Supplementary Material for "High-dimensional Spectrum Matching: Covariance Inference Based on Universality"

1 Preliminaries on random matrix theory

1.1 Notations

In this subsection, we introduce the notations necessary for proving the main theorems. We assume $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n \in \mathbb{R}^p$ have zero means and use the notation $\hat{\boldsymbol{\Sigma}} = \boldsymbol{X}^T\boldsymbol{X}/n$ for the sample covariance matrix. Recall the ratio $\phi = p/n$ and we assume that $p \asymp n^\gamma$ for some $\gamma \in (0, +\infty)$. Here we use $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$.

Denote $\boldsymbol{X}_i = \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_i$ where $\boldsymbol{Z}_i$ has identity covariance matrix for $i = 1, \cdots, n$. Consider the matrix

$$\boldsymbol{M}_n = \hat{\boldsymbol{\Sigma}} + \boldsymbol{R} = \sqrt{\frac{p}{n}}\boldsymbol{M}_n' \tag{1}$$

where

$$\boldsymbol{M}_n' = \frac{1}{\sqrt{np}}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{\Sigma}^{\frac{1}{2}} + \boldsymbol{R}', \quad \boldsymbol{R}' = \sqrt{\frac{n}{p}}\boldsymbol{R}. \tag{2}$$

$\boldsymbol{R}$ is a symmetric matrix and we allow $\boldsymbol{R}$ to have positive, negative, and zero eigenvalues. As demonstrated in the main text, the universality holds for $\boldsymbol{M}_n$ if and only if for $\boldsymbol{M}_n + a\boldsymbol{I}_p$ with some $a \in \mathbb{R}$. This means we can prove the results by replacing $\boldsymbol{R}$ with $\boldsymbol{R} + a\boldsymbol{I}_p$ with a large enough $a > 0$. We thus assume without loss of generality that $\boldsymbol{R}$ is positive definite.

Define $\boldsymbol{R}'' = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{R}'\boldsymbol{\Sigma}^{-1/2}$. Here, we first assume

$$\text{"}\boldsymbol{\Sigma} \text{ is invertible and } \boldsymbol{R} \text{ is positive-definite"}. \tag{3}$$

We will show how to remove (3) in Remark 2.3. Define $\boldsymbol{N} = \begin{bmatrix} (\boldsymbol{R}'')^{1/2} \\ \boldsymbol{Z}/(np)^{1/4} \end{bmatrix} \in \mathbb{R}^{(n+p)\times p}$, $\boldsymbol{Y} = \boldsymbol{N}\boldsymbol{\Sigma}^{1/2} = \begin{bmatrix} (\boldsymbol{R}'')^{1/2}\boldsymbol{\Sigma}^{1/2} \\ \boldsymbol{Z}\boldsymbol{\Sigma}^{1/2}/(np)^{1/4} \end{bmatrix} \in \mathbb{R}^{(n+p)\times p}$, we have

$$\boldsymbol{M}_n' = \boldsymbol{Y}^T\boldsymbol{Y}.$$

Define the Green functions

$$\boldsymbol{Q}_p(z) = \boldsymbol{H}_p^{-1}(z) = (\boldsymbol{Y}^T\boldsymbol{Y} - z\boldsymbol{I}_p)^{-1},\ \boldsymbol{Q}_{n+p}(z) = \boldsymbol{H}_{n+p}^{-1}(z) = (\boldsymbol{Y}\boldsymbol{Y}^T - z\boldsymbol{I}_{n+p})^{-1},$$

where

$$\boldsymbol{H}_p(z) = \boldsymbol{Y}^T\boldsymbol{Y} - z\boldsymbol{I}_p,\ \boldsymbol{H}_{n+p}(z) = \boldsymbol{Y}\boldsymbol{Y}^T - z\boldsymbol{I}_{n+p}.$$

We also define the Stieltjes transforms as

$$m_p(z) = \frac{1}{p}\mathrm{tr}\boldsymbol{Q}_p(z),\ m_{n+p}(z) = \frac{1}{n+p}\mathrm{tr}\boldsymbol{Q}_{n+p}(z).$$

Similar to Knowles & Yin (2017), we define the linearized matrix

$$\boldsymbol{H}(z) = \begin{bmatrix} -\boldsymbol{\Sigma}^{-1} & \boldsymbol{N}^T \\ \boldsymbol{N} & -z\boldsymbol{I}_{n+p} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{\Sigma}^{-1} & (\boldsymbol{R}'')^{\frac{1}{2}} & \frac{1}{(np)^{\frac{1}{4}}}\boldsymbol{Z}^T \\ (\boldsymbol{R}'')^{\frac{1}{2}} & -z\boldsymbol{I}_p & \boldsymbol{0} \\ \frac{1}{(np)^{\frac{1}{4}}}\boldsymbol{Z} & \boldsymbol{0} & -z\boldsymbol{I}_n \end{bmatrix},\ \boldsymbol{G}(z) = \boldsymbol{H}^{-1}(z). \tag{4}$$

As the linearized Green function $\boldsymbol{G}(z)$ has size $(2p+n)\times(2p+n)$, we introduce the following index sets.

Definition 1. We define the sets

$$\mathcal{I}_p = \{1, \cdots, p\},\ \mathcal{I}_{n+p} = \{p+1, \cdots, 2p+n\},\ \mathcal{I} = \mathcal{I}_p \cup \mathcal{I}_{n+p}.$$

To further distinguish different elements in $\mathcal{I}_{n+p}$, we also define

$$\mathcal{I}_{n+p,p} = \{p+1, \cdots, 2p\}, \ \mathcal{I}_{n+p,n} = \{2p+1, \cdots, 2p+n\}.$$

This partition coincides with the block size of $\boldsymbol{G}(z)$ in (4). We consistently use the letters $i, j \in \mathcal{I}_p$, $\mu, \nu \in \mathcal{I}_{n+p}$ and $s, t \in \mathcal{I}$.

By Definition 1, we can label the indices of $\boldsymbol{Z}$, $\boldsymbol{N}$ and $\boldsymbol{\Sigma}$ by

$$\boldsymbol{Z} = \{Z_{\mu i}, \mu \in \mathcal{I}_{n+p,n}, i \in \mathcal{I}_p\}, \ \boldsymbol{N} = \{N_{\mu i}, \mu \in \mathcal{I}_{n+p}, i \in \mathcal{I}_p\}, \ \boldsymbol{\Sigma} = \{\sigma_{ij}, i, j \in \mathcal{I}_p\}.$$

We will always take the following definition of generalized matrix multiplication, which simplifies our notations dramatically.

Definition 2. (Generalized matrix multiplication) For any matrices $\boldsymbol{A} = (a_{ij}, i \in \mathcal{I}_1, j \in \mathcal{I}_2)$, and $\boldsymbol{B} = (b_{jk}, j \in \mathcal{I}_3, k \in \mathcal{I}_4)$, we define the generalized matrix multiplication $\boldsymbol{AB}$ by

$$(\boldsymbol{AB})_{ik} = \sum_{j \in \mathcal{I}_2 \cap \mathcal{I}_3} a_{ij} b_{jk}, \ i \in \mathcal{I}_1, k \in \mathcal{I}_4.$$

By this definition, we do not require the size of $\boldsymbol{A}$ and $\boldsymbol{B}$ when defining $\boldsymbol{AB}$.

We now define the stochastic domination symbol $\prec$ as follows, which was introduced by Erdős et al. (2013) and widely applied in random matrix theory (Knowles & Yin 2017).

Definition 3. For two families of non-negative random variables $A = \left(A^{(n)}(u) \ : \ n \in \mathbb{N}, u \in U^{(n)}\right)$ and $B = \left(B^{(n)}(u) \ : \ n \in \mathbb{N}, u \in U^{(n)}\right)$, where $U^{(n)}$ is n-dependent parameter set, we say A is stochastically dominated by B uniformly in u if for any $\epsilon > 0$ and $D > 0$, we have

$$\sup_{u \in U^{(n)}} \mathbb{P}\left(A^{(n)}(u) > n^{\epsilon} B^{(n)}(u)\right) \leq n^{-D},$$

for large enough $n \geq n_0(\epsilon, D)$. We use the notation $A \prec B$ to represent this relationship. When A is a family of negative or complex random variables, we also write $A \prec B$ or $A = O_{\prec}(B)$ to indicate $|A| \prec B$.

1.2 Basic results

In this subsection, we establish some basic results for the Green function.

Lemma 1.1. *We have*

$$\boldsymbol{G}(z) = \begin{bmatrix} \boldsymbol{G}_p(z) & \frac{1}{z}\boldsymbol{G}_p(z)\boldsymbol{N}^T \\ \frac{1}{z}\boldsymbol{N}\boldsymbol{G}_p(z) & \frac{1}{z^2}\boldsymbol{N}\boldsymbol{G}_p(z)\boldsymbol{N}^T - \frac{1}{z} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}\boldsymbol{N}^T\boldsymbol{G}_{n+p}(z)\boldsymbol{N}\boldsymbol{\Sigma} - \boldsymbol{\Sigma} & \boldsymbol{\Sigma}\boldsymbol{N}^T\boldsymbol{G}_{n+p}(z) \\ \boldsymbol{G}_{n+p}(z)\boldsymbol{N}\boldsymbol{\Sigma} & \boldsymbol{G}_{n+p}(z) \end{bmatrix}, \tag{5}$$

where

$$\boldsymbol{G}_p(z) = z\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Q}_p(z)\boldsymbol{\Sigma}^{\frac{1}{2}},\ \boldsymbol{G}_{n+p}(z) = \boldsymbol{Q}_{n+p}(z). \tag{6}$$

Proof. The results follow from Schur's complement formula. $\square$

A direct result of Lemma 1.1 is that

$$\boldsymbol{Q}_{n+p}(z) = \frac{1}{z}\boldsymbol{Y}\boldsymbol{Q}_p(z)\boldsymbol{Y}^T - \frac{1}{z}\boldsymbol{I}_{n+p}. \tag{7}$$

To use the classical 'leave-one-out' method, we need to define the minors of the Greens function.

Definition 4. For any index set $\mathcal{S} \subset \mathcal{I}$, we define the minor matrix $\boldsymbol{H}^{(\mathcal{S})}(z) = \{H_{st}(z), s,t \in \mathcal{I}/\mathcal{S}\}$. We remark that $\boldsymbol{H}^{(\mathcal{S})}(z)$ is still an $(2p+n)\times(2p+n)$ matrix, with the undefined entries being 0. The minor Green function matrix is defined as $\boldsymbol{G}^{(\boldsymbol{\mathcal{S}})}(z) = \left(\boldsymbol{H}^{(\mathcal{S})}(z)\right)^{-1}$. Similarly definition holds for $\boldsymbol{G}_p^{(\boldsymbol{\mathcal{S}})}(z)$ and $\boldsymbol{G}_{n+p}^{(\boldsymbol{\mathcal{S}})}(z)$. For simplicity, we abbreviate $(\{s\}) = (s)$ and $(\{s,t\}) = (s,t)$.

We also define the augmented covariance matrix $\bar{\boldsymbol{\Sigma}} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{n+p} \end{bmatrix}$. We now state the resolvent identities.

Lemma 1.2. *(i) We have for $\mu \in \mathcal{I}_{n+p}$*

$$\frac{1}{G_{\mu\mu}(z)} = -z - \big(\boldsymbol{N}\boldsymbol{G}^{(\mu)}(z)\boldsymbol{N}^T\big)_{\mu\mu},$$

and for $\mu \neq \nu \in \mathcal{I}_{n+p}$,

$$G_{\mu\nu}(z) = -G_{\mu\mu}(z)\big(\boldsymbol{N}\boldsymbol{G}^{(\mu)}(z)\big)_{\mu\nu} = -G_{\nu\nu}(z)\big(\boldsymbol{G}^{(\nu)}(z)\boldsymbol{N}^T\big)_{\mu\nu} = -G_{\mu\mu}(z) - G_{\nu\nu}(z)\big(\boldsymbol{N}\boldsymbol{G}^{(\mu\nu)}(z)\boldsymbol{N}^T\big)_{\mu\nu}.$$

(ii) If $\boldsymbol{\Sigma}$ is diagonal, we also have for $i \in \mathcal{I}_p$,

$$\frac{1}{G_{ii}(z)} = -\frac{1}{\sigma_{ii}} - \big(\boldsymbol{N}^T\boldsymbol{G}^{(i)}(z)\boldsymbol{N}\big)_{ii},$$

and for $i \neq j \in \mathcal{I}_p$,

$$G_{ij}(z) = -G_{ii}(z)\big(\boldsymbol{N}^T\boldsymbol{G}^{(i)}(z)\big)_{ij} = -G_{jj}(z)\big(\boldsymbol{G}^{(j)}(z)\boldsymbol{N}\big)_{\mu\nu} = -G_{ii}(z) - G_{jj}(z)\big(\boldsymbol{N}^T\boldsymbol{G}^{(ij)}(z)\boldsymbol{N}\big)_{ij}.$$

(iii) For $i \in \mathcal{I}_p$ and $\mu \in \mathcal{I}_{n+p}$, we have

$$G_{i\mu}(z) = -G_{\mu\mu}(z)\big(\boldsymbol{G}^{(\mu)}(z)\boldsymbol{N}^T\big)_{i\mu},\ G_{\mu i}(z) = -G_{\mu\mu}(z)\big(\boldsymbol{N}\boldsymbol{G}^{(\mu)}(z)\big)_{\mu i}.$$

If in addition $\boldsymbol{\Sigma}$ is diagonal, we also have

$$G_{i\mu}(z) = -G_{ii}(z)\big(\boldsymbol{N}^T\boldsymbol{G}^{(i)}(z)\big)_{i\mu} = G_{ii}(z)G^{(i)}_{\mu\mu}(z)\Big(- N_{\mu i} + \big(\boldsymbol{N}^T\boldsymbol{G}^{(i\mu)}(z)\boldsymbol{N}^T\big)_{i\mu}\Big),$$

and

$$G_{\mu i}(z) = -G_{ii}(z)\big(\boldsymbol{G}^{(i)}(z)\boldsymbol{N}^T\big)_{\mu i} = G_{\mu\mu}(z)G^{(\mu)}_{ii}(z)\Big(- N_{\mu i} + \big(\boldsymbol{N}\boldsymbol{G}^{(\mu i)}(z)\boldsymbol{N}\big)_{\mu i}\Big).$$

(iv) For $r \in \mathcal{I}$ and $s, t \in \mathcal{I}/\{r\}$, we have

$$G^{(r)}_{st}(z) = G_{st}(z) - \frac{G_{sr}(z)G_{rt}(z)}{G_{rr}(z)}.$$

(v) All the above results hold if replacing $\boldsymbol{G}(z)$ by $\boldsymbol{G}^{(\mathcal{S})}(z)$ with some $\mathcal{S} \subset \mathcal{I}_{n+p}$ or $\mathcal{S} \subset \mathcal{I}$ and $\boldsymbol{\Sigma}$ is diagonal.

Proof. The proof is similar to Knowles & Yin (2017). $\square$

We define the generalized entries of a matrix as follows.

Definition 5. For vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^{\mathcal{I}}$, $t \in \mathcal{I}$ and an $\mathcal{I}\times\mathcal{I}$ matrix $\boldsymbol{A}$, we define its generalized entries as follows:

$$\boldsymbol{A}_{\boldsymbol{v}\boldsymbol{w}} = \boldsymbol{v}^T \boldsymbol{A}\boldsymbol{w},\ \ \boldsymbol{A}_{\boldsymbol{v}t} = \boldsymbol{v}^T \boldsymbol{A}\boldsymbol{e}_t,\ \ \boldsymbol{A}_{t\boldsymbol{w}} = \boldsymbol{e}_t^T \boldsymbol{A}\boldsymbol{w}.$$

Here $\boldsymbol{e}_t \in \mathbb{R}^{\mathcal{I}}$ is the natural basis vector with the t-th entry 1 and other entries 0.

We identify vectors $\boldsymbol{v} \in \mathbb{R}^p$ and $\boldsymbol{w} \in \mathbb{R}^{p+n}$ with their embeddings $\begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{0} \end{bmatrix}$ and $\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{w} \end{bmatrix}$ in $\mathbb{R}^{2p+n}$. With this, we can define quantities like $\boldsymbol{A}_{\boldsymbol{v}\boldsymbol{w}}$ when $\boldsymbol{v} \in \mathbb{R}^{\mathcal{I}_p}$ and $\boldsymbol{w} \in \mathbb{R}^{\mathcal{I}_{p+n}}$. We now represent the crucial estimation for proving anisotropic local law.

Lemma 1.3. *Under Assumptions 1, 2, 3, 4, and (3), for $\boldsymbol{v} \in \mathbb{R}^{\mathcal{I}_p}$, $\boldsymbol{w} \in \mathbb{R}^{\mathcal{I}_{p+n}}$, and $z = E + i\eta \in \mathcal{D}$ defined in (17) below, there exists constants C for the following estimation*

$$\sum_{\mu\in\mathcal{I}_{n+p}} \left|G_{\boldsymbol{w}\mu}(z)\right|^2 = \frac{\Im G_{\boldsymbol{w}\boldsymbol{w}}(z)}{\eta},$$

$$\sum_{i\in\mathcal{I}_p} \left|G_{\boldsymbol{v}i}(z)\right|^2 \leq \frac{C\left\|\boldsymbol{N}\boldsymbol{N}^T\right\|}{\eta} \Im G_{\boldsymbol{v}\boldsymbol{v}}(z) + 2\big(\boldsymbol{\Sigma}^2\big)_{\boldsymbol{v}\boldsymbol{v}},$$

$$\sum_{i\in\mathcal{I}_p} \left|G_{\boldsymbol{w}i}(z)\right|^2 \leq C\left\|\boldsymbol{N}\boldsymbol{N}^T\right\| \sum_{\mu\in\mathcal{I}_{n+p}} \left|G_{\boldsymbol{w}\mu}(z)\right|^2,$$

$$\sum_{\mu\in\mathcal{I}_{n+p}} \left|G_{\boldsymbol{v}\mu}(z)\right|^2 \leq C\left\|\boldsymbol{N}\boldsymbol{N}^T\right\| \sum_{i\in\mathcal{I}_p} \left|G_{\boldsymbol{v}i}(z)\right|^2.$$

Moreover, with high probability, we have

$$\left\|\boldsymbol{N}\boldsymbol{N}^T\right\| \leq C(\phi^{\frac{1}{2}} + \phi^{-\frac{1}{2}}). \tag{8}$$

Proof. See Knowles & Yin (2017) for $\gamma = 1$ and Ding & Wang (2025) for other cases. $\square$

In (8), by saying an event E happens with high probability, we mean $1 \prec 1_E$.

We now introduce the spectral decomposition of $\boldsymbol{G}(z)$. We define the singular value decomposition of $\boldsymbol{Y}^T$ as

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

where

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0 = \lambda_{p+1} = \cdots = \lambda_{p+n},$$

and $\{\boldsymbol{\xi}_k\}_{k=1}^{p}$ and $\{\boldsymbol{\zeta}_k\}_{k=1}^{n+p}$ are orthonormal bases of $\mathbb{R}^p$ and $\mathbb{R}^{n+p}$ respectively. Then for $\mu, \nu \in \mathcal{I}_{n+p}$ we have

$$G_{\mu\nu}(z) = \sum_{k=1}^{n+p} \frac{(\boldsymbol{\zeta}_k)_\mu (\boldsymbol{\zeta}_k)_\nu}{\lambda_k - z},$$

For $i, j \in \mathcal{I}_p$, we have

$$G_{ij}(z) = z \sum_{k=1}^{p} \frac{\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\xi}_k\right)_i \left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\xi}_k\right)_j}{\lambda_k - z} = -\Sigma_{ij} + \sum_{k=1}^{p} \frac{\lambda_k \left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\xi}_k\right)_i \left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\xi}_k\right)_j}{\lambda_k - z}.$$

Moreover, for $i \in \mathcal{I}_p$ and $\mu \in \mathcal{I}_{n+p}$ we have

$$G_{i\mu}(z) = G_{\mu i}(z) = \sum_{k=1}^{p} \frac{\sqrt{\lambda_k} \left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\xi}_k\right)_i (\boldsymbol{\zeta}_k)_\mu}{\lambda_k - z}.$$

Summarizing, defining

$$\mathbf{u}_k := \begin{pmatrix} 1_{k \leq p} \sqrt{\lambda_k} \boldsymbol{\xi}_k \\ 1_{k \leq n+p} \boldsymbol{\zeta}_k \end{pmatrix} \in \mathbb{R}^{2p+n},$$

we have

$$\boldsymbol{G}(z) = -\bar{\boldsymbol{\Sigma}} + \bar{\boldsymbol{\Sigma}}^{1/2} \sum_{k=1}^{n+p} \frac{\mathbf{u}_k \mathbf{u}_k^T}{\lambda_k - z} \bar{\boldsymbol{\Sigma}}^{1/2}. \tag{9}$$

We have the following lemma.

