, in: Séminaire de Probabilités XLIX, in: Lecture Notes in Math., vol. 2215, Springer, Cham, 2018, pp. 119–164.
[9] M. Capitaine, C. Donati-Martin, Spectrum of deformed random matrices and free probability, 2016, arXiv:1607.05560.
[10] X. Ding, High dimensional deformed rectangular matrices with applications in matrix denoising, Bernoulli 26 (1) (2020) 387–417.
[11] X. Ding, F. Yang, A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices, Ann. Appl. Probab. 28 (3) (2018) 1679–1738.
[12] X. Ding, F. Yang, Tracy-widom distribution for heterogeneous gram matrices with applications in signal detection, 2020, arXiv:2008.04166.
[13] X. Ding, F. Yang, Spiked separable covariance matrices and principal components, Ann. Statist. 49 (2) (2021) 1113–1138.
[14] X. Ding, F. Yang, Supplement to "Edge statistics of large dimensional deformed rectangular matrices", 2022.
[15] D. Donoho, M. Gavish, Minimax risk of matrix denoising by singular value thresholding, Ann. Statist. 42 (6) (2014) 2413–2440.
[16] D.L. Donoho, M. Gavish, a. Montanari, The phase transition of matrix recovery from Gaussian measurements matches the minimax MSE of matrix denoising, Proc. Natl. Acad. Sci. 110 (21) (2013) 8405–8410.
[17] R.B. Dozier, J.W. Silverstein, Analysis of the limiting spectral distribution of large dimensional information-plus-noise type matrices, J. Multivariate Anal. 98 (6) (2007a) 1099–1122.

[18] R.B. Dozier, J.W. Silverstein, On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices, J. Multivariate Anal. 98 (4) (2007b) 678–694.
[19] N. El Karoui, Tracy–widom limit for the largest eigenvalue of a large class of complex sample covariance matrices, Ann. Probab. 35 (2) (2007) 663–714.
[20] N. El Karoui, Spectrum estimation for large dimensional covariance matrices using random matrix theory, Ann. Statist. (2008) 2757–2790.
[21] L. Erdős, A. Knowles, H.-T. Yau, Averaging fluctuations in resolvents of random band matrices, Ann. Henri PoincarÉ 14 (2013a) 1837–1926.
[22] L. Erdős, A. Knowles, H.-T. Yau, J. Yin, Spectral statistics of erdős-rényi graphs II: Eigenvalue spacing and the extreme eigenvalues, Comm. Math. Phys. 314 (2012a) 587–640.
[23] L. Erdős, A. Knowles, H.-T. Yau, J. Yin, Spectral statistics of erdős-rényi graphs I: Local semicircle law, Ann. Probab. 41 (3B) (2013b) 2279–2375.
[24] L. Erdős, H.-T. Yau, A dynamical approach to random matrix theory, Courant Lect. Notes Math. 28 (2017).
[25] L. Erdős, H.-T. Yau, J. Yin, Rigidity of eigenvalues of generalized wigner matrices, Adv. Math. 229 (2012b) 1435–1515.
[26] J. Fan, K. Wang, Y. Zhong, Z. Zhu, Robust high dimensional factor models with applications to statistical machine learning, Statist. Sci. 36 (2) (2021) 303–327.
[27] X. Feng, X. He, Inference on low-rank data matrices with applications to microarray data, Ann. Appl. Stat. 3 (4) (2009) 1634–1654.
[28] M. Gavish, D.L. Donoho, Optimal shrinkage of singular values, IEEE Trans. Inform. Theory 63 (4) (2017) 2137–2152.
[29] J.Y. Hwang, J.O. Lee, K. Schnelli, Local law and Tracy–Widom limit for sparse sample covariance matrices, Ann. Appl. Probab. 29 (5) (2019) 3006–3036.
[30] J. Josse, S. Wager, Bootstrap-based regularization for low-rank matrix estimation, J. Mach. Learn. Res. 17 (124) (2016) 1–29.
[31] A. Knowles, J. Yin, The isotropic semicircle law and deformation of Wigner matrices, Comm. Pure Appl. Math. 66 (2013) 1663–1749.
[32] A. Knowles, J. Yin, The outliers of a deformed Wigner matrix, Ann. Probab. 42 (5) (2014) 1980–2031.
[33] A. Knowles, J. Yin, Anisotropic local laws for random matrices, Probab. Theory Related Fields (2016) 1–96.
[34] B. Landon, H.-T. Yau, Edge statistics of dyson Brownian motion, 2017, arXiv:1712.03881.