Lemma 1.4. *Under Assumptions 1, 2, 3, 4, and (3), $z = E + i\eta \in \mathcal{D}$ defined in (17) below, there exists a constant C such that*

$$\left\|\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{G}\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\right\| \le \frac{C(\phi^{\frac{1}{2}}+\phi^{-\frac{1}{2}})}{\eta}, \quad \left\|\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\partial_z\boldsymbol{G}\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\right\| \le \frac{C(\phi^{\frac{1}{2}}+\phi^{-\frac{1}{2}})}{\eta^2}.$$

Proof. The results are immediate using (9). Details are referred to Knowles & Yin (2017) for $\gamma = 1$ and Ding & Wang (2025) for other cases. □

1.3 Deterministic equivalence

In this subsection, we establish the basic properties of the deterministic equivalence of the Green function matrix $\boldsymbol{Q}(z)$ and the linearized Green function $\boldsymbol{G}(z)$. Inspired by Couillet et al. (2011), we define the deterministic equivalent Green function $\boldsymbol{Q}_p(z)$ as

$$\bar{\boldsymbol{Q}}_p(z) = \left(\boldsymbol{\Sigma}\boldsymbol{R}'' + \frac{1}{\phi^{\frac{1}{2}}(1+\phi^{\frac{1}{2}}e(z))}\boldsymbol{\Sigma} - z\boldsymbol{I}_p\right)^{-1}, \tag{10}$$

where $e(z)$ satisfies

$$e(z) = \frac{1}{p}\mathrm{tr}\left(\boldsymbol{\Sigma}\boldsymbol{R}'' + \frac{1}{\phi^{\frac{1}{2}}(1+\phi^{\frac{1}{2}}e(z))}\boldsymbol{\Sigma} - z\boldsymbol{I}_p\right)^{-1}\boldsymbol{\Sigma}, \tag{11}$$

and $\Im e(z) = \Im z$ when $\Im z \neq 0$ and $e(z) > 0$ when $z < 0$. We also define the corresponding deterministic equivalent Stieltjes transform $\bar{m}_p(z)$ as $\bar{m}_p(z) = \mathrm{tr}(\bar{\boldsymbol{Q}}_p(z))/p$. According to Couillet et al. (2011), $e(z)$ and $\bar{\boldsymbol{Q}}_p(z)$ are well-defined for every $z \in \mathbb{C}^+$, and $\bar{m}_p(z)$ serves as the Stieltjes transform of a measure ρ on $\mathbb{R}$. We further denote E_+ and E_- as the endpoints of ρ. Formally, if we define $\mathrm{supp}(\rho) = \{x \in \mathbb{R} \,:\, \rho([x-\epsilon, x+\epsilon]) > 0,\ \forall\epsilon > 0\}$, we have $E_+ = \sup\mathrm{supp}(\rho)$ and $E_- = \inf\mathrm{supp}(\rho)$. Combining results of Knowles & Yin (2017) for the $\gamma \le 1$ case and Ding & Wang (2025) for the $\gamma > 1$ case, it follows that $E_+ \asymp \phi^{1/2} + \phi^{-1/2}$ and $|E_-| \lesssim 1 + \phi^{-1/2}$.

However, the direct analysis for $\boldsymbol{Q}(z)$ is hard as $\boldsymbol{Q}^{-1}(z)$ is a quadratic function of $\boldsymbol{X}$. We then define quantities useful for the linearized Green function $\boldsymbol{G}(z)$. Recall the

Definition (4) of $\boldsymbol{G}(z)$ consisting of nine blocks. We correspondingly define its deterministic equivalence with nine blocks as follows,

$$\bar{\boldsymbol{G}}(z)=\begin{bmatrix}\bar{\boldsymbol{G}}_p(z) & * \\ * & \bar{\boldsymbol{G}}_{n+p}(z)\end{bmatrix}=\begin{bmatrix} z\boldsymbol{\Sigma}^{\frac{1}{2}}\bar{\boldsymbol{Q}}_p(z)\boldsymbol{\Sigma}^{\frac{1}{2}} & \boldsymbol{\Sigma}^{\frac{1}{2}}\bar{\boldsymbol{Q}}_p(z)\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{R}'')^{\frac{1}{2}} & \boldsymbol{0} \\ (\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}\bar{\boldsymbol{Q}}_p(z)\boldsymbol{\Sigma}^{\frac{1}{2}} & \frac{1}{z}\left((\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}\bar{\boldsymbol{Q}}_p(z)\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{R}'')^{\frac{1}{2}}-\boldsymbol{I}_p\right) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \tilde{m}(z)\boldsymbol{I}_n\end{bmatrix}, \tag{12}$$

where $\tilde{m}(z)=-z^{-1}(1+\phi^{1/2}e(z))^{-1}$.

Intuition: we give an intuitive explanation for the above definition of $\bar{\boldsymbol{G}}(z)$ and $\tilde{m}(z)$. This intuition serves as the guidance for our proof strategy later and the motivation for definitions of some intermediate quantities. The main technique we use is to repeatedly apply Schur's complement formula, which we call the "double Schur inversion" technique.

The standpoint is the result of the first application of Schur's complement formula, Lemma 1.1. To move on, recall the definition of $\boldsymbol{G}_{n+p}(z)$

$$\boldsymbol{G}_{n+p}(z)=(\boldsymbol{Y}\boldsymbol{Y}^T-z\boldsymbol{I}_{n+p})^{-1}=\begin{bmatrix}(\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma}(\boldsymbol{R}'')^{\frac{1}{2}}-z\boldsymbol{I}_p & \frac{1}{(np)^{\frac{1}{4}}}(\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{Z}^T \\ \frac{1}{(np)^{\frac{1}{4}}}\boldsymbol{Z}\boldsymbol{\Sigma}(\boldsymbol{R}'')^{\frac{1}{2}} & \frac{1}{\sqrt{np}}\boldsymbol{Z}\boldsymbol{\Sigma}\boldsymbol{Z}^T-z\boldsymbol{I}_n\end{bmatrix}^{-1}.$$

This allows us to apply Schur's complement formula at the second time to derive

$$\boldsymbol{G}_{n+p}(z)=\begin{bmatrix}\boldsymbol{A}(z) & \boldsymbol{B}(z) \\ \boldsymbol{B}^T(z) & \boldsymbol{G}_{n+p,n}(z)\end{bmatrix},$$

where $\boldsymbol{B}(z)=-((\boldsymbol{R}'')^{1/2}\boldsymbol{\Sigma}(\boldsymbol{R}'')^{1/2}-z\boldsymbol{I}_p)^{-1}(\boldsymbol{R}'')^{1/2}\boldsymbol{\Sigma}\boldsymbol{Z}^T\boldsymbol{G}_{n+p,n}(z)/(np)^{1/4}$. To give an expression of $\boldsymbol{G}_{n+p,n}(z)$, we use the identity $\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X}-z\boldsymbol{I})^{-1}\boldsymbol{X}^T=z(\boldsymbol{X}\boldsymbol{X}^T-z\boldsymbol{I})^{-1}+\boldsymbol{I}$ to obtain

$$\begin{aligned}\boldsymbol{G}_{n+p,n}(z)&=\left(\frac{1}{\sqrt{np}}\boldsymbol{Z}\boldsymbol{\Sigma}\boldsymbol{Z}^T-\frac{1}{\sqrt{np}}\boldsymbol{Z}\boldsymbol{\Sigma}(\boldsymbol{R}'')^{\frac{1}{2}}((\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma}(\boldsymbol{R}'')^{\frac{1}{2}}-z\boldsymbol{I}_p)^{-1}(\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{Z}^T-z\boldsymbol{I}_n\right)^{-1}\\&=\left(\frac{1}{\sqrt{np}}\boldsymbol{Z}(\boldsymbol{\Sigma}-\boldsymbol{\Sigma}(\boldsymbol{R}'')^{\frac{1}{2}}((\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma}(\boldsymbol{R}'')^{\frac{1}{2}}-z\boldsymbol{I}_p)^{-1}(\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma})\boldsymbol{Z}^T-z\boldsymbol{I}_n\right)^{-1}\\&=\left(\frac{1}{\sqrt{np}}\boldsymbol{Z}\boldsymbol{\Sigma}(z)\boldsymbol{Z}^T-z\boldsymbol{I}_n\right)^{-1}\end{aligned}$$

with

$$\boldsymbol{\Sigma}(z) = -z\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{R}''\boldsymbol{\Sigma}^{\frac{1}{2}} - z\boldsymbol{I}_p)^{-1},$$

defined in the main text. Similarly, we use the identity $\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X} - z\boldsymbol{I})^{-1} = (\boldsymbol{X}\boldsymbol{X}^T - z\boldsymbol{I})^{-1}\boldsymbol{X}$ to obtain the expression of $\boldsymbol{A}(z)$:

$$\begin{aligned}
\boldsymbol{A}(z) &= \frac{1}{\sqrt{np}}((\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma}(\boldsymbol{R}'')^{\frac{1}{2}} - z\boldsymbol{I}_p)^{-1}(\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{Z}^T\boldsymbol{G}_{n+p,n}(z)\boldsymbol{Z}\boldsymbol{\Sigma}(\boldsymbol{R}'')^{\frac{1}{2}}((\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma}(\boldsymbol{R}'')^{\frac{1}{2}} - z\boldsymbol{I}_p)^{-1} \\
&\quad + ((\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma}(\boldsymbol{R}'')^{\frac{1}{2}} - z\boldsymbol{I}_p)^{-1} \\
&= -(\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{R}''\boldsymbol{\Sigma}^{\frac{1}{2}} - z\boldsymbol{I}_p)^{-\frac{1}{2}}\boldsymbol{G}_{n+p,p}(z)(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{R}''\boldsymbol{\Sigma}^{\frac{1}{2}} - z\boldsymbol{I}_p)^{-\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{R}'')^{\frac{1}{2}} - \frac{1}{z}\boldsymbol{I}_p \\
&= \frac{1}{z}(\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}(z)\boldsymbol{G}_{n+p,p}(z)\boldsymbol{\Sigma}^{\frac{1}{2}}(z)(\boldsymbol{R}'')^{\frac{1}{2}} - \frac{1}{z}\boldsymbol{I}_p,
\end{aligned}$$

where we define

$$\boldsymbol{G}_{n+p,p}(z) = \left(\frac{1}{\sqrt{np}}\boldsymbol{\Sigma}^{\frac{1}{2}}(z)\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{\Sigma}^{\frac{1}{2}}(z) - z\boldsymbol{I}_p\right)^{-1}.$$

A simple calculation shows that

$$\boldsymbol{\Sigma}^{\frac{1}{2}}(z)\boldsymbol{G}_{n+p,p}(z)\boldsymbol{\Sigma}^{\frac{1}{2}}(z) = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Q}_p(z)\boldsymbol{\Sigma}^{\frac{1}{2}}. \tag{13}$$

From the above calculation, the z-dependent covariance matrix $\boldsymbol{\Sigma}(z)$ appears naturally. If we ignore the complex number z and view $\boldsymbol{\Sigma}(z)$ as a positive-definite matrix, the functions $\boldsymbol{G}_{n+p,n}(z)$ and $\boldsymbol{G}_{n+p,p}(z)$ are just the Green functions of the sample covariance matrix when the samples are generated from the population with covariance matrix $\boldsymbol{\Sigma}(z)$. In this view, we can intuitively expect the results from the classical deterministic equivalence that

$$\boldsymbol{G}_{n+p,n}(z) \approx \tilde{m}(z)\boldsymbol{I}_n,\ \boldsymbol{G}_{n+p,p}(z) \approx -(z(\boldsymbol{I}_p + \phi^{-\frac{1}{2}}\tilde{m}(z)\boldsymbol{\Sigma}(z)))^{-1}, \tag{14}$$

where $\tilde{m}(z)$ is the fixed point of the z-dependent deformed Marčenko-Pastur law,

$$\frac{1}{\tilde{m}(z)} = -z + \int \frac{\phi t}{\phi^{\frac{1}{2}} + \tilde{m}(z)t}\mathrm{d}\pi^z(t),$$

π^z is the empirical measure of the spectral of $\boldsymbol{\Sigma}(z)$. Here, we do not give the rigorous definition of the approximation symbol "$\approx$" in (14) and only pursue an intuition here for simplicity. A comparison of the definition of $e(z)$ (11) and some algebra transformation show that $\tilde{m}(z) = -z^{-1}(1 + \phi^{1/2}e(z))^{-1}$. This explains our definition of $\tilde{m}(z)$ in $\bar{\boldsymbol{G}}(z)$. Combining (13), (14) and the definition of $\bar{\boldsymbol{Q}}(z)$ (10) gives

$$\boldsymbol{\Sigma}^{\frac{1}{2}}(z)\boldsymbol{G}_{n+p,p}(z)\boldsymbol{\Sigma}^{\frac{1}{2}}(z) \approx \boldsymbol{\Sigma}^{\frac{1}{2}}\bar{\boldsymbol{Q}}_p(z)\boldsymbol{\Sigma}^{\frac{1}{2}}. \tag{15}$$

Take the above calculation in $\boldsymbol{G}_{n+p}(z)$, we have

$$\boldsymbol{G}_{n+p}(z) = \begin{bmatrix} \frac{1}{z}(\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}(z)\boldsymbol{G}_{n+p,p}(z)\boldsymbol{\Sigma}^{\frac{1}{2}}(z)(\boldsymbol{R}'')^{\frac{1}{2}} - \frac{1}{z}\boldsymbol{I}_p & \frac{1}{z}(\boldsymbol{R}'')^{\frac{1}{2}}\frac{1}{\sqrt{n}}\boldsymbol{\Sigma}(z)\boldsymbol{Z}^T\boldsymbol{G}_{n+p,n}(z) \\ \frac{1}{z}\boldsymbol{G}_{n+p,n}(z)\frac{1}{\sqrt{n}}\boldsymbol{Z}\boldsymbol{\Sigma}(z)(\boldsymbol{R}'')^{\frac{1}{2}} & \boldsymbol{G}_{n+p,n}(z) \end{bmatrix}.$$

Put pieces together, we use the above expression for $\boldsymbol{G}_{n+p}(z)$ in Lemma 1.1 to derive $\boldsymbol{G}(z)$ is equal to

$$\begin{bmatrix} z\boldsymbol{\Sigma}^{\frac{1}{2}}(z)\boldsymbol{G}_{n+p,p}(z)\boldsymbol{\Sigma}^{\frac{1}{2}}(z) & \boldsymbol{\Sigma}^{\frac{1}{2}}(z)\boldsymbol{G}_{n+p,p}(z)\boldsymbol{\Sigma}^{\frac{1}{2}}(z)(\boldsymbol{R}'')^{\frac{1}{2}} & \frac{1}{(np)^{\frac{1}{4}}}\boldsymbol{\Sigma}(z)\boldsymbol{Z}^T\boldsymbol{G}_{n+p,n}(z) \\ (\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}(z)\boldsymbol{G}_{n+p,p}(z)\boldsymbol{\Sigma}^{\frac{1}{2}}(z) & \frac{1}{z}[(\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}(z)\boldsymbol{G}_{n+p,p}(z)\boldsymbol{\Sigma}^{\frac{1}{2}}(z)(\boldsymbol{R}'')^{\frac{1}{2}} - \boldsymbol{I}_p] & \frac{1}{z(np)^{\frac{1}{4}}}(\boldsymbol{R}'')^{\frac{1}{2}}\boldsymbol{\Sigma}(z)\boldsymbol{Z}^T\boldsymbol{G}_{n+p,n}(z) \\ \boldsymbol{G}_{n+p,n}(z)\frac{1}{(np)^{\frac{1}{4}}}\boldsymbol{Z}\boldsymbol{\Sigma}(z) & \frac{1}{z(np)^{\frac{1}{4}}}\boldsymbol{G}_{n+p,n}(z)\boldsymbol{Z}\boldsymbol{\Sigma}(z)(\boldsymbol{R}'')^{\frac{1}{2}} & \boldsymbol{G}_{n+p,n}(z) \end{bmatrix} \tag{16}$$

Once we replace the z- dependent Green functions $\boldsymbol{G}_{n+p,n}(z)$ and $\boldsymbol{G}_{n+p,p}(z)$ by its intuitive approximation in (14) and (15) in (16), we can recover our deterministic formulation $\bar{\boldsymbol{G}}(z)$ (12).

Double Schur inversion The above formal calculations are valid except for the intuitive approximations (14) and (15), which are the hardest parts to prove the anisotropic local law. Here we name the rest calculation and the process of the derivation of (16) the "double Schur inversion" technique, meaning we apply the Schur inversion formula twice to derive the expression of the Green function $\boldsymbol{G}(z)$. This technique inspires the definition of $\tilde{m}(z)$ and $\boldsymbol{\Sigma}(z)$, which will be crucial in proving the anisotropic local law. It also illustrates the key observations that the effect of adding $\boldsymbol{R}$ to the sample covariance matrix can be characterized as changing $\boldsymbol{\Sigma}$ into $\boldsymbol{\Sigma}(z)$ in the sample covariance matrix. We call the former viewpoint "the structural form", and the latter viewpoint "the reduced form". Both forms have limitations and meet challenges in proving the local law. For example, the structural form has a complicated structure because of $\boldsymbol{R}$, while the reduced form $\boldsymbol{\Sigma}(z)$ has complex eigenvalues, both invalidating the proof of the classical local law for the covariance matrix. We will change our viewpoint repeatedly from these two forms to overcome these challenges.

2 Anisotropic local law

In this subsection, we establish the anisotropic local law, which forms the foundation for our universality results. Specifically, we aim to describe the fluctuations of $\lambda_1(\boldsymbol{M}_n') - E_+$ and $\lambda_p(\boldsymbol{M}_n') - E_-$. This requires us to analyze the convergence of the Stieltjes transform near the endpoints E_+ and E_-. We define the local domains $\mathcal{D}_+$ and $\mathcal{D}_-$ as

$$\mathcal{D}_\pm = \mathcal{D}_\pm(\tau) = \left\{ z = E + \mathrm{i}\eta \in \mathbb{C}^+ \ : \ |z| \geq \tau, |E - E_\pm| \leq \tau^{-1}, n^{-1+\tau} \leq \eta \leq \tau^{-1} \right\} \tag{17}$$

for a fixed parameter $0 < \tau < 1$. For $\gamma \leq 1$, let $\mathcal{D} = \mathcal{D}_+ \cup \mathcal{D}_-$, and for $\gamma > 1$, define $\mathcal{D} = \mathcal{D}_+$. We will focus on the behavior of the Green function on $\mathcal{D}$. This definition ensures that for $\gamma \leq 1$, both the largest and smallest eigenvalues of $\boldsymbol{M}_n'$ are controlled, so that $\sigma_1(\boldsymbol{M}_n') = \max\{|\lambda_1(\boldsymbol{M}_n')|, |\lambda_p(\boldsymbol{M}_n')|\}$ can also be controlled. While for the case $\gamma > 1$, we always have $|E_+| \gg |E_-|$, thus only the largest eigenvalue requires attention. With these definitions, we state the anisotropic local law as follows.

Theorem 2.1 (Anisotropic local law)**.** *Define the control parameter as* $\Psi(z) = \sqrt{\frac{\Im \tilde{m}(z)}{n\eta} + \frac{1}{n\eta}}$.

Suppose that Assumptions 1,2,3,4 and (3) hold.

(i) We have

$$\boldsymbol{u}^T \boldsymbol{\Phi} \bar{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{G}(z) - \bar{\boldsymbol{G}}(z)) \bar{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Phi} \boldsymbol{v} = O_{\prec}(\|\boldsymbol{u}\| \|\boldsymbol{v}\| \Psi(z)), \tag{18}$$

uniformly in vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{n+2p}$ and $z \in \mathcal{D}$, where the weight matrix $\boldsymbol{\Phi}$ is defined in (25) and the augemented covariance matrix is $\bar{\boldsymbol{\Sigma}} = diag(\boldsymbol{\Sigma}, \boldsymbol{I}_{n+p})$.

(ii) Moreover, we have the average local law

$$m_p(z) - \bar{m}_p(z) = O_{\prec}((p\eta)^{-1}), \tag{19}$$

uniformly in $z \in \mathcal{D}$.

The anisotropic local law provides a delicate characterization of the discrepancy between the random Green function and its deterministic counterpart, yielding a precise bound to analyze the behavior of the extreme eigenvalues of $\boldsymbol{M}_n'$. When $\boldsymbol{R} = \boldsymbol{0}$, similar results have been established for $\gamma = 1$ in Knowles & Yin (2017) and for $\gamma \geq 1$ in Ding et al. (2024), Ding & Wang (2025). As demonstrated in Ding & Wang (2025), the results for general $\gamma \geq 1$ hold without requiring the dimension p and sample size n to grow proportionally, resulting in different convergence rates for the blocks of $\boldsymbol{G}(z)$. Ding & Wang (2025) separately provides convergence rates for each block and offers a coarse bound for the anisotropic local law

$$G_{i\mu}(z) = O_{\prec}(\phi^{-\frac{1}{4}} \Psi(z)), \ i = 1, \cdots, p, \mu = p+1, \cdots, 2p+n.$$

In this study, we enhance these findings by introducing the weight matrix $\boldsymbol{\Phi}$, allowing us to express the anisotropic local law (18) in a more compact form. This reformulation refines the convergence rate and simplifies our proof.

2.1 Entry-wise local law for diagonal covariance

In this section, we will prove an entry-wise local law under the diagonal $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \cdots, \sigma_p)$ and $\boldsymbol{R}'' = \operatorname{diag}(\sigma_1 r_1, \cdots, \sigma_p r_p)$ assumption, which serves as a first step for establishing the final anisotropic local law. By using the term "entry-wise local law", we aim to establish the bound of $G_{st}(z) - \bar{G}_{st}(z)$ for $s, t \in \mathcal{I}$.

We first introduce some notations. Recall the eigenvalues of $\boldsymbol{\Sigma}(z)$ as $\lambda_i(z) = z\sigma_i/(z - \sigma_i r_i)$, we define for $i \in \mathcal{I}_p$,

$$\tilde{m}_i(z) = \frac{-\lambda_i(z)}{1 + \phi^{-\frac{1}{2}}\tilde{m}(z)\lambda_i(z)}.$$

We then have

$$\frac{1}{\tilde{m}(z)} = -z - \frac{\phi^{\frac{1}{2}}}{p}\sum_{i\in\mathcal{I}_p}\tilde{m}_i(z), \;\; \frac{1}{\tilde{m}_i(z)} = -\frac{1}{\lambda_i(z)} - \phi^{-\frac{1}{2}}\tilde{m}(z).$$

By Assumption 4 and the definition of $\tilde{m}_i(z)$, we also have

$$|\lambda_i(z)| \asymp 1, \;\; |\tilde{m}_i(z)| \asymp \min\{1, \phi^{\frac{1}{2}}\},$$

and

$$\Im\tilde{m}(z) \geq c\eta.$$

We now extend the notation σ_i and r_i for $i \in \mathcal{I}_p$ to $\sigma_\mu = 1$ for $\mu \in \mathcal{I}_{n+p}$ and $r_\mu = r_{\mu-p}$ for $\mu \in \mathcal{I}_{n+p,p}$. Similarly, we extend the notation of $\tilde{m}_\mu(z)$ by letting $m_\mu(z) = (\tilde{m}_{\mu-p}(z)r_{\mu-p}/z - 1)/z$ if $\mu \in \mathcal{I}_{n+p,p}$ and $m_\mu(z) = \tilde{m}(z)$ if $\mu \in \mathcal{I}_{n+p,n}$. Here we use the convention that $\mu - p \in \mathcal{I}_p$ for $\mu \in \mathcal{I}_{n+p,p}$. We define the random control parameters

$$\Lambda_p(z) = \max_{s,t\in\mathcal{I}_p}\frac{|G_{st}(z) - \delta_{t,s}\tilde{m}_t(z)|}{\sigma_s\sigma_t}, \;\; \Lambda_{o,p}(z) = \max_{s\neq t\in\mathcal{I}_p}\frac{|G_{st}(z)|}{\sigma_s\sigma_t},$$

$$\Lambda_n(z) = \max_{s,t\in\mathcal{I}_{n+p,n}}\frac{|G_{st}(z) - \delta_{t,s}\tilde{m}_t(z)|}{\sigma_s\sigma_t}, \;\; \Lambda_{o,n}(z) = \max_{s\neq t\in\mathcal{I}_{n+p,n}}\frac{|G_{st}(z)|}{\sigma_s\sigma_t},$$

and

$$\Lambda_{o,n,p}(z) = \max_{s \in \mathcal{I}_p, t \in \mathcal{I}_{n+p,n}} \frac{|G_{st}(z)|}{\sigma_s \sigma_t}.$$

Here δ_{ts} is the delta function that equals 1 if $s = t$ and equals 0 otherwise. These parameters control the diagonal and off-diagonal entries of the four blocks of $\bar{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{G}(z) - \bar{\boldsymbol{G}}(z))\bar{\boldsymbol{\Sigma}}^{-1}$ among the nine blocks. The reason we only control these four blocks comes from the expression (16) so that the rest of the five blocks can be easily controlled using these four-block control parameters. We also define the averaged control bound,

$$\Theta_p(z) = \left| \frac{1}{p} \sum_{i \in \mathcal{I}_p} G_{ii}(z) - \tilde{m}_i(z) \right|,$$
$$\Theta_n(z) = \left| \frac{1}{n} \sum_{i \in \mathcal{I}_{n+p,n}} G_{ii}(z) - \tilde{m}_i(z) \right|.$$

To control the above parameters at the same time, we define the combined bound

$$\Lambda(z) = \min\{\phi^{\frac{1}{2}}, \phi^{\frac{3}{2}}\}\Lambda_p(z) + \Lambda_n(z) + \min\{\phi^{\frac{1}{4}}, \phi^{\frac{3}{4}}\}\Lambda_{o,n,p}(z),$$

$$\Lambda_o(z) = \min\{\phi^{\frac{1}{2}}, \phi^{\frac{3}{2}}\}\Lambda_{o,p}(z) + \Lambda_{o,n}(z) + \min\{\phi^{\frac{1}{4}}, \phi^{\frac{3}{4}}\}\Lambda_{o,n,p}(z),$$

$$\Theta(z) = \min\{\phi^{\frac{1}{2}}, \phi^{\frac{3}{2}}\}\Theta_p(z) + \Theta_n(z).$$

Notice we multiply different weights before each control parameter. When $\gamma \neq 1$, these weights are not comparable as each block also has an incomparable size. We use these weights to balance the effect of the block size of different control parameters.

2.1.1 Useful lemmas

To prove the entry-wise local law, we first prove some lemmas. The first one is the trivial bound.

Lemma 2.2. *We have for $z \in \mathcal{D}$,*

$$\Theta(z) \lesssim \Lambda(z).$$

For $s \in \mathcal{I}$, we introduce the conditional expectation

$$\mathbb{E}_s[\cdot] =: \mathbb{E}[\cdot|\boldsymbol{H}^s],$$

where $\boldsymbol{H}^s$ is defined in Definition 4. Using Lemma 1.2 and the above definition, we have the following identities.

Lemma 2.3. *For $z \in \mathcal{D}$, we have*

(i) for $i \in \mathcal{I}_p$,

$$\frac{1}{G_{ii}(z)} = -\frac{1}{\lambda_i(z)} - \frac{\phi^{-\frac{1}{2}}}{n} tr\boldsymbol{G}^{(i)}_{n+p,n}(z) - Z_i(z),\ Z_i(z) = (1-\mathbb{E}_i)\big(\boldsymbol{N}^T\boldsymbol{G}^{(i)}(z)\boldsymbol{N}\big)_{ii}.$$

(ii) for $\mu \in \mathcal{I}_{n+p,p}$,

$$\frac{1}{G_{\mu\mu}(z)} = -z - r_\mu\big(\boldsymbol{G}^{(\mu)}_p(z)\big)_{\mu\mu}.$$

(iii) for $\mu \in \mathcal{I}_{n+p,n}$,

$$\frac{1}{G_{\mu\mu}(z)} = -z - \frac{\phi^{-\frac{1}{2}}}{n} tr\boldsymbol{G}^{(\mu)}_p(z) - Z_\mu(z),\ Z_\mu(z) = (1-\mathbb{E}_\mu)\big(\boldsymbol{N}\boldsymbol{G}^{(\mu)}(z)\boldsymbol{N}^T\big)_{\mu\mu}.$$

Proof. For (i), we first calculate $\mathbb{E}_i[\big(\boldsymbol{N}^T\boldsymbol{G}^{(i)}(z)\boldsymbol{N}\big)_{ii}]$ for $i \in \mathcal{I}_p$. To achieve this, notice

$$\big(\boldsymbol{N}^T\boldsymbol{G}^{(i)}(z)\boldsymbol{N}\big)_{ii} = \big[\boldsymbol{R}^{\frac{1}{2}}_{i\cdot}, \frac{1}{(np)^{\frac{1}{4}}}\boldsymbol{Z}^T_{\cdot i}\big]\boldsymbol{G}^{(i)}_{n+p}(z)\begin{bmatrix} \boldsymbol{R}^{\frac{1}{2}}_{\cdot i} \\ \frac{1}{(np)^{\frac{1}{4}}}\boldsymbol{Z}_{\cdot i}\end{bmatrix},$$

where we adopt the convention that $\boldsymbol{Z}_{\cdot i}$ and $\boldsymbol{Z}_{i\cdot}$ represent the i-th column and the i-th row of a matrix $\boldsymbol{Z}$ respectively. It then follows from (16) and definition of $\boldsymbol{G}^{(i)}(z)$ that

$$\mathbb{E}_i[\big(\boldsymbol{N}^T\boldsymbol{G}^{(i)}(z)\boldsymbol{N}\big)_{ii}] = -\frac{r_i}{z} + \frac{\phi^{-\frac{1}{2}}}{n}\mathrm{tr}\boldsymbol{G}^{(i)}_{n+p,n}(z).$$

Using Lemma 1.2, we have

$$\begin{aligned}\frac{1}{G_{ii}(z)} &= -\frac{1}{\sigma_i} + \frac{r_i}{z} - \frac{\phi^{-\frac{1}{2}}}{n}\operatorname{tr}\boldsymbol{G}^{(i)}_{n+p,n}(z) - Z_i(z)\\ &= -\frac{1}{\lambda_i(z)} - \frac{\phi^{-\frac{1}{2}}}{n}\operatorname{tr}\boldsymbol{G}^{(i)}_{n+p,n}(z) - Z_i(z).\end{aligned}$$

i.e. (i) follows.

For (ii), we have for $\mu \in \mathcal{I}_{n+p,p}$,

$$\big(\boldsymbol{N}\boldsymbol{G}^{(\mu)}(z)\boldsymbol{N}^T\big)_{\mu\mu} = r_{\mu\mu}\big(\boldsymbol{G}^{(\mu)}_p(z)\big)_{\mu\mu}.$$

For (iii), we have for $\mu \in \mathcal{I}_{n+p,n}$,

$$\mathbb{E}_\mu[\big(\boldsymbol{N}\boldsymbol{G}^{(\mu)}(z)\boldsymbol{N}^T\big)_{\mu\mu}] = \mathbb{E}_\mu[\frac{1}{(np)^{\frac{1}{2}}}\boldsymbol{Z}_{\mu\cdot}\boldsymbol{G}^{(\mu)}_p(z)\boldsymbol{Z}^T_{\mu\cdot}] = \frac{\phi^{-\frac{1}{2}}}{n}\operatorname{tr}\boldsymbol{G}^{(\mu)}_p(z).$$

The results follow using Lemma 1.2 directly. □

We define the z-dependent event $\mathcal{E}(z) = \{\Lambda(z) \le \frac{1}{\ln n}\}$ and the control parameter

$$\Psi_{\Theta,n}(z) = \sqrt{\frac{\Im\tilde{m}(z)+\Theta(z)}{n\eta}},\ \Psi_{\Theta,p}(z) = \sqrt{\frac{\Im\tilde{m}(z)+\Theta(z)}{p\eta}},\ \Psi_{\Theta,n,p}(z) = \sqrt{\frac{\Im\tilde{m}(z)+\Theta(z)}{\sqrt{np}\eta}}.$$

We prove the following estimate.

Lemma 2.4. *Under Assumptions 1, 2, 3, 4, and (3), we have for $t \in \mathcal{I}_p \bigcup \mathcal{I}_{n+p,n}$ and $z \in \mathcal{D}$,*

$$1_{\mathcal{E}(z)}(|Z_t(z)| + \Lambda_o(z)) \prec \Psi_{\Theta,n}(z),$$

and

$$1_{\{\eta\ge 1\}}(z)(|Z_t(z)| + \Lambda_o(z)) \prec \Psi_{\Theta,n}(z).$$

Proof. We only provide proof of the first inequality. The second inequality follows with a similar proof. Notice that for $S \subset \mathcal{I}_p \bigcup \mathcal{I}_{n+p,n}$ and $t \in \mathcal{I}_p \bigcup \mathcal{I}_{n+p,n}/S$ satisfying $|S| \le C$, we have

$$1_{\mathcal{E}(z)}|G_{tt}^{(S)}| \asymp \sigma_{tt}.$$

We first estimate $\Lambda_o(z)$. For $i \neq j \in \mathcal{I}_p$, using Lemma 1.2 and the large deviation argument in Knowles & Yin (2017), we have

$$\begin{aligned}
&1_{\mathcal{E}(z)}\left|1-\frac{G_{ii}(z)G_{jj}^{(i)}(z)}{z^2}\right|\left|G_{ij}(z)\right| \\
\leq &1_{\mathcal{E}(z)}\left|G_{ii}(z)G_{jj}^{(i)}(z)\right|\left|\frac{1}{\sqrt{np}}\sum_{\mu,\nu\in\mathcal{I}_{n+p,n}} Z_{\mu,i}G_{\mu,\nu}^{(ij)}(z)Z_{\nu,j}\right| \\
\prec &1_{\mathcal{E}(z)}\sigma_{ii}\sigma_{jj}\left(\frac{1}{np}\sum_{\mu,\nu\in\mathcal{I}_{n+p,n}}\left|G_{\mu,\nu}^{(ij)}(z)\right|^2\right)^{\frac{1}{2}} \\
\prec &1_{\mathcal{E}(z)}\max\{1,\phi^{-1}\}\sigma_{ii}\sigma_{jj}\left(\frac{1}{np\eta}\sum_{\mu\in\mathcal{I}_{n+p,n}}\Im G_{\mu,\mu}^{(ij)}(z)\right)^{\frac{1}{2}} \\
\prec &1_{\mathcal{E}(z)}\max\{1,\phi^{-1}\}\sigma_{ii}\sigma_{jj}\left(\frac{1}{np\eta}\sum_{\mu\in\mathcal{I}_{n+p,n}}\Im G_{\mu,\mu}(z)+\frac{\Lambda_{0,n,p}^2(z)}{p\eta}\right)^{\frac{1}{2}} \\
\leq &\max\{1,\phi^{-1}\}\sigma_{ii}\sigma_{jj}\left(\frac{\Im\tilde{m}(z)+\Theta_n(z)+\Lambda_{o,n,p}^2(z)}{p\eta}\right)^{\frac{1}{2}}.
\end{aligned}$$

Similarly, for $i \in \mathcal{I}_p$, $\mu \in \mathcal{I}_{n+p,n}$, we have

$$\begin{aligned}
&1_{\mathcal{E}(z)}|G_{i\mu}(z)| \\
=&|G_{ii}(z)G_{\mu\mu}^{(i)}(z)||(\frac{(1-\frac{1}{z})}{(np)^{-\frac{1}{2}}}Z_{\mu i}+\frac{1}{\sqrt{np}}\sum_{j\in\mathcal{I}_p,\nu\in\mathcal{I}_{n+p,n}} Z_{\nu i}G_{\nu j}^{(i\mu)}(z)Z_{\mu j})| \\
\prec &1_{\mathcal{E}(z)}\sigma_{ii}\sigma_{\mu\mu}\left(|\frac{(1-\frac{1}{z})}{(np)^{\frac{1}{4}}}Z_{\mu i}|+(\frac{1}{np}\sum_{j\in\mathcal{I}_p,\nu\in\mathcal{I}_{n+p,n}}|G_{\nu,j}^{(i\mu)}(z)|^2)^{\frac{1}{2}}\right) \\
\prec &1_{\mathcal{E}(z)}\sigma_{ii}\sigma_{\mu\mu}\left(|\frac{(1-\frac{1}{z})}{(np)^{\frac{1}{4}}}Z_{\mu i}|+(\frac{\phi^{\frac{1}{2}}+\phi^{-\frac{1}{2}}}{np\eta}\sum_{\nu\in\mathcal{I}_{n+p,n}}\Im G_{\nu,\nu}^{(i\mu)}(z))^{\frac{1}{2}}\right) \\
\prec &1_{\mathcal{E}(z)}\sigma_{ii}\sigma_{\mu\mu}\left(|\frac{(1-\frac{1}{z})}{(np)^{\frac{1}{4}}}Z_{\mu i}|+(\frac{\phi^{\frac{1}{2}}+\phi^{-\frac{1}{2}}}{np\eta}\sum_{\nu\in\mathcal{I}_{n+p,n}}\Im G_{\nu,\nu}(z)+\frac{(\phi^{\frac{1}{2}}+\phi^{\frac{1}{2}})\Lambda_{o,n}^2(z)}{n\eta})^{\frac{1}{2}}\right) \\
\prec &\max\{1,\phi^{-\frac{1}{2}}\}\sigma_{ii}\sigma_{\mu\mu}\left(\frac{\phi^{\frac{1}{2}}\Im\tilde{m}(z)+\phi^{\frac{1}{2}}\Theta_n(z)+\phi^{\frac{1}{2}}\Lambda_{o,n}^2(z)}{n\eta}\right)^{\frac{1}{2}} \\
= &\max\{1,\phi^{-\frac{1}{2}}\}\sigma_{ii}\sigma_{\mu\mu}\left(\frac{\Im\tilde{m}(z)+\Theta_n(z)+\Lambda_{o,n}^2(z)}{\sqrt{np}\eta}\right)^{\frac{1}{2}}.
\end{aligned}$$

For $\mu \neq \nu \in \mathcal{I}_{n+p,n}$, we have

$$\begin{aligned}
&1_{\mathcal{E}(z)}|G_{\mu\nu}(z)| \\
=&1_{\mathcal{E}(z)}|G_{\mu\mu}(z)G^{(\mu)}_{\nu\nu}(z)|\Big|\frac{1}{\sqrt{np}}\sum_{i,j\in\mathcal{I}_p} Z_{\mu i}G^{(\mu\nu)}_{ij}(z)Z_{\nu j}\Big| \\
\prec&1_{\mathcal{E}(z)}\sigma_{\mu\mu}\sigma_{\nu\nu}\Bigg(|\frac{1}{np}\sum_{i,j\in\mathcal{I}_p}|G^{(\mu\nu)}_{ij}(z)|^2\Bigg)^{\frac{1}{2}} \\
\prec&\sigma_{\mu\mu}\sigma_{\nu\nu}\Bigg(\frac{\Im\tilde{m}(z)+\min\{\phi^{\frac{1}{2}},\phi^{\frac{3}{2}}\}\Theta_p(z)+\min\{\phi^{\frac{1}{2}},\phi^{\frac{3}{2}}\}\Lambda^2_{o,n,p}(z)}{n\eta}\Bigg)^{\frac{1}{2}}.
\end{aligned}$$

Combining above gives

$$\begin{aligned}
\Lambda_{o,n}(z) \prec&\Psi_{\Theta,n}(z)+\frac{1}{\sqrt{n\eta}}\Lambda_{o,n,p}(z), \\
\min\{\phi^{\frac{1}{2}},\phi^{\frac{3}{2}}\}\Lambda_{o,p}(z) \prec&\Psi_{\Theta,n}(z)+\frac{1}{\sqrt{n\eta}}\Lambda_{o,n,p}(z), \\
\min\{\phi^{\frac{1}{4}},\phi^{\frac{3}{4}}\}\Lambda_{o,n,p}(z) \prec&\Psi_{\Theta,n}(z)+\frac{1}{\sqrt{n\eta}}\Lambda_{o,n}(z),
\end{aligned}$$

which leads to

$$1_{\mathcal{E}(z)}\Lambda_o(z) \prec \Psi_{\Theta,n}(z).$$

We then estimate $|Z_t(z)|$. For $i \in \mathcal{I}_p$, we have

$$\begin{aligned}
&1_{\mathcal{E}(z)}|Z_i(z)| = 1_{\mathcal{E}(z)}|(1-\mathbb{E}_i)\big(\boldsymbol{N}^T\boldsymbol{G}^{(i)}(z)\boldsymbol{N}\big)_{ii}| \\
=&1_{\mathcal{E}(z)}\big|(1-\mathbb{E}_i)\frac{1}{\sqrt{np}}\sum_{\mu,\nu\in\mathcal{I}_{n+p,n}} Z_{\mu,i}G^{(i)}_{\mu,\nu}(z)Z_{\nu,i}\big| \\
\leq&1_{\mathcal{E}(z)}\big|\frac{1}{\sqrt{np}}\sum_{\mu\in\mathcal{I}_{n+p,n}}(Z^2_{\mu,i}-1)G^{(i)}_{\mu,\mu}(z)\big| + 1_{\mathcal{E}(z)}\big|\frac{1}{\sqrt{np}}\sum_{\mu\neq\nu\in\mathcal{I}_{n+p,n}} Z_{\mu,i}G^{(i)}_{\mu,\nu}(z)Z_{\nu,i}\big| \\
\prec&1_{\mathcal{E}(z)}\big(\frac{1}{np}\sum_{\mu,\nu\in\mathcal{I}_{n+p,n}}|G^{(i)}_{\mu,\nu}(z)|^2\big)^{\frac{1}{2}} \\
\leq&1_{\mathcal{E}(z)}\big(\frac{1}{np\eta}\sum_{\mu\in\mathcal{I}_{n+p,n}}\Im G^{(i)}_{\mu,\mu}(z)+\frac{\Lambda^2_o(z)}{p\eta}\big)^{\frac{1}{2}} \\
\leq&\Bigg(\frac{\Im\tilde{m}(z)+\Theta_n(z)+\Lambda^2_o(z)}{p\eta}\Bigg)^{\frac{1}{2}} \prec \Psi_{\Theta,n}(z).
\end{aligned}$$

For $\mu \in \mathcal{I}_{n+p,n}$, we have

$$
\begin{aligned}
&1_{\mathcal{E}(z)}|Z_\mu(z)| = |(1-\mathbb{E}_\mu)\big(\boldsymbol{N}\boldsymbol{G}^{(\mu)}(z)\boldsymbol{N}^T\big)_{\mu\mu}| \\
=&1_{\mathcal{E}(z)}\big|(1-\mathbb{E}_\mu)\frac{1}{\sqrt{np}}\sum_{i,j\in\mathcal{I}_p} Z_{\mu,i}G^{(\mu)}_{i,j}(z)Z_{\mu,j}\big| \\
\leq&1_{\mathcal{E}(z)}\big|\frac{1}{\sqrt{np}}\sum_{i\in\mathcal{I}_p}(Z^2_{\mu,i}-1)G^{(\mu)}_{i,i}(z)\big| + 1_{\mathcal{E}(z)}\big|\frac{1}{\sqrt{np}}\sum_{i\neq j\in\mathcal{I}_p} Z_{\mu,i}G^{(\mu)}_{i,j}(z)Z_{\mu,j}\big| \\
\prec&1_{\mathcal{E}(z)}\big(\frac{1}{np}\sum_{i,j\in\mathcal{I}_p}|G^{(\mu)}_{i,j}(z)|^2\big)^{\frac{1}{2}} \\
\leq&\left(\frac{\Im\tilde{m}(z)+\min\{\phi^{\frac{1}{2}},\phi^{\frac{3}{2}}\}\Theta_p(z)+\Lambda_o^2(z)}{n\eta}\right)^{\frac{1}{2}} \prec \Psi_\Theta(z).
\end{aligned}
$$

These two show that

$$1_{\mathcal{E}(z)}|Z_t(z)| \prec \Psi_{\Theta,n}(z).$$

This finishes the proof of the first inequality. □

To improve the rate of Lemma 2.4, we give a refined estimation for the average of $Z_i(z)$. Define $m_n(z) = \frac{1}{n}\mathrm{tr}\boldsymbol{G}_{n+p,n}(z)$. We have the following results.