[35] J.O. Lee, K. Schnelli, Edge universality for deformed Wigner matrices, Rev. Math. Phys. 27 (08) (2015) 1550018.
[36] J.O. Lee, J. Yin, A necessary and sufficient condition for edge universality of wigner matrices, Duke Math. J. 163 (2014) 117–173.
[37] M. Lettau, M. Pelger, Estimating latent asset-pricing factors, J. Econometrics 218 (1) (2020) 1–31.
[38] P. Loubaton, P. Vallet, Almost sure localization of the eigenvalues in a Gaussian information plus noise model–application to the spiked models, Electron. J. Probab. 16 (2011) 1934–1959.
[39] V.V. Mani, R. Bose, Direction of arrival estimation and beamforming of multiple coherent UWB signals, in: 2010 IEEE International Conference on Communications, 2010, pp. 1–5.
[40] R.R. Nadakuditi, OptShrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage, IEEE Trans. Inform. Theory 60 (5) (2014) 3002–3018.
[41] R.R. Nadakuditi, J.W. Silverstein, 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. Sign. Proces. 4 (3) (2010) 468–480.
[42] Q. Nadeem, A. Kammoun, M. Debbah, M. Alouini, Design of 5G full dimension massive MIMO systems, IEEE Trans. Commun. 66 (2) (2018) 726–740.
[43] B. Nadler, Finite sample approximation results for principal component analysis: A matrix perturbation approach, Ann. Statist. 36 (6) (2008) 2791–2817.
[44] M. Nejati, S. Samavi, H. Derksen, K. Najarian, Denoising by low-rank and sparse representations, J. Vis. Commun. Image Represent. 36 (2016) 28–39.
[45] A. Onatski, Testing hypotheses about the number of factors in large factor models, Econometrica 77 (5) (2009) 1447–1479.
[46] N.S. Pillai, J. Yin, Universality of covariance matrices, Ann. Appl. Probab. 24 (3) (2014) 935–1001.
[47] E. Richard, P.-a. Savalle, N. Vayatis, Estimation of simultaneously sparse and low rank matrices, in: Proceedings of the 29th International Coference on International Conference on Machine Learning, in: ICML'12, 2012, pp. 51–58.
[48] L. Sun, B. Jeon, A novel subspace spatial-spectral low rank learning method for hyperspectral denoising, in: 2017 IEEE Visual Communications and Image Processing, VCIP, 2017, pp. 1–4.
[49] D.W. Tufts, A.A. Shah, Estimation of a signal waveform from noisy data using low-rank approximation to a data matrix, IEEE Trans. Signal Process. 41 (4) (1993) 1716–1721.
[50] A.M. Tulino, S. Verdú, et al., Random matrix theory and wireless communications, Found. Trends Commun. Inf. Theory 1 (1) (2004) 1–182.
[51] P. Vallet, P. Loubaton, X. Mestre, Improved subspace estimation for multivariate observations of high dimension: The deterministic signals case, IEEE Trans. Inform. Theory 58 (2) (2012) 1043–1068.
[52] D. Verenzuela, E. Bjoernson, L. Sanguinetti, Spectral efficiency of superimposed pilots in uplink massive MIMO systems, in: GLOBECOM 2017 - 2017 IEEE Global Communications Conference, 2017, pp. 1–6.
[53] J. Vinogradova, E. Björnson, E.G. Larsson, On the separability of signal and interference-plus-noise subspaces in blind pilot decontamination, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2016, pp. 3421–3425.
[54] F. Yang, Edge universality of separable covariance matrices, Electron. J. Probab. 24 (2019) 1–57.
[55] D. Yang, Z. Ma, a. Buja, Rate optimal denoising of simultaneously sparse and low rank matrices, J. Mach. Learn. Res. 17 (1) (2016) 3163–3189.
[56] G. Ye, M. Tang, J.-F. Cai, Q. Nie, X. Xie, Low-rank regularization for learning gene expression programs, PLoS One 8 (12) (2013) 1–9.
[57] A.W. Yu, W. Ma, Y. Yu, J. Carbonell, S. Sra, Efficient structured matrix rank minimization, in: Advances in Neural Information Processing Systems, vol. 27, 2014, pp. 1350–1358.
[58] Z. Zhang, W. Liu, W. Leng, A. Wang, H. Shi, Interference-plus-noise covariance matrix reconstruction via spatial power spectrum sampling for robust adaptive beamforming, IEEE Signal Process. Lett. 23 (1) (2016) 121–125.
[59] Z. Zhang, G. Pan, Tracy-widom law for the extreme eigenvalues of large signal-plus-noise matrices, 2020, arXiv:2009.12031.