Lemma 2.5. *Define the average*

$$[Z]_n(z) = \frac{1}{n}\sum_{\mu\in\mathcal{I}_{n+p,n}} Z_\mu(z),\ [Z]_p(z) = \frac{\phi^{\frac{1}{2}}}{p}\sum_{i\in\mathcal{I}_p}\frac{\lambda_i^2(z)}{(1+\phi^{-\frac{1}{2}}m_n(z)\lambda_i(z))^2}Z_i(z).$$

Under Assumptions 1, 2, 3, 4, and (3), we have for $z \in \mathcal{D}$,

$$1_{\mathcal{E}(z)}f(m_n(z),z) = 1_{\mathcal{E}(z)}([Z]_n(z)+[Z]_p(z)+O_\prec(\Psi^2_{\Theta,n}(z))),$$

and

$$1_{\eta\geq 1}f(m_n(z),z) = 1_{\eta\geq 1}\bigg([Z]_n(z)+[Z]_p(z)+O_\prec\big(\frac{1}{n}\big)\bigg),$$

where the function $f(m,z)$ is defined as

$$f(m,z) = -\frac{1}{m}+\frac{\phi^{\frac{1}{2}}}{p}\sum_{i=1}^{p}\frac{\lambda_i(z)}{1+\phi^{-\frac{1}{2}}m\lambda_i(z)} - z.$$

Proof. For simplicity, we only give the proof for the first one. The second one can be obtained in a similar way.

Using Lemma 2.3, we have for $i \in \mathcal{I}_p$,

$$1_{\mathcal{E}(z)} \frac{1}{G_{ii}(z)} = 1_{\mathcal{E}(z)} \Big(- \frac{1}{\lambda_i(z)} - \phi^{-\frac{1}{2}} m_n(z) - Z_i(z) + O_{\prec}(\phi^{-\frac{1}{2}} \Psi^2_{\Theta,n,p}(z)) \Big),$$

and for $\mu \in \mathcal{I}_{n+p,n}$,

$$1_{\mathcal{E}(z)} \frac{1}{G_{\mu\mu}(z)} = 1_{\mathcal{E}(z)} \Big(- z - \frac{\phi^{\frac{1}{2}}}{p} \sum_{i \in \mathcal{I}_{n+p,n}} G_{ii}(z) - Z_\mu(z) + O_{\prec}(\phi^{\frac{1}{2}} \Psi^2_{\Theta,n,p}(z)) \Big).$$

Moreover, Lemma 2.4 shows that for $\mu \in \mathcal{I}_{n+p,n}$,

$$1_{\mathcal{E}(z)} |G_{\mu\mu}(z) - m_n(z)| \prec \Psi_{\Theta,n}(z),$$

which gives

$$1_{\mathcal{E}(z)} \frac{1}{n} \sum_{\nu \in \mathcal{I}_{n+p,n}} \frac{1}{G_{\mu\mu}(z)} = 1_{\mathcal{E}(z)} \frac{1}{m_n(z)} + O_{\prec}(\Psi^2_{\Theta,n}(z)).$$

Combining the above and Assumption 4 shows

$$\begin{aligned} 1_{\mathcal{E}(z)} \frac{1}{m_n(z)} &= 1_{\mathcal{E}(z)} \Big(- z + \frac{\phi^{\frac{1}{2}}}{p} \sum_{i \in \mathcal{I}_p} \frac{-\lambda_i(z)}{1 + \phi^{-\frac{1}{2}} m_n(z) \lambda_i(z) + \lambda_i(z) Z_i(z) + O_{\prec}(\phi^{-\frac{1}{2}} \lambda_i(z) \Psi^2_{\Theta,n,p}(z))} \\ &\quad - [Z]_n(z) + O_{\prec}(\Psi^2_{\Theta,n}(z)) \Big) \\ &= 1_{\mathcal{E}(z)} \Big(- z + \frac{\phi^{\frac{1}{2}}}{p} \sum_{i \in \mathcal{I}_p} \frac{-\lambda_i(z)}{1 + \phi^{-\frac{1}{2}} m_n(z) \lambda_i(z)} - [Z]_n(z) - [Z]_p(z) + O_{\prec}(\Psi^2_{\Theta,n}(z)) \Big). \end{aligned}$$

This finishes the proof. $\square$

We now analyze the properties of $f(m, z)$ to utilize Lemma 2.5. We focus on the largest edge E_+, while the case for E_- for $\gamma \leq 1$ can be dealt in the same way.

Lemma 2.6. *Under Assumptions 1, 2, 3, 4, and (3), we have for $z \in \mathcal{D}$,*

$$|\tilde{m}(E_+)| \asymp \min\{1, \phi^{\frac{1}{2}}\}, \ \left| \frac{\partial^2}{\partial m^2} f(\tilde{m}(E_+), E_+) \right| \asymp \max\{1, \phi^{-2}\}.$$

Moreover, there exists τ' and C for all $|\zeta - \tilde{m}(E_+)| \leq \tau'$,

$$\left|\frac{\partial^3}{\partial m^3} f(\zeta, z)\right| \leq C \max\{1, \phi^{-3}\}.$$

Proof. By the definition of $\tilde{m}(z)$ and E_+, we have

$$f(\tilde{m}(E_+), E_+) = 0, \ \frac{\partial}{\partial m} f(\tilde{m}(E_+), E_+) = 0.$$

This combined with Assumption 2 gives $|\tilde{m}(E_+)| \asymp 1$. We also have

$$\frac{\partial}{\partial m} f(\tilde{m}, z) = \frac{1}{\tilde{m}^2} - \frac{1}{p}\sum_{i=1}^{p} \frac{1}{(\phi^{-\frac{1}{2}}\tilde{m} + \lambda_i^{-1}(z))^2},$$

$$\frac{\partial^2}{\partial m^2} f(\tilde{m}, z) = \frac{-2}{\tilde{m}^3} + \frac{2\phi^{-\frac{1}{2}}}{p}\sum_{i=1}^{p} \frac{1}{(\phi^{-\frac{1}{2}}\tilde{m} + \lambda_i^{-1}(z))^3},$$

$$\frac{\partial^3}{\partial m^3} f(\tilde{m}, z) = \frac{6}{\tilde{m}^4} - \frac{6}{p}\sum_{i=1}^{p} \frac{\phi^{-1}}{(\phi^{-\frac{1}{2}}\tilde{m} + \lambda_i^{-1}(z))^4}.$$

This gives that

$$\frac{\partial^2}{\partial m^2} f(\tilde{m}(E_+), E_+) = -\frac{2}{p}\sum_{i=1}^{p} \frac{1}{\tilde{m}(E_+)} \frac{\lambda_i^{-1}(E_+)}{(\phi^{-\frac{1}{2}}\tilde{m}(E_+) + \lambda_i^{-1}(z))^3},$$

the estimations are easily obtained from the expressions of the derivatives. □

Another useful lemma shows the behavior of $\tilde{m}(z)$, $\bar{m}_p(z)$, and thus $\rho(z)$, the density of $\bar{m}_p(z)$.

Lemma 2.7. *Under Assumptions 1, 2, 3, 4, and (3), we have for $z \in \mathcal{D}$, there exists τ' such that*

$$\tilde{m}(z) - \tilde{m}(E_+) \asymp \min\{1, \phi\}\sqrt{|z - E_+|}, \ \bar{m}_p(z) - \bar{m}_p(E_+) \asymp \sqrt{|z - E_+|},$$

and

$$\rho(z) \asymp \min \sqrt{|z - E_+|},$$

for $|z - E_+| \leq \tau'$.

Proof. We have

$$\frac{\partial}{\partial z} f(\tilde{m}, z) = -\frac{\phi^{-\frac{1}{2}}}{p} \sum_{i=1}^{p} \frac{\sigma_i^2 r_i}{(z - \sigma_i r_i + \phi^{-\frac{1}{2}} \tilde{m}(z)\sigma_i)^2} - 1,$$

which gives

$$|\frac{\partial}{\partial z} f(\tilde{m}(E_+), E_+)| \asymp 1.$$

By implicit function theorem, there exists function $z(\tilde{m}(E_+))$ such that

$$|\frac{\partial^2 z(\tilde{m}(E_+))}{\partial \tilde{m}^2}| = |\frac{\frac{\partial^2}{\partial m^2} f(\tilde{m}(E_+), E_+)}{\frac{\partial}{\partial z} f(\tilde{m}(E_+), E_+)}| \asymp \max\{1, \phi^{-2}\}.$$

Thus, for $|z - E_+| \le \tau'$,

$$z - E_+ = \frac{1}{2} \frac{\partial^2 z(\tilde{m}(E_+))}{\partial m^2} (\tilde{m}(z) - \tilde{m}(E_+))^2 + O(|\tilde{m}(z) - \tilde{m}(E_+)|^3),$$

and the estimation for $\tilde{m}(z) - \tilde{m}(E_+) \asymp \min\{1, \phi\}\sqrt{|z - E_+|}$ follows. By the definition of $\bar{m}_p(z) = \operatorname{tr}\left(\boldsymbol{\Sigma}\boldsymbol{R}'' - \phi^{-1/2} z\tilde{m}(z)\boldsymbol{\Sigma} - z\boldsymbol{I}_p\right)^{-1}/p$, we can show the similar estimation for $\bar{m}_p(z) - \bar{m}_p(E_+)$. The behavior of the density $\rho(z)$ can be derived by $\bar{m}_p(z)$ using a standard argument as in Knowles & Yin (2017). □

We now give a finer characterization of the behavior of $\tilde{m}(z)$ in $\mathcal{D}$.

Lemma 2.8. *Under Assumptions 1, 2, 3, 4, and (3), for uniformly $z = E + i\eta \in \mathcal{D}_+$ and $\gamma \ge 1$, we have for some positive constant $c > 0$,*

$$|\tilde{m}(z)| \asymp 1, \quad \Im\tilde{m}(z) \ge c\eta, \tag{20}$$

and

$$\Im\tilde{m}(z) \asymp \begin{cases} \sqrt{|E - E_+| + \eta}, & E \in supp(\rho), \\ \dfrac{\eta}{\sqrt{|E - E_+| + \eta}}, & E \notin supp(\rho). \end{cases} \tag{21}$$

Moreover, we have

$$\tilde{m}'(z) \asymp \frac{1}{\sqrt{|E - E_+| + \eta}}. \tag{22}$$

When $\gamma \leq 1$, we have for uniformly $z \in \mathcal{D}_+ \cup \mathcal{D}_-$,

$$\tilde{m}(z) + \phi^{\frac{1}{2}} \lambda_i^{-1}(z) \asymp \phi.$$

Proof. The proof is standard with Lemma 2.7. We refer to Alex et al. (2014) and Knowles & Yin (2017) for proof. □

Finally, the above lemma gives the property of stability, which is important in establishing the local law.

Lemma 2.9. *Suppose $\delta : \mathcal{D} \to (0, \infty)$ satisfies $n^{-2} \leq \delta(z) \leq \ln^{-1} n$ and δ is Lipschitz continuous with Lipschitz constants n^2. Moreover, the function $\eta \to \delta(E + i\eta)$ is nonincreasing for $\eta > 0$. Suppose $u : \mathcal{D} \to \mathbb{C}$ is the Stielthes transform of a probability measure on $[0, C]$ with a large enough constant C. Then under Assumptions 1, 2, 3, 4, and (3), suppose for some z and all $z' \in L(z) = \{z' \in \mathcal{D} | \Re z' = \Re z, \Im z' \in [\Im z, 1] \cap [n^{-10}\mathbb{N}]\}$ we have*

$$|f(u(z'), z')| \leq \delta(z'),$$

then we have for some constant C' independent of z and n,

$$|u(z) - \tilde{m}(z)| \leq \frac{C'\delta(z)}{\sqrt{|z - E| + \eta + \delta(z)}}.$$

Proof. The stability can be established using the same method as in Alex et al. (2014) and Knowles & Yin (2017). The main inputs are the Lemmas 2.6, 2.7, and 2.8. □

This lemma shows that once we can control the term $f(m_n(z), z)$, we can bound the difference $m_n(z) - \tilde{m}(z)$. This is the case where we use the Lemma 2.5 to bound $f(m_n(z), z)$. In the next subsection, we will show how this method works.

2.1.2 Weak entry-wise local law for diagonal covariance

Lemma 2.10. *Under Assumptions 1, 2, 3, 4, and (3), we have for $z = E + i\eta \in \mathcal{D}$ and $\eta \geq 1$,*

$$\Lambda(z) \prec n^{-\frac{1}{4}}.$$

Proof. The result is immediate using Lemmas 1.4 and 2.9. □

We now prove the weak entry-wise local law for diagonal covariance.

Proposition 1 (Weak entry-wise local law for diagonal covariance)**.** Under Assumptions 1, 2, 3, 4, and (3), we have for $z = E + i\eta \in \mathcal{D}$,

$$\Lambda(z) \prec (n\eta)^{-\frac{1}{4}}.$$

Proof. The $\eta \geq 1$ case has been dealt with by Lemma 2.10, and we only need to focus on the small η case. We use a continuity argument as in Alex et al. (2014). The main inputs are Lemmas 2.4, 2.5, and 2.9. □

2.1.3 Strong entry-wise local law for diagonal covariance

To improve the rates of the weak entry-wise local law Proposition 1, we use the stability Lemmas 2.9 and prove the following fluctuation averaging lemma.

Lemma 2.11. *Under Assumptions 1, 2, 3, 4, and (3), suppose Φ is a positive, n-dependent function on $\mathcal{D}$ satisfying $n^{-1/2} \leq \Phi(z) \leq n^{-c}$ for some $c > 0$. Suppose moreover that $\Lambda_n(z) \prec n^{-c}$, $\Lambda_{o,n}(z) \prec \Phi(z)$. Then we have*

$$|[Z]_n(z)| \prec (\Phi(z))^2.$$

Similarly, if for some other $p^{-1/2} \leq \Phi'(z) \leq p^{-c}$, and we have $\Lambda_p(z) \prec p^{-c}$, $\Lambda_{o,p}(z) \prec \Phi'(z)$. Then

$$\phi^{-\frac{1}{2}}|[Z]_p(z)| \prec (\Phi'(z))^2.$$

Proof. The proof is similar to the method used in Alex et al. (2014) and Ding & Wang (2025) once we write

$$Z_t(z) = (1 - \mathbb{E}_t)\frac{1}{G_{tt}(z)},$$

for $t \in \mathcal{I}_p \bigcup \mathcal{I}_{n+p,n}$ and notice

$$m_n^{(i)}(z) = m_n(z) + O_P(\Lambda_0^2(z))$$

for $i \in \mathcal{I}_p$. □

Lemma 2.11 directly leads to the following results.

Lemma 2.12. *Suppose $\Theta \prec (n\eta)^{-t}$ for some $0 < t < 1$ uniformly in $z \in \mathcal{D}$, then we have*

$$|[Z]_p(z)| + |[Z]_n(z)| \prec \frac{\Im\tilde{m}(z) + (n\eta)^{-t}}{n\eta}.$$

Proof. Notice if we define

$$\Phi(z) = \sqrt{\frac{\Im\tilde{m}(z) + (n\eta)^{-t}}{n\eta}},$$

we have by Lemma 2.4 and Proposition 1,

$$\Lambda_{o,n}(z) \prec \Phi(z),\ \Lambda(z) \prec (n\eta)^{-\frac{1}{4}}.$$

By Lemma 2.11, we thus get

$$|[Z]_n(z)| \prec \frac{\Im\tilde{m}(z) + (n\eta)^{-t}}{n\eta}.$$

Similarly, define

$$\Phi'(z) = \sqrt{\frac{\Im \tilde{m}(z) + (n\eta)^{-t}}{p\eta}},$$

we have

$$\Lambda_{o,p}(z) \prec \Phi'(z), \ \Lambda(z) \prec (p\eta)^{-\frac{1}{4}}.$$

Using above Lemma 2.11 again, we obtain

$$|[Z]_p(z)| \prec \frac{\Im \tilde{m}(z) + (n\eta)^{-t}}{n\eta}.$$

□

We can now prove the strong entry-wise local law.

Proposition 2 (Strong entry-wise local law for diagonal covariance)**.** Under Assumptions 1, 2, 3, 4, and (3), we have for $z = E + i\eta \in \mathcal{D}$,

$$\Lambda(z) \prec \Psi(z),$$

where

$$\Psi(z) = \sqrt{\frac{\Im \tilde{m}(z)}{n\eta}} + \frac{1}{n\eta}.$$

Proof. Notice by the weak local law, the event $\mathcal{E}$ happens with high probability. Since $\Theta(z) \prec (n\eta)^{-1/4}$ by Lemma 2.4, we take

$$\Phi(z) = \sqrt{\frac{\Im \tilde{m}(z) + (n\eta)^{-\frac{1}{4}}}{n\eta}} + \frac{1}{n\eta}.$$

By Lemma 2.5, we have

$$f(m_n(z), z) \prec \frac{\Im \tilde{m}(z) + (n\eta)^{-\frac{1}{4}}}{n\eta}.$$

By stability Lemma 2.9, we get

$$|m_n(z) - \tilde{m}(z)| \prec \frac{\Im \tilde{m}(z)}{n\eta\sqrt{|z-E|+\eta}} + \frac{1}{(n\eta)^{\frac{5}{8}}} \prec \frac{1}{(n\eta)^{\frac{5}{8}}},$$

as $\Im \tilde{m}(z) = O(\sqrt{|z-E|+\eta})$ by Lemma 2.8. This and a simple union-bound argument lead to

$$\Theta_n(z) \prec \frac{1}{(n\eta)^{\frac{5}{8}}}.$$

This then gives

$$\Lambda_{o,n}(z) \prec \sqrt{\frac{\Im \tilde{m}(z) + (n\eta)^{-\frac{5}{8}}}{n\eta}},$$

which in turn improves the raw bound of $\Phi(z)$ above. In general, we can prove the following self-improving estimate

$$\Theta_n(z) \prec (n\eta)^{-\tau} \;\Rightarrow\; \Theta_n(z) \prec (n\eta)^{-\frac{1}{2}-\frac{\tau}{2}}.$$

This implies the result

$$\Theta_n(z) \prec \frac{1}{n\eta}.$$

Similar arguments give the

$$\Theta_p(z) \prec \frac{\phi^{-\frac{1}{2}}}{n\eta},$$

and thus

$$\Theta(z) = \phi^{\frac{1}{2}}\Theta_p(z) + \Theta_n(z) \prec \frac{1}{n\eta}.$$

Combining with Lemma 2.4 completes the proof. □

2.1.4 Average local law for diagonal covariance

Based on the above results, we proceed to prove the average local law for the diagonal covariance case. Recall we have

$$m_p(z) = \frac{1}{z}\frac{1}{p}\sum_{i\in\mathcal{I}_p}\frac{G_{ii}(z)}{\sigma_i},$$

and

$$\bar{m}_p(z) = \frac{1}{z}\frac{1}{p}\sum_{i\in\mathcal{I}_p}\frac{\tilde{m}_i(z)}{\sigma_i} = -\frac{1}{p}\sum_{i\in\mathcal{I}_p}\frac{1}{\phi^{-\frac{1}{2}}z\tilde{m}(z)\sigma_i + z - \sigma_i r_i}.$$

Using $\Theta(z) \prec \frac{1}{n\eta}$, we have

$$\frac{1}{G_{ii}(z)} = (-\frac{1}{\lambda_i(z)} - \phi^{-\frac{1}{2}}m_n(z) - Z_i(z) + O_{\prec}(\phi^{-\frac{1}{2}}\Psi^2(z)),$$

i.e.

$$G_{ii}(z) = -\frac{1}{p}\sum_{i\in\mathcal{I}_p}\frac{z\sigma_i}{\phi^{-\frac{1}{2}}zm_n(z)\sigma_i + z\sigma_i Z_i(z) + z - \sigma_i r_i} + O_{\prec}(\phi^{-\frac{1}{2}}\Psi^2(z)).$$

This gives

$$m_p(z) = -\frac{1}{p}\sum_{i\in\mathcal{I}_p}\frac{1}{\phi^{-\frac{1}{2}}zm_n(z)\sigma_i + z - \sigma_i r_i} + \frac{1}{p}\sum_{i\in\mathcal{I}_p}\frac{z\sigma_i}{(\phi^{-\frac{1}{2}}zm_n(z)\sigma_i + z - \sigma_i r_i)^2}Z_i(z) + O_{\prec}(\phi^{-1}\Psi^2(z)).$$

Similar to Lemma 2.12, we can show that

Lemma 2.13. *Under Assumptions 1, 2, 3, 4, and (3), we have for $z \in \mathcal{D}$,*

$$\frac{1}{p}\sum_{i\in\mathcal{I}_p}\frac{z\sigma_i}{(\phi^{-\frac{1}{2}}zm_n(z)\sigma_i + z - \sigma_i r_i)^2}Z_i(z) \prec \phi^{-1}\Psi^2(z).$$

This gives the following results.

Proposition 3 (Average local law for diagonal covariance)**.** Under Assumptions 1, 2, 3, 4, and (3), we have for $z = E + i\eta \in \mathcal{D}$,

$$|m_p(z) - \bar{m}_p(z)| \prec \frac{1}{p\eta}.$$

Proof. Directly calculation gives

$$|m_p(z) - \bar{m}_p(z)| = \left| -\frac{1}{p}\sum_{i\in\mathcal{I}_p} \frac{1}{\phi^{-\frac{1}{2}} z m_n(z)\sigma_i + z - \sigma_i r_i} + \frac{1}{p}\sum_{i\in\mathcal{I}_p} \frac{1}{\phi^{-\frac{1}{2}} z \tilde{m}(z)\sigma_i + z - \sigma_i r_i} \right|$$
$$+ \left| \frac{1}{p}\sum_{i\in\mathcal{I}_p} \frac{z\sigma_i}{(\phi^{-\frac{1}{2}} z m_n(z)\sigma_i + z - \sigma_i r_i)^2} Z_i(z) \right| + O_{\prec}(\phi^{-1}\Psi^2(z))$$
$$\prec \frac{\phi^{-\frac{1}{2}}}{n\eta} + \phi^{-1}\Psi^2(z) \prec \frac{\phi^{-1}}{n\eta} = \frac{1}{p\eta}.$$

□

2.2 Anisotropic local law for Gaussian entries

We have proved the entry-wise local law for diagonal $\mathbf{\Sigma}, \boldsymbol{R}$. It remains to prove the entry-wise local law and the anisotropic local law for general $\mathbf{\Sigma}, \boldsymbol{R}$. Denote

$$\boldsymbol{G}(z) = \begin{bmatrix} \boldsymbol{G}_{11}(z) & \boldsymbol{G}_{12}(z) & \boldsymbol{G}_{13}(z) \\ \boldsymbol{G}_{21}(z) & \boldsymbol{G}_{22}(z) & \boldsymbol{G}_{23}(z) \\ \boldsymbol{G}_{31}(z) & \boldsymbol{G}_{32}(z) & \boldsymbol{G}_{33}(z) \end{bmatrix}, \ \bar{\boldsymbol{G}}(z) = \begin{bmatrix} \bar{\boldsymbol{G}}_{11}(z) & \bar{\boldsymbol{G}}_{12}(z) & \bar{\boldsymbol{G}}_{13}(z) \\ \bar{\boldsymbol{G}}_{21}(z) & \bar{\boldsymbol{G}}_{22}(z) & \bar{\boldsymbol{G}}_{23}(z) \\ \bar{\boldsymbol{G}}_{31}(z) & \bar{\boldsymbol{G}}_{32}(z) & \bar{\boldsymbol{G}}_{33}(z) \end{bmatrix}.$$

We can represent the anisotropic law as the following six equations: when $\gamma \geq 1$,

$$\begin{aligned}
\boldsymbol{u}_1^T(\boldsymbol{G}_{11}(z) - \bar{\boldsymbol{G}}_{11}(z))\boldsymbol{u}_2 &= O_{\prec}(\phi^{-\frac{1}{2}}\|\boldsymbol{u}_1\|\|\boldsymbol{u}_2\|\Psi(z)), \\
\boldsymbol{u}_1^T(\boldsymbol{G}_{22}(z) - \bar{\boldsymbol{G}}_{22}(z))\boldsymbol{u}_2 &= O_{\prec}(\phi^{-2}\|\boldsymbol{u}_1\|\|\boldsymbol{u}_2\|\Psi(z)), \\
\boldsymbol{v}_1^T(\boldsymbol{G}_{33}(z) - \bar{\boldsymbol{G}}_{33}(z))\boldsymbol{v}_2 &= O_{\prec}(\|\boldsymbol{v}_1\|\|\boldsymbol{v}_2\|\Psi(z)), \qquad (23) \\
\boldsymbol{u}_1^T(\boldsymbol{G}_{12}(z) - \bar{\boldsymbol{G}}_{12}(z))\boldsymbol{u}_2 &= O_{\prec}(\phi^{-\frac{5}{4}}\|\boldsymbol{u}_2\|\|\boldsymbol{u}_1\|\Psi(z)), \\
\boldsymbol{u}_1^T(\boldsymbol{G}_{13}(z) - \bar{\boldsymbol{G}}_{13}(z))\boldsymbol{v}_1 &= O_{\prec}(\phi^{-\frac{1}{4}}\|\boldsymbol{u}_1\|\|\boldsymbol{v}_1\|\Psi(z)), \\
\boldsymbol{u}_1^T(\boldsymbol{G}_{23}(z) - \bar{\boldsymbol{G}}_{23}(z))\boldsymbol{v}_1 &= O_{\prec}(\phi^{-1}\|\boldsymbol{u}_1\|\|\boldsymbol{v}_1\|\Psi(z)),
\end{aligned}$$

where $\boldsymbol{u}_1, \boldsymbol{u}_2$ are vectors in $\mathbb{R}^p$ and $\boldsymbol{v}_1, \boldsymbol{v}_2$ are vectors in $\mathbb{R}^n$; when $\gamma < 1$,

$$\begin{aligned}
\boldsymbol{u}_1^T(\boldsymbol{G}_{11}(z) - \bar{\boldsymbol{G}}_{11}(z))\boldsymbol{u}_2 &= O_\prec(\phi^{-\frac{3}{2}}\|\boldsymbol{u}_1\|\|\boldsymbol{u}_2\|\Psi(z)),\\
\boldsymbol{u}_1^T(\boldsymbol{G}_{22}(z) - \bar{\boldsymbol{G}}_{22}(z))\boldsymbol{u}_2 &= O_\prec(\phi^{-1}\|\boldsymbol{u}_1\|\|\boldsymbol{u}_2\|\Psi(z)),\\
\boldsymbol{v}_1^T(\boldsymbol{G}_{33}(z) - \bar{\boldsymbol{G}}_{33}(z))\boldsymbol{v}_2 &= O_\prec(\|\boldsymbol{v}_1\|\|\boldsymbol{v}_2\|\Psi(z)),\\
\boldsymbol{u}_1^T(\boldsymbol{G}_{12}(z) - \bar{\boldsymbol{G}}_{12}(z))\boldsymbol{u}_2 &= O_\prec(\phi^{-\frac{5}{4}}\|\boldsymbol{u}_2\|\|\boldsymbol{u}_1\|\Psi(z)),\\
\boldsymbol{u}_1^T(\boldsymbol{G}_{13}(z) - \bar{\boldsymbol{G}}_{13}(z))\boldsymbol{v}_1 &= O_\prec(\phi^{-\frac{3}{4}}\|\boldsymbol{u}_1\|\|\boldsymbol{v}_1\|\Psi(z)),\\
\boldsymbol{u}_1^T(\boldsymbol{G}_{23}(z) - \bar{\boldsymbol{G}}_{23}(z))\boldsymbol{v}_1 &= O_\prec(\phi^{-\frac{1}{2}}\|\boldsymbol{u}_1\|\|\boldsymbol{v}_1\|\Psi(z)).
\end{aligned} \tag{24}$$

As far as we are aware, the only anisotropic local law when $\frac{p}{n}$ can diverge to infinity is established in Ding et al. (2024), Ding & Wang (2025). Different from the six equations here, they only considered the equations for $\boldsymbol{G}_{11}(z)$ and $\boldsymbol{G}_{33}(z)$, and for non-diagonal entries they showed that

$$G_{i\mu}(z) = O_\prec(\phi^{-\frac{1}{4}}\Psi(z)),\ i = 1, \cdots, p, \mu = p+1, \cdots, 2p+n.$$

These results are strictly weaker than (24). The key observation is that the six equations (24) can be written compactly by only one equation with a better structure. Define the weight matrix

$$\Phi = \begin{bmatrix} \min\{\phi^{\frac{1}{4}}, \phi^{\frac{3}{4}}\} & & \\ & \max\{\phi, \phi^{\frac{1}{2}}\} & \\ & & 1 \end{bmatrix}, \tag{25}$$

we can show that (24) is equivalent to

$$\boldsymbol{u}^T\Phi(\boldsymbol{G}(z) - \bar{\boldsymbol{G}}(z))\Phi\boldsymbol{v} = O_\prec(\|\boldsymbol{u}\|\|\boldsymbol{v}\|\Psi(z)), \tag{26}$$

for any vectors $\boldsymbol{u}, \boldsymbol{v}$ in $\mathbb{R}^{2p+n}$.

We first show that the entry-wise local law for diagonal covariance adapts to this framework.

Proposition 4 (Entry-wise local law for diagonal covariance, new form)**.** Under assumptions of Proposition 2, we have for $z = E + i\eta \in \mathcal{D}$ and $s, t \in \mathcal{I}$,

$$\left(\Phi(\boldsymbol{G}(z) - \bar{\boldsymbol{G}}(z))\Phi\right)_{st} \prec \Psi(z).$$

Proof. From the expression of $\boldsymbol{G}(z)$ (16) and $\bar{\boldsymbol{G}}(z)$ (12), we have

$$\left(\Phi(\boldsymbol{G}(z) - \bar{\boldsymbol{G}}(z))\Phi\right)_{st} \lesssim \Lambda(z) \prec \Psi(z).$$

uniformly for $z = E + i\eta \in \mathcal{D}$ and $s, t \in \mathcal{I}$. □

This form of entry-wise local law preserves the structure of the Green function, showing great convenience for proving anisotropic local law below. We now proceed to prove (26) for Gaussian entries. We first prove the following lemma.

Lemma 2.14. *Assume $\boldsymbol{Z}$ to be Gaussian. If the entry-wise local law holds for diagonal $\boldsymbol{D}(\boldsymbol{\Sigma})$ and $\boldsymbol{D}(\boldsymbol{R})$ satisfying Assumptions 1, 2, 3, 4, and (3), then the entry-wise local law holds if replacing $\boldsymbol{D}(\boldsymbol{\Sigma})$ and $\boldsymbol{D}(\boldsymbol{R})$ by $\boldsymbol{\Sigma}$ and $\boldsymbol{R}$. Here $\boldsymbol{D}(\boldsymbol{A})$ for a matrix $\boldsymbol{A}$ is a diagonal matrix with diagonal entries being the eigenvalues of $\boldsymbol{A}$.*

Proof. Suppose $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{D}(\boldsymbol{\Sigma})\boldsymbol{U}^T, \boldsymbol{R} = \boldsymbol{U}\boldsymbol{D}(\boldsymbol{R})\boldsymbol{U}^T$ where $\boldsymbol{U}$ is orthogonal matrix. Notice we have

$$\begin{bmatrix} \boldsymbol{U} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{n+p} \end{bmatrix} \boldsymbol{G}^{\boldsymbol{D}(\boldsymbol{\Sigma}),\boldsymbol{D}(\boldsymbol{R})}(z) \begin{bmatrix} \boldsymbol{U}^T & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{n+p} \end{bmatrix} =_{\mathrm{d}} \boldsymbol{G}^{\boldsymbol{\Sigma},\boldsymbol{R}}(z),$$

where $=_{\mathrm{d}}$ represents the same distribution and we use the notation $\boldsymbol{G}^{\boldsymbol{\Sigma},\boldsymbol{R}}(z)$ for Grenn function with covariance matrix $\boldsymbol{\Sigma}$ and the addition term $\boldsymbol{R}$. Following the proof strategy of the Proposition 4, we only need to prove the first, third, and fifth equation in (24) to prove the entry-wise local law. We thus need to show

$$|\sum_{k,l \in \mathcal{I}_p} U_{ik} \frac{G_{kl}(z) - \delta_{kl}\tilde{m}_k(z)}{\sigma_k \sigma_l} U_{lj}| \prec \phi^{-\frac{1}{2}}\Psi(z),$$

$$|\sum_{k\in\mathcal{I}_p} U_{ik}\frac{G_{k\mu}(z)}{\sigma_k}| + |\sum_{k\in\mathcal{I}_p}\frac{G_{\mu k}(z)}{\sigma_k\sigma_l}U_{lj}| \prec \phi^{-\frac{1}{4}}\Psi(z),$$

$$|G_{\mu\nu}(z) - \delta_{\mu\nu}\tilde{m}_\mu(z)| \prec \Psi(z),$$

for $i,j \in \mathcal{I}_p$ and $\mu,\nu \in \mathcal{I}_{n+p,n}$, where $\boldsymbol{D}(\boldsymbol{\Sigma}) = \operatorname{diag}(\sigma_1,\cdots,\sigma_p)$ we abbreviate $\boldsymbol{G}(z) = \boldsymbol{G}^{\boldsymbol{D}(\boldsymbol{\Sigma}),\boldsymbol{D}(\boldsymbol{R})}(z)$ for simplicity. Notice the last one is satisfied by the assumption of the entry-wise local law for the diagonal matrices $\boldsymbol{D}(\boldsymbol{\Sigma})$ and $\boldsymbol{D}(\boldsymbol{R})$, and we focus on the first two estimates.

We begin with the first estimate. By assumptions of $\boldsymbol{G}_{kk}(z) - \tilde{m}_k(z) \prec \sigma_k^2\phi^{-1/2}\Psi^2(z)$ and $\boldsymbol{U}$ is orthogonal, we have

$$\begin{aligned}\sum_{k,l\in\mathcal{I}_p} U_{ik}\frac{G_{kl}(z) - \delta_{kl}\tilde{m}_k(z)}{\sigma_k\sigma_l}U_{lj} &= \sum_{k\in\mathcal{I}_p} U_{ik}\frac{G_{kk}(z) - \tilde{m}_k(z)}{\sigma_k^2}U_{kj} + \sum_{k\neq l\in\mathcal{I}_p} U_{ik}\frac{G_{kl}(z)}{\sigma_k\sigma_l}U_{lj} \\ &= O_\prec(\phi^{-\frac{1}{2}}\Psi(z)) + \sum_{k\neq l\in\mathcal{I}_p} U_{ik}\frac{G_{kl}(z)}{\sigma_k\sigma_l}U_{lj}.\end{aligned}$$

We thus only need to prove

$$\sum_{k\neq l\in\mathcal{I}_p} U_{ik}\frac{G_{kl}(z)}{\sigma_k\sigma_l}U_{lj} \prec \phi^{-\frac{1}{2}}\Psi(z).$$

This can be done by the polynomialization method developed in Alex et al. (2014). The second estimate can be proceeded in a similar way by showing that $\sum_{k\in\mathcal{I}_p} U_{ik}G_{k\mu}(z)/\sigma_k \prec \phi^{-1/4}\Psi(z)$ and we refer Alex et al. (2014) and Knowles & Yin (2017) for more details. □

We can now show the anisotropic local law for Gaussian entries.

Proposition 5 (Anisotropic local law for Gaussian entries)**.** Under Assumptions 1, 2, 3, 4, (3) and $\boldsymbol{Z}$ is Gaussian, we have for $z \in \mathcal{D}$ and $\boldsymbol{u},\boldsymbol{v} \in \mathbb{R}^{2p+n}$,

$$\boldsymbol{u}^T\Phi(\boldsymbol{G}(z) - \bar{\boldsymbol{G}}(z))\Phi\boldsymbol{v} = O_\prec(\|\boldsymbol{u}\|\|\boldsymbol{v}\|\Psi(z)). \tag{27}$$

Proof. By Proposition 4 and Lemma 2.14, we can prove the entry-wise local law for $\boldsymbol{G}^{\boldsymbol{O}_1\boldsymbol{\Sigma}\boldsymbol{O}_1^T,\boldsymbol{O}_1\boldsymbol{R}\boldsymbol{O}_2^T}(z)$ where $\boldsymbol{O}_1,\boldsymbol{O}_2$ are orthogonal matrices with dimension p. Notice that similar to the proof of Lemma 2.14, we have

$$\begin{bmatrix} \boldsymbol{O}_1 & & \\ & \boldsymbol{O}_2 & \\ & & \boldsymbol{O}_3 \end{bmatrix} \boldsymbol{G}^{\boldsymbol{\Sigma},\boldsymbol{R}}(z) \begin{bmatrix} \boldsymbol{O}_1^T & & \\ & \boldsymbol{O}_2^T & \\ & & \boldsymbol{O}_3^T \end{bmatrix} =_{\mathrm{d}} \boldsymbol{G}^{\boldsymbol{O}_1\boldsymbol{\Sigma}\boldsymbol{O}_1^T,\boldsymbol{O}_1\boldsymbol{R}\boldsymbol{O}_2^T}(z),$$

where $\boldsymbol{O}_3$ is orthogonal matrix with dimension n. This shows that the entry-wise local law also holds for $\begin{bmatrix} \boldsymbol{O}_1 & & \\ & \boldsymbol{O}_2 & \\ & & \boldsymbol{O}_3 \end{bmatrix} \boldsymbol{G}^{\boldsymbol{\Sigma},\boldsymbol{R}}(z) \begin{bmatrix} \boldsymbol{O}_1^T & & \\ & \boldsymbol{O}_2^T & \\ & & \boldsymbol{O}_3^T \end{bmatrix}$, i.e. for $z = E + i\eta \in \mathcal{D}$ and $s, t \in \mathcal{I}$,

$$\left(\begin{bmatrix} \boldsymbol{O}_1 & & \\ & \boldsymbol{O}_2 & \\ & & \boldsymbol{O}_3 \end{bmatrix} \Phi(\boldsymbol{G}(z) - \bar{\boldsymbol{G}}(z))\Phi \begin{bmatrix} \boldsymbol{O}_1^T & & \\ & \boldsymbol{O}_2^T & \\ & & \boldsymbol{O}_3^T \end{bmatrix} \right)_{st} \prec \Psi(z),$$

where we use the fact that $\begin{bmatrix} \boldsymbol{O}_1 & & \\ & \boldsymbol{O}_2 & \\ & & \boldsymbol{O}_3 \end{bmatrix}$ and $\boldsymbol{\Phi}$ commutes. A continuous argument shows the above inequality also holds uniformly for all orthogonal matrices $\boldsymbol{O}_1, \boldsymbol{O}_2$ and $\boldsymbol{O}_3$. This leads to

$$\boldsymbol{u}^T \Phi(\boldsymbol{G}(z) - \bar{\boldsymbol{G}}(z))\Phi \boldsymbol{v} = O_{\prec}(\|\boldsymbol{u}\|\|\boldsymbol{v}\|\Psi(z)).$$

for uniformly $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{2p+n}$. □

2.3 Anisotropic local law for general entries

In this section, we prove the anisotropic local law for general entries. We will follow the strategies of Knowles & Yin (2017) and prove our results in two steps. By polarization identity, it suffices to prove that

$$(\boldsymbol{G}_{\bar{\Sigma}^{-1}\Phi\boldsymbol{v},\bar{\Sigma}^{-1}\Phi\boldsymbol{v}}(z) - \bar{\boldsymbol{G}}_{\bar{\Sigma}^{-1}\Phi\boldsymbol{v},\bar{\Sigma}^{-1}\Phi\boldsymbol{v}}(z)) = O_{\prec}(\Psi(z)), \tag{28}$$

where we use the generalized entries notations of Definition 5.

To proceed, we define a discrete size so that we can use a self-improved argument. Fix a small $\delta > 0$, constant $C_0 > 0$ such that $C_0\delta \leq \tau/2$, and $\eta > 1/n$. We define

$$\eta_l = \eta n^{\delta l}, l = 0, \cdots, L-1;\ \eta_L = 1,$$

where

$$L = L(\eta) = \max\{l \in \mathbb{N} : \eta n^{\delta(l-1)} < 1\}.$$

Notice we have $L \leq 1/\delta + 1$.

For simplicity, we define $\bar{\boldsymbol{v}} = \bar{\Sigma}^{-1}\Phi\boldsymbol{v}$. Consider a $n^{-\epsilon}$-net $\hat{\mathcal{D}}$ of $\mathcal{D}$, such that $|\hat{\mathcal{D}}| \leq n^{2\epsilon}$ and satisfies

$$E + i\eta \in \hat{\mathcal{D}} \ \Rightarrow\ E + i\eta_l \in \hat{\mathcal{D}},\ l = 1, \cdots, L$$

Let

$$\hat{\mathcal{D}}_m = \{z \in \hat{\mathcal{D}} : \Im z \geq n^{\delta m}\}.$$

We define two sets of events.

$$(A_m) : \text{for uniformly } z \in \hat{\mathcal{D}}_m, \boldsymbol{v} \in \mathbb{R}^{2p+n}, \Im \boldsymbol{G}_{\bar{\boldsymbol{v}}\bar{\boldsymbol{v}}}(z) \prec \Im\tilde{m}(z) + n^{C_0\delta}\Psi(z),$$

and

$$(C_m) : \text{for uniformly } z \in \hat{\mathcal{D}}_m, \boldsymbol{v} \in \mathbb{R}^{2p+n}, |\boldsymbol{G}_{\bar{\boldsymbol{v}}\bar{\boldsymbol{v}}}(z) - \bar{\boldsymbol{G}}_{\bar{\boldsymbol{v}}\bar{\boldsymbol{v}}}(z)| \prec n^{C_0\delta}\Psi(z).$$

The starting point of the self-improving process is the following.

Lemma 2.15. *Under Assumptions 1, 2, 3, 4, (3), the event* (A_0) *holds.*

Proof. For $z \in \hat{\mathcal{D}}_0$,

$$\Im \boldsymbol{G}_{\bar{\boldsymbol{v}}\bar{\boldsymbol{v}}}(z) \leq \Im \bar{\boldsymbol{G}}_{\bar{\boldsymbol{v}}\bar{\boldsymbol{v}}}(z) + \|\bar{\Sigma}^{-1}(\boldsymbol{G}(z) - \bar{\boldsymbol{G}}(z))\bar{\Sigma}^{-1}\| \prec 1.$$

Since $\Im \tilde{m}(z) \asymp 1$, we get the result. □

Another important observation is the next lemma.

Lemma 2.16. *Under Assumptions 1, 2, 3, 4, (3), the event* (C_m) *implies* (A_m) *for* $1 \leq m \leq 1/\delta$.

Proof. The result follows easily once we notice

$$|\bar{\boldsymbol{G}}_{\bar{\boldsymbol{v}}\bar{\boldsymbol{v}}}(z)| \prec \Im \tilde{m}(z),$$

by Assumption 4. □

Thus, it remains to prove the reverse inclusion results.

Lemma 2.17. *Under Assumptions 1, 2, 3, 4, (3), the event* (A_{m-1}) *implies* (C_m) *for* $1 \leq m \leq 1/\delta$.

Once Lemma 2.17 can be shown, we conclude from the above lemmas that $|\boldsymbol{G}_{\bar{\boldsymbol{v}}\bar{\boldsymbol{v}}}(z) - \bar{\boldsymbol{G}}_{\bar{\boldsymbol{v}}\bar{\boldsymbol{v}}}(z)| \prec n^{C_0\delta}\Psi(z)$ for all $z \in \hat{\mathcal{D}}$. Since δ can be chosen arbitrarily small, we can show that (28) holds for all $z \in \mathcal{D}$ and the proof of the anisotropic local law is complete. To show this, we need to estimate

$$F_{\boldsymbol{v}}(\boldsymbol{Z}, z) = |\boldsymbol{G}_{\bar{\boldsymbol{v}}\bar{\boldsymbol{v}}}(z) - \bar{\boldsymbol{G}}_{\bar{\boldsymbol{v}}\bar{\boldsymbol{v}}}(z)|.$$

It can be seen that the Lemma 2.17 is implied by the following lemma by Markov inequality.

Lemma 2.18. *Under Assumptions 1, 2, 3, 4, (3), and assume the event* (A_{m-1}) *holds, we have for some fixed even number* r *and all* $z \in \hat{\mathcal{D}}_m$*,*

$$\mathbb{E}[F_{\boldsymbol{v}}^r(\boldsymbol{Z}, z)] \leq (n^{C_0\delta}\boldsymbol{\Psi}(z))^r.$$

The rest of this subsection will be devoted to proving Lemma 2.18. We follow the strategy of Knowles & Yin (2017) using the Bernoulli interpolation and the "word" argument to prove the lemma. We first define the interpolation. We introduce the notation $Z^0 = G^{\text{Gaussian}}$ and $Z^1 = Z$ for random variable Z. For $u \in \{0, 1\}$, define $\rho_{\mu i}^u$ to be the law of $Z_{\mu i}^u$. For $\theta \in (0, 1)$, we define the interpolation law

$$\rho_{\mu i}^\theta = \theta \rho_{\mu i}^1 + (1-\theta)\rho_{\mu i}^0.$$

Let $(\boldsymbol{Z}^0, \boldsymbol{Z}^\theta, \boldsymbol{Z}^1)$ be a triple of independent $\mathcal{I}_{n+p,n} \times \mathcal{I}_p$ matrices with law

$$\Pi_{i\in\mathcal{I}_p}\Pi_{\mu\in\mathcal{I}_{n+p,n}}\rho_{\mu i}^u(\mathrm{d}Z_{\mu i}^u)$$

with $u \in \{0, \theta, 1\}$. We also define the matrix $\boldsymbol{Z}_{(\mu i)}^{\theta,\lambda}$ by replacing the (μi)-entry of $\boldsymbol{Z}^\theta$ by λ and keep the remaining entries unchanged. With this, we define the Green's functions $\boldsymbol{G}^\theta(z)$ and $\boldsymbol{G}_{(\mu i)}^{\theta,\lambda}(z)$ by replacing $\boldsymbol{Z}$ in $\boldsymbol{G}(z)$ (4) by $\boldsymbol{Z}^\theta$ and $\boldsymbol{Z}_{(\mu i)}^{\theta,\lambda}$ respectively.

The key lemma we shall use intensively is the following one.

Lemma 2.19. *Consider the deterministic* Ψ *satisfying* $\Psi \geq n^{-c}$ *for some* c*. Let* X *be a random variable such that* $X \prec \Psi$ *and* $\mathbb{E}X^2 \leq n^c$*. Then we have*

$$\mathbb{E}[X] \prec \Psi.$$

Proof. We choose any $\epsilon > 0$, and calculate

$$\mathbb{E}X = \mathbb{E}X1_{(X\leq N^\epsilon\Psi)} + \mathbb{E}X1_{(X>N^\epsilon\Psi)} \leq N^\epsilon\Psi + \sqrt{\mathbb{E}X^2}\sqrt{\mathbb{P}\left(X > N^\epsilon\Psi\right)} \leq N^\varepsilon\Psi + N^{\frac{c-D}{2}},$$

for arbitrary $D \geq 0$. The result follows by choosing $D \geq 3C$. $\square$

Lemma 2.19 allows us to derive the bounds of the form $\mathbb{E}[X] \prec \Psi$ from $X \prec \Psi$. Similar to Knowles & Yin (2017), we will not always mention the assumption $\mathbb{E}X^2 \leq n^c$ as it can be easily verified by Lemma 1.3.

We will prove Lemma 2.18 using the interpolation between $\boldsymbol{Z}^0$ and $\boldsymbol{Z}^1$. We begin to prove the inequality for $\boldsymbol{Z}^0$.

Lemma 2.20. *Under Assumptions 1, 2, 3, 4, (3), and assume the event* (A_{m-1}) *holds, we have for some fixed even number* r *and all* $z \in \hat{\mathcal{D}}_m$,

$$\mathbb{E}[F_{\boldsymbol{v}}^r(\boldsymbol{Z}^0, z)] \leq (n^{C_0\delta}\boldsymbol{\Psi}(z))^r.$$

Proof. We have by anisotropic law for Gaussian case that

$$F_{\boldsymbol{v}}^r(\boldsymbol{Z}^0, z) \prec \boldsymbol{\Psi}^r(z).$$

It then suffices to prove $\mathbb{E}[F_{\boldsymbol{v}}^{2r}(\boldsymbol{Z}^0, z)] \leq n^C$ for some C, which is easy to obtain. $\square$

With this starting point, we give the following interpolation lemma, whose proof can be referred to Knowles & Yin (2017).

Lemma 2.21. *For any measurable* F*, we have*

$$\mathbb{E}[F(\boldsymbol{Z}^1)] - \mathbb{E}[F(\boldsymbol{Z}^0)] = \int_0^1 \sum_{i \in \mathcal{I}_p} \sum_{\mu \in \mathcal{I}_{n+p,n}} [\mathbb{E}[F(\boldsymbol{Z}_{(\mu i)}^{\theta, Z_{\mu i}^1})] - \mathbb{E}[F(\boldsymbol{Z}_{(\mu i)}^{\theta, Z_{\mu i}^0})]]\mathrm{d}\theta.$$

We shall take $F = F_{\boldsymbol{v}}^r(\boldsymbol{Z}, z)$ in the following proof. By Gronwall type argument, it then suffices to show for uniformly $\boldsymbol{v}, \boldsymbol{\Sigma}, \theta, z \in \hat{\mathcal{D}}_m$, if (A_{m-1}) holds, we have

$$\sum_{i \in \mathcal{I}_p} \sum_{\mu \in \mathcal{I}_{n+p,n}} [\mathbb{E}[F_{\boldsymbol{v}}^r(\boldsymbol{Z}_{(\mu i)}^{\theta, Z_{\mu i}^1}, z)] - \mathbb{E}[F_{\boldsymbol{v}}^r(\boldsymbol{Z}_{(\mu i)}^{\theta, Z_{\mu i}^0}, z)]] = O((n^{C_0\delta}\boldsymbol{\Psi}(z))^r + \sup_{\boldsymbol{w} \in \mathbb{S}} \mathbb{E}[F_{\boldsymbol{w}}^r(\boldsymbol{Z}^\theta, z)]), \tag{29}$$

where $\mathbb{S}$ denotes the unit ball. Once (29) is established, we can prove Lemma 2.18 using Lemmas 2.20 and 2.21. The strategy to prove (29) is to find a function $A(\cdot, z)$ such that for $u = 0, 1$, $\theta \in [0, 1]$, $z \in \hat{\mathcal{D}}_m$ and $\boldsymbol{w} \in \mathbb{S}$,

$$\sum_{i\in\mathcal{I}_p}\sum_{\mu\in\mathcal{I}_{n+p,n}}[\mathbb{E}[F_{\boldsymbol{v}}^r(\boldsymbol{Z}_{(\mu i)}^{\theta,Z_{\mu i}^u}, z)] - \mathbb{E}[A(\boldsymbol{Z}_{(\mu i)}^{\theta,0}, z)]] = O((n^{C_0\delta}\boldsymbol{\Psi}(z))^r + \sup_{\boldsymbol{w}\in\mathbb{S}}\mathbb{E}[F_{\boldsymbol{w}}^r(\boldsymbol{Z}^{\theta}, z)]).$$

In order to make use of the event (A_{m-1}), we derive the following rough bound.

Lemma 2.22. *Under Assumptions 1, 2, 3, 4, (3), for any $z = E + i\eta$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{2p+n}$, we have*

$$|\boldsymbol{G}_{\Phi\boldsymbol{x},\Phi\boldsymbol{y}}(z) - \bar{\boldsymbol{G}}_{\Phi\boldsymbol{x},\Phi\boldsymbol{y}}(z)| \prec n^{2\delta}\sum_{l=1}^{L}(\Im\boldsymbol{G}_{\Phi\boldsymbol{x},\Phi\boldsymbol{x}}(E+i\eta_l) + \Im\boldsymbol{G}_{\Phi\boldsymbol{y},\Phi\boldsymbol{y}}(E+i\eta_l)) + |\bar{\boldsymbol{\Sigma}}\boldsymbol{x}||\bar{\boldsymbol{\Sigma}}\boldsymbol{y}|.$$

Proof. Recall the spectral decomposition of $\boldsymbol{G}(z)$ in (9) and the definition of $\bar{\boldsymbol{G}}(z)$ (12). Under Assumption 4, we have

$$|\boldsymbol{G}_{\Phi\boldsymbol{x},\Phi\boldsymbol{y}}(z) - \bar{\boldsymbol{G}}_{\Phi\boldsymbol{x},\Phi\boldsymbol{y}}(z)| \le \sum_k \frac{\langle\Phi\boldsymbol{x},\boldsymbol{u}_k\rangle^2}{|\lambda_k - z|} + \sum_k \frac{\langle\Phi\boldsymbol{y},\boldsymbol{u}_k\rangle^2}{|\lambda_k - z|} + C|\bar{\boldsymbol{\Sigma}}\boldsymbol{x}||\bar{\boldsymbol{\Sigma}}\boldsymbol{y}|.$$

It suffices to estimate the first term. Setting $\eta_{-1} = 0$ and $\eta_{L+1} = \infty$, we define the subsets of indices

$$U_l = \{k : \eta_{l-1} \le |\lambda_k - E| < \eta_l\} \quad l = 0, 1, \cdots, L+1.$$

We split the summation

$$\sum_k \frac{\langle\Phi\boldsymbol{x},\boldsymbol{u}_k\rangle^2}{|\lambda_k - z|} = \sum_{l=0}^{L+1}\sum_{k\in U_l}\frac{\langle\Phi\boldsymbol{x},\boldsymbol{u}_k\rangle^2}{|\lambda_k - z|},$$

and treat each l separately. For $l = 1, \ldots, L$ we find

$$\sum_{k\in U_l}\frac{\langle\Phi\boldsymbol{x},\boldsymbol{u}_k\rangle^2}{|\lambda_k - z|} \le \sum_{k\in U_l}\frac{\langle\Phi\boldsymbol{x},\boldsymbol{u}_k\rangle^2\eta_l}{(\lambda_k - E)^2} \le 2\sum_{k\in U_l}\frac{\langle\Phi\boldsymbol{x},\boldsymbol{u}_k\rangle^2\eta_l}{(\lambda_k - E)^2 + \eta_{l-1}^2}$$
$$\le \frac{2\eta_l}{\eta_{l-1}}\Im\boldsymbol{G}_{\Phi\boldsymbol{x},\Phi\boldsymbol{x}}(E + i\eta_{l-1}) \le 2n^{\delta}\Im\boldsymbol{G}_{\Phi\boldsymbol{x},\Phi\boldsymbol{x}}(E + i\eta_{l-1}),$$

where the third step follows from (9). Using that the map $y \mapsto y\Im \boldsymbol{G}_{\Phi\boldsymbol{x},\Phi\boldsymbol{x}}(E+iy)$ is nondecreasing, we have for $l=1,\dots,L$,

$$\sum_{k\in U_l} \frac{\langle \Phi\boldsymbol{x}, \boldsymbol{u}_k\rangle^2}{|\lambda_k - z|} \le 2n^{2\delta}\Im \boldsymbol{G}_{\Phi\boldsymbol{x},\Phi\boldsymbol{x}}\left(E + i\eta_{\max\{l,1\}}\right).$$

Next, we estimate

$$\sum_{k\in U_0} \frac{\langle \Phi\boldsymbol{x}, \boldsymbol{u}_k\rangle^2}{|\lambda_k - z|} \le \sum_{k\in U_0} \frac{\langle \Phi\boldsymbol{x}, \boldsymbol{u}_k\rangle^2 2\eta}{(\lambda_k - E)^2 + \eta^2} = 2\Im \boldsymbol{G}_{\Phi\boldsymbol{x},\Phi\boldsymbol{x}}(E+i\eta) \le 2n^{\delta}\Im \boldsymbol{G}_{\Phi\boldsymbol{x},\Phi\boldsymbol{x}}(E+i\eta_1).$$

Finally, we have

$$\sum_{k\in U_{L+1}} \frac{\langle \Phi\boldsymbol{x}, \boldsymbol{u}_k\rangle^2}{|\lambda_k - z|} \le 2\sum_{k\in U_{L+1}} \frac{\langle \Phi\boldsymbol{x}, \boldsymbol{u}_k\rangle^2 |\lambda_k - E|\,\eta_L}{(\lambda_k - E)^2 + \eta_L^2} \prec \sum_{k\in U_{L+1}} \frac{\langle \Phi\boldsymbol{x}, \boldsymbol{u}_k\rangle^2 \eta_L}{(\lambda_k - E)^2 + \eta_L^2} \le \Im \boldsymbol{G}_{\Phi\boldsymbol{x},\Phi\boldsymbol{x}}(E+i\eta_L),$$

This concludes the proof. $\square$

With Lemma 2.22, we can prove the estimate below.

Lemma 2.23. *Under Assumptions 1, 2, 3, 4, (3), and suppose* (A_{m-1}) *holds, we have for any* $z = E + i\eta$,

$$\bar{\Sigma}^{-1}\Phi(\boldsymbol{G}(z) - \bar{\boldsymbol{G}}(z))\Phi\bar{\Sigma}^{-1} = O_{\prec}(n^{2\delta}),$$

and for deterministic $\boldsymbol{v} \in \mathbb{S}$,

$$\Im \boldsymbol{G}_{\bar{\boldsymbol{v}}\bar{\boldsymbol{v}}}(z) \prec n^{2\delta}(\Im\tilde{m}(z) + n^{C_0\delta}\boldsymbol{\Psi}(z)).$$

Proof. Let $E + i\eta \in \hat{\mathcal{D}}_m$. Then we have $E + i\eta_l \in \hat{\mathcal{D}}_{m-1}$ for $l = 1,\cdots,L(\eta)$. Therefore the event (A_{m-1}) and Assumption 4 yield, for any deterministic $\boldsymbol{v} \in \mathbb{S}$,

$$\Im \boldsymbol{G}_{\bar{\boldsymbol{v}}\bar{\boldsymbol{v}}}(E+i\eta_l) \prec |\boldsymbol{v}|^2 + \Im\left\langle \bar{\boldsymbol{v}}, \bar{\boldsymbol{G}}\left(E+i\eta_l\right)\bar{\boldsymbol{v}}\right\rangle \lesssim |\boldsymbol{v}|^2 = 1.$$

The first estimate now follows easily using Lemma 2.22.

In order to prove the second estimate, we use the fact that if $s(z)$ is the Stieltjes transform of any probability measure, the map $\eta \mapsto \eta\Im s(E+i\eta)$ is nondecreasing and the map $\eta \mapsto \eta^{-1}\Im s(E+i\eta)$ is nonincreasing. Let $z=E+i\eta \in \hat{\mathcal{D}}_m$ and $z_1 = E+i\eta_1 \in \hat{\mathcal{D}}_{m-1}$, we therefore have

$$\Im \boldsymbol{G}_{\bar{\boldsymbol{v}}\bar{\boldsymbol{v}}}(z) \le n^{\delta}\Im \boldsymbol{G}_{\bar{\boldsymbol{v}}\bar{\boldsymbol{v}}}(z_1) \prec n^{\delta}\left(\Im \tilde{m}\left(z_1\right)+n^{C_0\delta}\Psi\left(z_1\right)\right) \le n^{2\delta}\left(\Im \tilde{m}(z)+n^{C_0\delta}\Psi(z)\right)$$

where in the second step we used the event (A_{m-1}). □

We now develop the main expansion to prove Lemma 2.18. We define the matrix $\boldsymbol{\Delta}^{\lambda}_{(\mu i)}$ so that only the $(i\mu)$-th entry and (μi)-th entry have value λ, and all other entries are 0. We then have the expansion for λ, λ' and $k \in \mathbb{N}$,

$$\boldsymbol{G}^{\theta,\lambda'}_{(\mu i)}(z) = \boldsymbol{G}^{\theta,\lambda}_{(\mu i)}(z) + \sum_{k=1}^{K} \boldsymbol{G}^{\theta,\lambda}_{(\mu i)}(z)(\boldsymbol{\Delta}^{\frac{\lambda-\lambda'}{(np)^{1/4}}}_{(\mu i)}\boldsymbol{G}^{\theta,\lambda}_{(\mu i)}(z))^k + \boldsymbol{G}^{\theta,\lambda'}_{(\mu i)}(z)(\boldsymbol{\Delta}^{\frac{\lambda-\lambda'}{(np)^{1/4}}}_{(\mu i)}\boldsymbol{G}^{\theta,\lambda}_{(\mu i)}(z))^{K+1}. \tag{30}$$

The following Lemma provides a prior bound for entries of $\boldsymbol{G}^{\theta,\lambda}_{(\mu i)}(z)$.

Lemma 2.24. *Under Assumptions 1, 2, 3, 4, (3) and (A_{m-1}). Suppose y is a random variable satisfying $|y| \prec 1$, we have for any $z = E+i\eta \in \hat{\mathcal{D}}_m$,*

$$\bar{\Sigma}^{-1}\Phi(\boldsymbol{G}^{\theta,y}_{(\mu i)}(z) - \bar{\boldsymbol{G}}(z))\Phi\bar{\Sigma}^{-1} = O_{\prec}(n^{2\delta}),$$

for uniformly $i \in \mathcal{I}_p$ and $\mu \in \mathcal{I}_{n+p,n}$.

Proof. We begin with (30). Take $K=10$, $\lambda' = y$ and $\lambda = \boldsymbol{Z}^{\theta}_{(\mu i)}$so that $\boldsymbol{G}^{\theta,\lambda}_{(\mu i)} = \boldsymbol{G}^{\theta}$. By Lemma 2.23, we can see the estimate holds for $y=\lambda$. In particular, $\bar{\Sigma}^{-1}\Phi\boldsymbol{G}^{\theta,\lambda'}_{(\mu i)}(z)\Phi, \Phi\boldsymbol{G}^{\theta,\lambda'}_{(\mu i)}(z)\Phi\bar{\Sigma}^{-1} \prec n^{2\delta}$. It thus suffices to control the difference $\bar{\Sigma}^{-1}\Phi(\boldsymbol{G}^{\theta,\lambda'}_{(\mu i)}(z) - \boldsymbol{G}^{\theta,\lambda}_{(\mu i)}(z))\Phi\bar{\Sigma}^{-1}$. Using (30) and $|\lambda'-\lambda| \prec 1$ we can derive the results. □

For simplicity, we denote

$$f_{(\mu i)}(\lambda) = F^{r}_{\boldsymbol{v}}(\boldsymbol{Z}^{\theta,(np)^{1/4}\lambda}_{(\mu i)}, z),$$

as we will fix $\boldsymbol{v}, \theta, z \in \hat{\mathcal{D}}_m$ in the following. We also denote $f^{(n)}_{(\mu i)}(\lambda)$ by the n-th derivative of $f_{(\mu i)}(\lambda)$. The following lemma is easily derived from the last lemma.

Lemma 2.25. *Under Assumptions 1, 2, 3, 4, (3) and* (A_{m-1})*. Suppose* y *is a random variable satisfying* $|y| \prec (np)^{-1/4}$*, we have for any* $z = E + i\eta \in \hat{\mathcal{D}}_m$*,*

$$|f_{(\mu i)}^{(k)}(y)| \prec n^{2\delta(r+k)},$$

and the Taylor expansion,

$$f_{(\mu i)}(y) = \sum_{k=1}^{4r} \frac{y^k}{k!} f_{(\mu i)}^{(0)}(0) + O_{\prec}(\Psi^r(z)).$$

Using the above lemma, we have

$$\begin{aligned}
&\mathbb{E}[F_{\boldsymbol{v}}^r(\boldsymbol{Z}_{(\mu i)}^{\theta, Z^1_{(\mu i)}})] - \mathbb{E}[F_{\boldsymbol{v}}^r(\boldsymbol{Z}_{(\mu i)}^{\theta,0})] = \mathbb{E}[f_{(\mu i)}(\frac{Z^1_{\mu i}}{(np)^{1/4}})] - \mathbb{E}[f_{(\mu i)}(0)] \\
=&\mathbb{E}[f_{(\mu i)}(0)] + \frac{1}{2(np)^{1/2}}\mathbb{E}[f_{(\mu i)}^{(2)}(0)] + \sum_{k=3}^{4r} \frac{1}{k!}\mathbb{E}[f_{(\mu i)}^{(k)}(0)]\mathbb{E}[(\frac{Z^1_{\mu i}}{(np)^{1/4}})^k] + O_{\prec}(\Psi^r(z)).
\end{aligned}$$

It thus suffices to prove that for $k = 3, \cdots, 4r$, we have

$$(np)^{-\frac{k}{4}} \sum_{i \in \mathcal{I}_p} \sum_{\mu \in \mathcal{I}_{n+p,n}} |\mathbb{E}[f_{(\mu i)}^{(k)}(0)]| = O((n^{C_0\delta}\Psi(z))^r + \sup_{\boldsymbol{w} \in \mathbb{S}} \mathbb{E}[F_{\boldsymbol{w}}^r(\boldsymbol{Z}^\theta, z)]). \tag{31}$$

We first notice that we can turn (31) into proving (32).

Lemma 2.26. *Under Assumptions 1, 2, 3, 4, (3) and* (A_{m-1})*. If we have for any* $z = E + i\eta \in \hat{\mathcal{D}}_m$ *and for* $k = 3, \cdots, 4r$*,*

$$(np)^{-\frac{k}{4}} \sum_{i \in \mathcal{I}_p} \sum_{\mu \in \mathcal{I}_{n+p,n}} |\mathbb{E}[f_{(\mu i)}^{(k)}(\frac{Z^\theta_{\mu i}}{(np)^{1/4}})]| = O((n^{C_0\delta}\Psi(z))^r + \sup_{\boldsymbol{w} \in \mathbb{S}} \mathbb{E}[F_{\boldsymbol{w}}^r(\boldsymbol{Z}^\theta, z)]), \tag{32}$$

then (31) holds.

Proof. Notice the identity

$$\mathbb{E}[f_{(\mu i)}^{(l)}(0)] = \mathbb{E}[f_{(\mu i)}^{(l)}(\zeta)] - \sum_{k=1}^{4r-l} \frac{1}{k!}\mathbb{E}[f_{(\mu i)}^{(l+l)}(0)]\mathbb{E}[\zeta^k] + O_{\prec}(n^{\frac{l}{2}}\Psi^r(z)),$$

for $l \leq 4r$ and we denote $\zeta = \frac{Z^{\theta}_{\mu i}}{(np)^{1/4}}$. repeatedly use this identity and we derive

$$\begin{aligned}
\mathbb{E}[f^{(k)}_{(\mu i)}(0)] &= \mathbb{E}[f^{(k)}_{(\mu i)}(\zeta)] - \sum_{k_1 \geq 1} 1_{\{k+k_1 \leq 4r\}} \frac{1}{k_1!} \mathbb{E}[f^{(k+k_1)}_{(\mu i)}(0)] \mathbb{E}[\zeta^{k_1}] + O_{\prec}(n^{\frac{k}{2}} \Psi^r(z)) \\
&= \mathbb{E}[f^{(k)}_{(\mu i)}(\zeta)] - \sum_{k_1 \geq 1} 1_{\{k+k_1 \leq 4r\}} \frac{1}{k_1!} \mathbb{E}[f^{(k+k_1)}_{(\mu i)}(\zeta)] \mathbb{E}[\zeta^{k_1}] \\
&+ \sum_{k_1, k_2 \geq 1} 1_{\{k+k_1+k_2 \leq 4r\}} \frac{\mathbb{E}[\zeta^{k_1}]}{k_1!} \frac{\mathbb{E}[\zeta^{k_2}]}{k_2!} \mathbb{E}[f^{(k+k_1+k_2)}_{(\mu i)}(0)] + O_{\prec}(n^{\frac{k}{2}} \Psi^r(z)) \\
&= \cdots = \sum_{q=0}^{4r-k} (-1)^q \sum_{k_1, \cdots, k_q} 1_{\{k+\sum_{j=1}^q k_j \leq 4r\}} \prod_{j=1}^{q} \frac{\mathbb{E}[\zeta^{k_q}]}{k_q!} \mathbb{E}[f^{(k+\sum_{j=1}^q k_j)}_{(\mu i)}(\zeta)] + O_{\prec}(n^{\frac{k}{2}} \Psi^r(z)).
\end{aligned}$$

This gives the result. □

It thus only remains to prove (32). Since only $\boldsymbol{Z}^{\theta}$ is involved, we abuse the notation to replace $\boldsymbol{Z}^{\theta}$ by $\boldsymbol{Z}$. In other words, we will prove

Lemma 2.27. *Under Assumptions 1, 2, 3, 4, (3) and* (A_{m-1}). *We have for any* $z = E + i\eta \in \hat{\mathcal{D}}_m$ *and for* $k = 3, \cdots, 4r$,

$$\sum_{i \in \mathcal{I}_p} \sum_{\mu \in \mathcal{I}_{n+p,n}} |\mathbb{E}[(\frac{\partial}{\partial Z_{\mu i}})^k F^r_{\boldsymbol{v}}(\boldsymbol{Z})]| = O((n^{C_0 \delta} \boldsymbol{\Psi}(z))^r + \sup_{\boldsymbol{w} \in \mathbb{S}} \mathbb{E}[F^r_{\boldsymbol{w}}(\boldsymbol{Z}, z)]).$$

Proof. The case $k \geq 4$ can be proved using the strategy involving an algebra structure called 'words', and extra efforts are needed to deal with the $k = 3$ case with the algebra structure 'lagged words'. We refer to Knowles & Yin (2017) for details. □

Combining the above lemmas completes the proof of Lemma 2.18, thus demonstrating the anisotropic local law for general entries. In summary, we prove the following proposition.

Proposition 6 (Anisotropic local law for general entries)**.** Under Assumptions 1, 2, 3, 4, (3), we have for $z \in \mathcal{D}$ and $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{2p+n}$,

$$\boldsymbol{u}^T \Phi(\boldsymbol{G}(z) - \bar{\boldsymbol{G}}(z)) \Phi \boldsymbol{v} = O_{\prec}(\|\boldsymbol{u}\| \|\boldsymbol{v}\| \Psi(z)). \tag{33}$$

Following the similar strategy of proving the anisotropic local law, we show the averaged local law for general entries.

Proposition 7 (Averaged local law for general entries)**.** Under Assumptions 1, 2, 3, 4, (3), we have for $z \in \mathcal{D}$,

$$|m_p(z) - \bar{m}_p(z)| \prec \frac{1}{p\eta}. \tag{34}$$

Proof. Note that we have proved this averaged local law for diagonal covariance in Proposition 3 and it can extend easily to Gaussian entries. We can follow the proof strategy for the anisotropic local law to derive averaged local laws for general entries. Fortunately, we do not need to iterate like the above for the final bound, as the prior bound given by the anisotropic local law is strong enough. Define

$$\tilde{F}(\boldsymbol{Z}) = |m_p(z) - \bar{m}_p(z)|.$$

Following the argument above, we find it suffices to prove

$$\sum_{i \in \mathcal{I}_p} \sum_{\mu \in \mathcal{I}_{n+p,n}} |\mathbb{E}[(\frac{\partial}{\partial Z_{\mu i}})^k \tilde{F}^p(\boldsymbol{Z})]| = O((n^{C_0 \delta} \boldsymbol{\Psi}(z))^r + \mathbb{E}[\tilde{F}^r(\boldsymbol{Z}, z)]),$$

for $3 \leq k \leq 4r$. This can be established in a similar manner as in Knowles & Yin (2017), and the average local law follows. $\square$

Remark. Throughout the proof, we make the assumption of (3) for simplicity. We now show that the above propositions still hold without (3). To remove the positive-definite of $\boldsymbol{R}$, we demonstrate that the universality holds for $\boldsymbol{M}_n$ if and only if for $\boldsymbol{M}_n + a\boldsymbol{I}_p$ with some $a \in \mathbb{R}$. This means we can prove the results by replacing $\boldsymbol{R}$ with $\boldsymbol{R} + a\boldsymbol{I}_p$ with a large enough $a > 0$. We can thus assume without loss of generality that $\boldsymbol{R}$ is positive definite. We now show how to remove the assumption that $\boldsymbol{\Sigma}$ is invertible. If $\boldsymbol{\Sigma}$ has zero eigenvalues, we can instead consider the local law with covariance $\boldsymbol{\Sigma} + \epsilon\boldsymbol{I}_p$ with small but positive ϵ. Our results above hold for $\boldsymbol{\Sigma} + \epsilon \boldsymbol{I}_p$. The key observation is that our proof only requires the covariance to be invertible, but does not depend on the smallest eigenvalues of the covariance. Thus, we can let $\epsilon \to 0$, and the local law remains valid.

Combining Propositions 7, 28, and Remark 2.3, we finish the proof of the anisotropic local law.

3 Proof of Theorem 3.1

3.1 Rigidity of eigenvalues

Building on the anisotropic local law theorem, we establish our rigidity result before the universal results. Our first step involves deriving key implications from Theorem 2.1. Recall that ρ is the measure on $\mathbb{R}$ whose Stieltjes transform is the deterministic equivalence $\bar{m}_p(z)$. We define the quantile sequence of ρ as $w_1 \geq \cdots \geq w_p$ such that $\int_{w_i}^{+\infty} \rho(x)\mathrm{d}x = (i - 1/2)/p$ for $i = 1, \cdots, p$. We aim to demonstrate that the eigenvalues $\lambda_i(\boldsymbol{M}_n')$ of $\boldsymbol{M}_n'$ remain close to w_i for $i = 1, \cdots, p$.

Lemma 3.1 (Rigidity of eigenvalues)**.** *Suppose that Assumptions 1,2,3,4 hold.*

(i) When $\gamma \leq 1$, we have for any fixed integer number $k \geq 1$,

$$|\lambda_i(\boldsymbol{M}_n') - w_i|,\ |\lambda_{p-i}(\boldsymbol{M}_n') - w_{p-i}| \prec (\min\{i, p+1-i\})^{-\frac{1}{3}} p^{-\frac{2}{3}}, \tag{35}$$

uniformly for $1 \leq i \leq k$.

(ii) When $\gamma > 1$, we have

$$|\lambda_i(\boldsymbol{M}_n') - w_i| \prec (\min\{i, n+1-i\})^{-\frac{1}{3}} n^{-\frac{2}{3}}, \tag{36}$$

uniformly for $1 \leq i \leq n$.

For the case $\gamma \leq 1$, the proof follows from a routine application of the Helffer-Sjöstrand functional calculus, see Alex et al. (2014). We now focus on the $\gamma > 1$ case.

When $\gamma > 1$, we have

$$\phi^{\frac{1}{2}} - \phi^{-\frac{1}{2}} \lesssim \lambda_i(\phi^{\frac{1}{2}}\hat{\boldsymbol{\Sigma}}) - \|\boldsymbol{R}'\| \leq \lambda_i(\boldsymbol{M}_n') \leq \lambda_i(\phi^{\frac{1}{2}}\hat{\boldsymbol{\Sigma}}) + \|\boldsymbol{R}'\| \lesssim \phi^{\frac{1}{2}} + \phi^{-\frac{1}{2}},\ 1 \leq i \leq n,$$

i.e.

$$\lambda_i(\boldsymbol{M}_n') \asymp \phi^{\frac{1}{2}},\ 1 \leq i \leq n,$$

while

$$\lambda_i(\boldsymbol{M}_n') \leq \lambda_i(\phi^{\frac{1}{2}}\hat{\boldsymbol{\Sigma}}) + \|\boldsymbol{R}'\| = \|\boldsymbol{R}'\| \lesssim \phi^{-\frac{1}{2}} \to 0,\ n+1 \leq i \leq p.$$

This demonstrates the large gap between the first n eigenvalues and the last $p-n$ eigenvalues of $\boldsymbol{M}_n'$. We can thus truncate at the level of a constant and only keep the largest n eigenvalues. Recall that

$$m_p(z) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{\lambda_i - z}.$$

Since we are interested in the largest n eigenvalues, which only stands for a vanishing proportion $\frac{n}{p}$ among all eigenvalues, the fast convergence rate (34) does not provide much information. To solve this problem and focus on the first n out of p eigenvalues, we define

$$m_{p,n}(z) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - z}. \tag{37}$$

Moreover, we choose constants $c_1 < c_2$ and define a cut-off function $K_{c_1,c_2}(z)$, which is 0 on $[0, c_1]$ and 1 on $[c_2, \infty)$ satisfying

$$|K_{c_1,c_2}'(z)| \lesssim 1,\ |K_{c_1,c_2}^{''}(z)| \lesssim 1.$$

For properly chosen c_1, c_2, we have for large enough n,

$$m_{p,n}(z) = \frac{\phi}{p} \sum_{i=1}^{p} \frac{K_{c_1,c_2}(\lambda_i)}{\lambda_i - z}.$$

For simplicity, we write $K = K_{c_1,c_2}$. Recall we have

$$\bar{m}(z) = \int_{\mathbb{R}} \frac{1}{\lambda - z} \bar{\rho}(\lambda) \mathrm{d}\lambda,$$

we consider the cut-off version

$$\bar{m}_c(z) = \int_{\mathbb{R}} \frac{1}{\lambda - z} \bar{\rho}_c(\lambda) \mathrm{d}\lambda,$$

where $\bar{\rho}_c(\lambda)$ is defined as

$$\bar{\rho}_c(\lambda) = \frac{K(\lambda)\bar{\rho}(\lambda)}{\int_{\mathbb{R}} K(\lambda)\bar{\rho}(\lambda)\mathrm{d}\lambda}.$$

This is the limiting Stieltjes transform for the largest n eigenvalues, i.e. limit of $m_{p,n}(z)$.

We have the following result.

Lemma 3.2. *Under Assumptions 1, 2, 3, 4, we have for $z \in \mathcal{D}$,*

$$|m_{p,n}(z) - \bar{m}_c(z)| \prec \frac{1}{n\eta}.$$

Proof. Denote the empirical density $\rho_e(z) = \frac{1}{p}\sum_{i=1}^{p} \delta_{\lambda_i}(z)$. By Proposition 7, we have

$$|m_p(z) - \bar{m}_p(z)| \prec \frac{1}{p\eta},$$

Apply the routine argument, we can show that

$$\int_{\mathbb{R}} \frac{K(\lambda)}{\lambda - z} (\rho_e(\lambda) - \rho(\lambda)) \mathrm{d}\lambda \prec \frac{1}{p\eta}.$$

Using the expression of ρ given above, we have

$$\phi - \int_{\mathbb{R}} K(\lambda)\bar{\rho}(\lambda)\mathrm{d}\lambda \prec \frac{1}{n},$$

which gives

$$\bar{\rho}_c(\lambda) - \phi K(\lambda)\bar{\rho}(\lambda) \prec \frac{1}{n\eta},$$

and thus

$$\bar{m}_c(z) - \phi \int_{\mathbb{R}} \frac{K(\lambda)}{\lambda - z}\bar{\rho}(\lambda)\mathrm{d}\lambda \prec \frac{1}{n\eta}.$$

Notice that

$$m_{p,n}(z) = \phi \int_{\mathbb{R}} \frac{K(\lambda)}{\lambda - z}\rho_e(\lambda)\mathrm{d}\lambda,$$

the result follows from

$$|m_{p,n}(z) - \bar{m}_c(z)| \prec \phi \int_{\mathbb{R}} \frac{K(\lambda)}{\lambda - z}(\rho_e(\lambda) - \bar{\rho}(\lambda))\mathrm{d}\lambda + \frac{1}{n\eta} \prec \frac{1}{n\eta}.$$

□

With Lemma 3.2, we can follow the routine Helffer-Sjöstrand functional calculus and establish the rigidity results when $\gamma > 1$. We omit further details.

3.2 Green function comparison

We now prove the following Green function comparison results on the edge using the averaged local law.

Lemma 3.3 (Green function comparison)**.** *Under Assumptions 1, 2, 3, 4, consider the function* $F : \mathbb{R} \to \mathbb{R}$ *such that*

$$\sup_{x\in\mathbb{R}} |F^{(k)}(x)|(1 + |x|)^{C_1} \leq C_1, k = 1, 2, 3, 4,$$

for some $C_1 > 0$*. For* $\epsilon > 0$ *small enough and constants* $E, E_1 < E_2$*,*

$$|E - E_+|, |E_1 - E_+|, |E_2 - E_+| \leq N^{-\frac{2}{3}+\epsilon},$$

where $N = \min\{n, p\}$*, set* $\eta \le N^{-2/3-\epsilon}$*, we have*

(i) when $\gamma \le 1$*,*

$$|\mathbb{E}F(p\eta\Im m_p(z)) - \mathbb{E}^{Gau}F(p\eta\Im m_p(z))| \le Cp^{-\frac{1}{6}+C\epsilon},\ z = E + i\eta,$$

and

$$|\mathbb{E}[F(p\int_{E_1}^{E_2}\Im m_p(y+i\eta)\mathrm{d}y)] - \mathbb{E}^{Gau}[F(p\int_{E_1}^{E_2}\Im m_p(y+i\eta)\mathrm{d}y)]| \le Cp^{-\frac{1}{6}+C\epsilon},$$

for some constant $C > 0$*. Similar results hold if we replace* E_+ *by* E_-*.*

(ii) when $\gamma > 1$*,*

$$|\mathbb{E}F(n\eta\Im m_{p,n}(z)) - \mathbb{E}^{Gau}F(n\eta\Im m_{p,n}(z))| \le Cn^{-\frac{1}{6}+C\epsilon},\ z = E + i\eta,$$

and

$$|\mathbb{E}[F(n\int_{E_1}^{E_2}\Im m_{p,n}(y+i\eta)\mathrm{d}y)] - \mathbb{E}^{Gau}[F(n\int_{E_1}^{E_2}\Im m_{p,n}(y+i\eta)\mathrm{d}y)]| \le Cn^{-\frac{1}{6}+C\epsilon},$$

for some constant $C > 0$*.*

Proof. The proof is similar to Erdős et al. (2012). The main inputs are the averaged local law 7 and Lemma 3.2. □

The Green function comparison Lemma 3.3 essentially bounds the difference between a smooth function of $m_p(z)$ and its Gaussian counterpart. To utilize this theorem, we introduce the following proposition that connects the distribution of the extreme eigenvalue with the expectation of the Stieltjes transform.

Lemma 3.4. *Consider* $\eta = N^{-2/3-9\epsilon}/2$ *where* $N = \min\{n, p\}$ *for some* $\epsilon > 0$*, and* $|E - E_+| \le N^{-2/3+\epsilon}$*. For a smooth cut-off function* $K \ge 0$ *such that* $K(x) = 0$ *when* $x \ge 2/9$ *and* $K(x) = 1$ *when* $x \le 1/9$*, and any small* ϵ *and any large* $D > 0$*, we have for* n *large enough,* *(i) when $\gamma \leq 1$,*

$$\mathbb{E}[K(\frac{p}{\pi}\int_{E-p^{6\epsilon}\eta}^{E_+ + 2p^{-\frac{2}{3}+\epsilon}} \Im m_p(y+i\eta)\mathrm{d}y)] \leq \mathbb{P}(\lambda_1(\boldsymbol{M}_n') \leq E)$$
$$\leq \mathbb{E}[K(\frac{p}{\pi}\int_{E+p^{6\epsilon}\eta}^{E_+ + 2p^{-\frac{2}{3}+\epsilon}} \Im m_p(y+i\eta)\mathrm{d}y)] + p^{-D}.$$

Similar results hold for $\lambda_p(\boldsymbol{M}_n')$ if we replace E_+ by E_-.

(ii) when $\gamma > 1$,

$$\mathbb{E}[K(\frac{n}{\pi}\int_{E-n^{6\epsilon}\eta}^{E_+ + 2n^{-\frac{2}{3}+\epsilon}} \Im m_{p,n}(y+i\eta)\mathrm{d}y)] \leq \mathbb{P}(\lambda_1(\boldsymbol{M}_n') \leq E)$$
$$\leq \mathbb{E}[K(\frac{n}{\pi}\int_{E+n^{6\epsilon}\eta}^{E_+ + 2n^{-\frac{2}{3}+\epsilon}} \Im m_{p,n}(y+i\eta)\mathrm{d}y)] + n^{-D}.$$

Proof. See Erdős et al. (2012). □

We now provide proof for Theorem 3.1. Fix

$$E = E_+ + tN^{-2/3}, \qquad \eta = N^{-2/3-9\epsilon}/2.$$

We first consider the case $\gamma \leq 1$. Since the Gaussian coupling also satisfies Assumptions 1–4, Lemma 3.4 applies under $\mathbb{P}^{\mathrm{Gau}}$ as well. Hence,

$$\mathbb{P}^{\mathrm{Gau}}(\lambda_1(\boldsymbol{M}_n') \leq E) \leq \mathbb{E}^{\mathrm{Gau}}\left[K\left(\frac{p}{\pi}\int_{E+p^{6\epsilon}\eta}^{E_+ + 2p^{-2/3+\epsilon}} \Im m_p(y+i\eta)\, dy\right)\right] + p^{-D}.$$

Now apply Lemma 3.3 with

$$F(x) = K(x/\pi), \qquad E_1 = E + p^{6\epsilon}\eta, \qquad E_2 = E_+ + 2p^{-2/3+\epsilon}.$$

Then

$$\mathbb{E}^{\mathrm{Gau}}\left[K\left(\frac{p}{\pi}\int_{E+p^{6\epsilon}\eta}^{E_+ + 2p^{-2/3+\epsilon}} \Im m_p(y+i\eta)\, dy\right)\right]$$
$$\leq \mathbb{E}\left[K\left(\frac{p}{\pi}\int_{E+p^{6\epsilon}\eta}^{E_+ + 2p^{-2/3+\epsilon}} \Im m_p(y+i\eta)\, dy\right)\right] + Cp^{-1/6+C\epsilon}.$$

Applying Lemma 3.4 again to the original model, with E replaced by $E + 2p^{6\epsilon}\eta$, yields

$$\mathbb{E}\left[K\left(\frac{p}{\pi}\int_{E+p^{6\epsilon}\eta}^{E_{+}+2p^{-2/3+\epsilon}} \Im m_p(y+i\eta)\,dy\right)\right] \leq \mathbb{P}(\lambda_1(\boldsymbol{M}_n') \leq E + 2p^{6\epsilon}\eta).$$

Combining the above three displays, we obtain

$$\mathbb{P}^{\mathrm{Gau}}(\lambda_1(\boldsymbol{M}_n') \leq E) \leq \mathbb{P}(\lambda_1(\boldsymbol{M}_n') \leq E + 2p^{6\epsilon}\eta) + Cp^{-1/6+C\epsilon} + p^{-D}.$$

Since $\gamma \leq 1$ implies $p \asymp N$, and $2p^{6\epsilon}\eta = p^{6\epsilon}N^{-2/3-9\epsilon} \leq N^{-2/3-\epsilon}$ for N large enough, we have

$$\begin{aligned}\mathbb{P}^{\mathrm{Gau}}\big(N^{2/3}(\lambda_1(\boldsymbol{M}_n') - E_{+}) \leq t\big) \leq\; & \mathbb{P}\big(N^{2/3}(\lambda_1(\boldsymbol{M}_n') - E_{+}) \leq t + N^{-\epsilon}\big) \\ & + N^{-1/6+C\epsilon},\end{aligned}$$

where we choose D large enough and absorb the term p^{-D} into the error $N^{-1/6+C\epsilon}$.

When $\gamma > 1$, the proof is identical, with p and m_p replaced by n and $m_{p,n}$, respectively. This proves the upper bound in Theorem 3.1. The lower bound follows similarly.

4 Proof of Corollary 1

The convergence to the Tracy-Widom distribution for Gaussian entries has been proved by Karoui (2003). By our universality Theorem 3.1, this also holds for general entries.

5 Proof of Theorem 3.2

5.1 Anti-concentration inequality

We first define the Lévy anti-concentration function for random variable T as

$$a(T,\delta) = \sup_{t\in\mathbb{R}} \mathbb{P}(t \leq T \leq t+\delta), \tag{38}$$

for $\delta > 0$. We shall provide bounds for $a(T,\delta)$.

Lemma 5.1 (Anti-concentration inequality). *Under Assumption 1, for any admissible pair* $(\mathbf{\Sigma}, \mathbf{\Sigma}_0)$ *and* $\delta \le N^{-2/3}$ *where* $N = \min\{n, p\}$, *we have*

$$a(T, \delta) \prec N^{\frac{2}{3}}\delta. \tag{39}$$

We assume Assumption 4 first. Similar to the anti-concentration function (38), we define the local anti-concentration function

$$a(X, \delta, \mathcal{D}) = \sup_{t \in \mathcal{D}} \mathbb{P}(t \le X \le t + \delta), \tag{40}$$

for some subset $\mathcal{D} \subset \mathbb{R}$. Using rigidity of $\lambda_1(\boldsymbol{M}_n')$, we can show that for any $\epsilon > 0$, there exists $E_1 < E_+ < E_2$ such that $|E_1 - E_+|, |E_2 - E_+| \le N^{-2/3+\epsilon}$,

$$a(\lambda_1(\boldsymbol{M}_n'), \delta) \le a(\lambda_1(\boldsymbol{M}_n'), \delta, [E_1, E_2]) + N^{-D}.$$

Using Lemma 3.4, for $\eta = N^{-2/3-9\epsilon}/2$, we have

$$\mathbb{P}(E_1 \le \lambda_1 \le E_2) \le \mathbb{E}[\frac{N}{\pi} \int_{E_1 - N^{6\epsilon}\eta}^{E_2 + N^{6\epsilon}\eta} \Im m_N(y + i\eta) \mathrm{d}y] + N^{-D}.$$

We thus have

$$a(\lambda_1(\boldsymbol{M}_n'), \delta, [E_1, E_2]) \le \sup_{E \in [E_1, E_2]} \mathbb{E}[\frac{N}{\pi} \int_{E - N^{6\epsilon}\eta}^{E + \delta + N^{6\epsilon}\eta} \Im m_N(y + i\eta) \mathrm{d}y] + N^{-D}.$$

Using the averaged local law and estimation on $\bar{m}_N$, we have

$$\mathbb{E}[\frac{N}{\pi} \int_{E - N^{6\epsilon}\eta}^{E + \delta + N^{6\epsilon}\eta} \Im m_N(y + i\eta) \mathrm{d}y] \lesssim \mathbb{E}[\frac{N}{\pi} \int_{E - N^{6\epsilon}\eta}^{E + \delta + N^{6\epsilon}\eta} \Im \bar{m}_N(y + i\eta) \mathrm{d}y] \le N^{\frac{2}{3}+\frac{1}{2}\epsilon}(\delta + N^{6\epsilon}\eta) \lesssim N^{\frac{2}{3}+\frac{1}{2}\epsilon}\delta.$$

This shows that for $\delta \le N^{-2/3}$, by choosing $D > \epsilon/2$, we have for any small $\epsilon > 0$,

$$a(\lambda_1(\boldsymbol{M}_n'), \delta) \lesssim N^{\frac{2}{3}+\frac{1}{2}\epsilon}\delta.$$

In other words,

$$a(\lambda_1(\boldsymbol{M}_n'), \delta) \prec N^{\frac{2}{3}}\delta.$$

This proves (31). The proof for $\lambda_p(\boldsymbol{M}_n')$ is similar, and we omit it. Now we assume Assumption 4′. In this case, when $|E_+| > |E_-|$, the largest eigenvalue argument stills holds. Similar argument holds when $|E_+| < |E_-|$.

5.2 Spectrum matching

We first prove the following lemma.

Lemma 5.2. *(i) Under Assumptions 1, 2, 3, 4, there* $\delta > 0$ *such that*

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\big(\lambda_1(\boldsymbol{M}_n) \le t\big) - \mathbb{P}^{Gau}\big(\lambda_1(\boldsymbol{M}_n) \le t\big) \right| \le N^{-\delta}. \tag{41}$$

where $N = \min\{n, p\}$. *If* $\gamma \le 1$, *similar results also hold for* $\lambda_p(\boldsymbol{M}_n)$.

(ii) Under Assumptions 1, 2, 3, 4′, there $\delta > 0$ *such that*

$$\sup_{t\ge 0} \left| \mathbb{P}\big(\sigma_1(\boldsymbol{M}_n) \le t\big) - \mathbb{P}^{Gau}\big(\sigma_1(\boldsymbol{M}_n) \le t\big) \right| \le N^{-\delta}. \tag{42}$$

Proof. By Theorem 3.1, we have

$$\begin{aligned}
&\mathbb{P}(N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}_n') - E_+) \le t) \\
&\quad \le \mathbb{P}^{\mathrm{Gau}}(N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}_n') - E_+) \le t + N^{-\epsilon}) + N^{-\frac{1}{6}+C\epsilon} \\
&\quad \le \mathbb{P}^{\mathrm{Gau}}(N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}_n') - E_+) \le t) + a(\lambda_1(\boldsymbol{M}_n'), N^{-\frac{2}{3}-\epsilon}) + N^{-\frac{1}{6}+C\epsilon},
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}(N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}_n') - E_+) \le t) \\
&\quad \ge \mathbb{P}^{\mathrm{Gau}}(N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}_n') - E_+) \le t - N^{-\epsilon}) - N^{-\frac{1}{6}+C\epsilon} \\
&\quad \ge \mathbb{P}^{\mathrm{Gau}}(N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}_n') - E_+) \le t) - a(\lambda_1(\boldsymbol{M}_n'), N^{-\frac{2}{3}-\epsilon}) - N^{-\frac{1}{6}+C\epsilon}.
\end{aligned}$$

Together with Lemma 5.1, it shows

$$\begin{aligned}
&\left| \mathbb{P}(N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}_n') - E_+) \le t) - \mathbb{P}^{\mathrm{Gau}}(N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}_n') - E_+) \le t) \right| \\
&\le a(\lambda_1(\boldsymbol{M}_n'), N^{-\frac{2}{3}-\epsilon}) + N^{-\frac{1}{6}+C\epsilon} \prec N^{-\epsilon} + N^{-\frac{1}{6}+C\epsilon}.
\end{aligned}$$

Equivalently, this gives for any $\epsilon' > 0$ and $t > 0$, we have

$$\left| \mathbb{P}(N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}_n') - E_+) \le t) - \mathbb{P}^{\mathrm{Gau}}(N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}_n') - E_+) \le t) \right| \le N^{-\epsilon+\epsilon'} + N^{-\frac{1}{6}+C\epsilon+\epsilon'}.$$

As both ϵ and ϵ' can be arbitrarily small, we can choose $6C\epsilon < 1$ and $\epsilon' < \min\{\epsilon, C\epsilon - 1/6\}$, so that there exists $\delta > 0$ such that

$$\left|\mathbb{P}(N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}'_n) - E_+) \le t) - \mathbb{P}^{\text{Gau}}(N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}'_n) - E_+) \le t)\right| \le N^{-\delta}.$$

As a result,

$$\begin{aligned}
&\sup_{t\in\mathbb{R}}\left|\mathbb{P}\big(\lambda_1(\boldsymbol{M}_n) \le t\big) - \mathbb{P}^{\text{Gau}}\big(\lambda_1(\boldsymbol{M}_n) \le t\big)\right| \\
=&\sup_{t\in\mathbb{R}}\left|\mathbb{P}(N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}'_n) - E_+) \le t) - \mathbb{P}^{\text{Gau}}(N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}'_n) - E_+) \le t)\right| \le N^{-\delta}.
\end{aligned}$$

Similar proof applies to $\lambda_p(\boldsymbol{M}'_n)$ when $\gamma \le 1$. This proves the first part.

For the largest singular values $\|\boldsymbol{M}_n\|_{\text{op}} = \max\{\lambda_1(\boldsymbol{M}'_n), -\lambda_p(\boldsymbol{M}'_n)\}$. We have $\lambda_1(\boldsymbol{M}'_n) \to E_+$ and $\lambda_p(\boldsymbol{M}'_n) \to E_-$ almost surely. Thus when $|E_+| > |E_-|$, we have with probability 1, $\lambda_1(\boldsymbol{M}'_n) > -\lambda_p(\boldsymbol{M}'_n)$ and $\|\boldsymbol{M}_n\|_{\text{op}} = \lambda_1(\boldsymbol{M}'_n)$ for n large enough. In this case, we only need to ensure the regularity of E_+. Similar argument holds when $|E_+| < |E_-|$. We can thus replace Assumption 4 by Assumption $4'$ when only dealing with $\|\boldsymbol{M}_n\|_{\text{op}}$. This completes the proof. $\square$

Using

$$\mathbb{P}\Bigg(\hat{\boldsymbol{\Sigma}} \in \mathbf{B}_{\text{op}}(\boldsymbol{\Sigma}_0, t)\Bigg) = \mathbb{P}\Bigg(\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\|_{\text{op}} \le t\Bigg),$$

the uniform Gaussian approximation bound is immediate with the Lemma 5.2. To prove the second part, we first notice that

$$\begin{aligned}
&\sup_{0\le\alpha\le 1}\left|\mathbb{P}\Bigg(T \ge \hat{q}^{\text{sm},B}_{\boldsymbol{\Sigma},\boldsymbol{\Sigma}_0}(\alpha)\Bigg|\boldsymbol{Y}^1, \cdots, \boldsymbol{Y}^B\Bigg) - \alpha\right| \\
&\lesssim \rho_n(\boldsymbol{\Sigma}_0) + \sup_{0\le\alpha\le 1}\left|\mathbb{P}\Bigg(\tilde{T}^{\text{sm}} \ge \hat{q}^{\text{sm},B}_{\boldsymbol{\Sigma},\boldsymbol{\Sigma}_0}(\alpha)\Bigg|\boldsymbol{Y}^1, \cdots, \boldsymbol{Y}^B\Bigg) - \alpha\right|,
\end{aligned}$$

where $\tilde{T}^{\text{sm}}$ is the i.i.d. copy of T^{sm} and is independent of $\boldsymbol{Y}^1, \cdots, \boldsymbol{Y}^B$. We now present the following lemma to describe the second term on the right side.

Lemma 5.3. *For i.i.d. continuous random variables $X, X_1, X_2, \cdots$. If we define the sample upper α-quantile for $X_1, X_2, \cdots, X_B$ as $\hat{q}_X^B(\alpha)$, we have with probability approaching* 1,

$$\sup_{0\le\alpha\le 1}\left|\mathbb{P}\Big(X \ge \hat{q}_X^B(\alpha)\big|X_1, X_2, \cdots, X_B\Big) - \alpha\right| \lesssim B^{-\frac{1}{2}}.$$

Proof. Let the distribution function of X as F, and F^{-1} as its quantile. Then we have $U = F(X)$ and $U_b = F(X_b)$, $b = 1, \cdots$ are i.i.d. random variables with uniform distribution on $[0, 1]$. It can be obtained easily that

$$\mathbb{P}\Big(X \ge \hat{q}_X^B(\alpha)\big|X_1, X_2, \cdots, X_n\Big) = \mathbb{P}\Big(U \ge \hat{q}_U^B(\alpha)\big|F^{-1}(U_1), F^{-1}(U_2), \cdots, F^{-1}(U_n)\Big),$$

where $\hat{q}_U^B(\alpha) = F(\hat{q}_X^B(\alpha))$ is the sample upper α-quantile for $U_1, U_2, \cdots, U_B$. Direct calculation gives that with probability approaching 1, we have

$$\sup_{0\le\alpha\le 1}\left|\mathbb{P}\Big(U \ge \hat{q}_U^B(\alpha)\big|F^{-1}(U_1), F^{-1}(U_2), \cdots, F^{-1}(U_n)\Big) - \alpha\right| \lesssim B^{-\frac{1}{2}}.$$

This gives the result. □

Notice that the statistic T^{sm} whose entries are Gaussian has a continuous distribution. We can thus apply the above Lemma 5.3. This completes the proof.

6 Proof of Theorem 3.3

Recall $\|\boldsymbol{A}\|_{S_1} = \sum_{i=1}\sigma_i(\boldsymbol{A})$, we have

$$\sup_{\boldsymbol{A}:\|\boldsymbol{A}\|_{S_1}\le 1}\langle \boldsymbol{A}, \boldsymbol{\Sigma}\rangle = \|\boldsymbol{\Sigma}\|_{\text{op}}.$$

This shows that

$$\left\{\boldsymbol{\Sigma}\,\middle|\forall \boldsymbol{A}\in\mathbb{R}^{p\times p},\ \langle \boldsymbol{A}, \boldsymbol{\Sigma}\rangle \in \Big[\langle \boldsymbol{A}, \hat{\boldsymbol{\Sigma}}\rangle - \hat{q}^{\text{sm},B}_{\widehat{\text{Sp}}(\boldsymbol{\Sigma})}(\alpha)\ \|\boldsymbol{A}\|_{S_1}, \langle \boldsymbol{A}, \hat{\boldsymbol{\Sigma}}\rangle + \hat{q}^{\text{sm},B}_{\widehat{\text{Sp}}(\boldsymbol{\Sigma})}(\alpha)\ \|\boldsymbol{A}\|_{S_1}\Big]\right\}$$
$$= \left\{\boldsymbol{\Sigma}\,\middle|\big\|\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}\big\|_{\text{op}} \le \hat{q}^{\text{sm},B}_{\widehat{\text{Sp}}(\boldsymbol{\Sigma})}(\alpha)\right\}.$$

We thus have

$$\sup_{0\le\alpha\le1}\bigg|\mathbb{P}\bigg(\forall \boldsymbol{A}\in\mathbb{R}^{p\times p},\ \langle \boldsymbol{A},\boldsymbol{\Sigma}\rangle \in \bigg[\langle \boldsymbol{A},\hat{\boldsymbol{\Sigma}}\rangle - \hat{q}^{\mathrm{sm},B}_{\widehat{\mathrm{Sp}}(\boldsymbol{\Sigma})}(\alpha)\ \|\boldsymbol{A}\|_{S_1}$$

$$\langle \boldsymbol{A},\hat{\boldsymbol{\Sigma}}\rangle + \hat{q}^{\mathrm{sm},B}_{\widehat{\mathrm{Sp}}(\boldsymbol{\Sigma})}(\alpha)\ \|\boldsymbol{A}\|_{S_1}\bigg]\bigg|\boldsymbol{Y}^1,\cdots,\boldsymbol{Y}^B\bigg) - (1-\alpha)\bigg|$$

$$= \sup_{0\le\alpha\le1}\bigg|\mathbb{P}\bigg(\big\|\boldsymbol{\Sigma}-\hat{\boldsymbol{\Sigma}}\big\|_{\mathrm{op}} \le \hat{q}^{\mathrm{sm},B}_{\widehat{\mathrm{Sp}}(\boldsymbol{\Sigma})}(\alpha)\bigg) - (1-\alpha)\bigg| \le \rho_n(\boldsymbol{\Sigma}_0) + \sup_{0\le\alpha\le1}\bigg|\mathbb{P}\bigg(\tilde{T}^{\mathrm{sm}} \le \hat{q}^{\mathrm{sm},B}_{\widehat{\mathrm{Sp}}(\boldsymbol{\Sigma})}(\alpha)\bigg) - (1-\alpha)\bigg|.$$

Similar to Lemma 5.3, we now provide another lemma to bound the second term on the right side.

Lemma 6.1. *For random variable X with distribution function F_1 and i.i.d. continuous random variables $X_1, X_2, \cdots$ with distribution function F_2 independent of X. If we define the sample upper α-quantile for $X_1, X_2, \cdots, X_B$ as $\hat{q}_X^B(\alpha)$, we have with probability approaching* 1,

$$\sup_{0\le\alpha\le1}\bigg|\mathbb{P}\bigg(X \le \hat{q}^B(\alpha)\big|X_1, X_2,\cdots,X_B\bigg) - (1-\alpha)\bigg| \lesssim d_K(F_1,F_2) + B^{-\frac{1}{2}},$$

where $d_K(F_1, F_2)$ is the Kolmogorov distance between F_1 and F_2.

Proof. Define $g = F_1^{-1}\circ F_2$ and $X_b' = g(X_b)$ for $b = 1, \cdots$. We then have i.i.d. random variable $X, X_1', X_2', \cdots$. By Lemma 5.3, we have with probability approaching 1,

$$\sup_{0\le\alpha\le1}\bigg|\mathbb{P}\bigg(X \le g^{-1}(\hat{q}_X^B(\alpha))\big|X_1, X_2,\cdots,X_B\bigg) - (1-\alpha)\bigg| \lesssim B^{-\frac{1}{2}}.$$

Notice that

$$\mathbb{P}\bigg(X \le g^{-1}(\hat{q}_X^B(\alpha))\big|X_1, X_2,\cdots,X_B\bigg) = F_1\circ F_2^{-1}\circ F_1(\hat{q}_X^B(\alpha)),$$

$$\mathbb{P}\bigg(X \le \hat{q}_X^B(\alpha))\big|X_1, X_2,\cdots,X_B\bigg) = F_1(\hat{q}_X^B(\alpha)),$$

and that

$$\big|F_1\circ F_2^{-1}\circ F_1(\hat{q}_X^B(\alpha)) - F_1(\hat{q}_X^B(\alpha))\big| \lesssim \sup_{x\in[0,1]}\big|F_1\circ F_2^{-1}(x) - x\big| \le d_{\mathrm{K}}(F_1,F_2),$$

we thus have with probability approaching 1,

$$\sup_{0\leq\alpha\leq 1}\left|\mathbb{P}\Big(X\leq \hat{q}_X^B(\alpha)\big|X_1,X_2,\cdots,X_B\Big)-(1-\alpha)\right|\lesssim d_{\mathrm{K}}(F_1,F_2)+B^{-\frac{1}{2}}.$$

This completes the proof. $\square$

We define $F^{\mathrm{sm}}_{\boldsymbol{\Sigma}}$ as the distribution function of T^{sm} when $\boldsymbol{Y}_1,\cdots,\boldsymbol{Y}_B$ are generated with rows following the distribution $\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})$. We then have

$$d_{\mathrm{K}}(F^{\mathrm{sm}}_{\boldsymbol{\Sigma}_1},F^{\mathrm{sm}}_{\boldsymbol{\Sigma}_2})\lesssim d_{\mathrm{Sp}}(\mathrm{Sp}(\boldsymbol{\Sigma}_1),\mathrm{Sp}(\boldsymbol{\Sigma}_2)),$$

where $d^2_{\mathrm{Sp}}(\mathrm{Sp}(\boldsymbol{\Sigma}_1),\mathrm{Sp}(\boldsymbol{\Sigma}_2))=\mathrm{tr}(\mathrm{Sp}(\boldsymbol{\Sigma}_1)-\mathrm{Sp}(\boldsymbol{\Sigma}_2))^2/p$ is defined in the main text. Apply Lemma 6.1 we have

$$\sup_{0\leq\alpha\leq 1}\left|\mathbb{P}\Big(\tilde{T}^{\mathrm{sm}}\leq \hat{q}^{\mathrm{sm},B}_{\widehat{\mathrm{Sp}}(\boldsymbol{\Sigma})}(\alpha)\Big)-(1-\alpha)\right|\lesssim d_{\mathrm{Sp}}(\widehat{\mathrm{Sp}}(\boldsymbol{\Sigma}),\mathrm{Sp}(\boldsymbol{\Sigma}))+B^{-\frac{1}{2}}.$$

7 Proof of Theorem 3.4

To prove the universality of

$$T^{\mathrm{ExS}}=f\left(\left\{\left(\sigma_k\left(\boldsymbol{\Sigma}_{2,1}^{-\frac{1}{2}}(\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_{1,1})\boldsymbol{\Sigma}_{2,1}^{-\frac{1}{2}}\right)\right)_{k=1}^{k_1},\cdots,\left(\sigma_k\left(\boldsymbol{\Sigma}_{2,M}^{-\frac{1}{2}}(\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_{1,M})\boldsymbol{\Sigma}_{2,M}^{-\frac{1}{2}}\right)\right)_{k=1}^{k_M}\right\}\right)$$

with a general function f, matrices $(\boldsymbol{\Sigma}_{1,m},\boldsymbol{\Sigma}_{2,m})$ and integer numbers k_m for $m=1,\cdots,M$. Define $\boldsymbol{M}'_{n,m}=\phi^{-1/2}\boldsymbol{\Sigma}_{2,m}^{-1/2}(\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_{1,m})\boldsymbol{\Sigma}_{2,m}^{-1/2}$ for $m_{p,n,m}(z)$ for $m=1,\cdots,M$, and the largest edge $E_{+,m}$ for $\boldsymbol{M}'_{n,m}$. We will prove a generalized Green function comparison theorem similar to Lemma 3.3.

Lemma 7.1 (Generalized Green function comparison)**.** *Under Assumptions 1, 2, 3, 4, consider the function* $F:\mathbb{R}^K\to\mathbb{R}$ *where* $K=\sum_{m=1}^M k_m$*, such that*

$$\sup_{x\in\mathbb{R}^K}|F^{(k)}(x)|(1+|x|)^{C_1}\leq C_1, k=1,2,3,4,$$

for some $C_1 > 0$. *For* $\epsilon > 0$ *small enough and constants* $E_{1,k_1} < E_{1,k_1-1} < \cdots < E_{1,0}$, $\cdots$, $E_{M,k_M} < E_{M,k_M-1} < \cdots < E_{M,0}$,

$$|E_{1,k_1} - E_{+,1}|, \cdots, |E_{1,0} - E_{+,1}|, \cdots, |E_{M,k_M} - E_{+,M}|, \cdots, |E_{M,0} - E_{+,M}| \le N^{-\frac{2}{3}+\epsilon},$$

where $N = \min\{n, p\}$. *Set* $\eta \le N^{-2/3-\epsilon}$, *we have*

(i) when $\gamma \le 1$,

$$\begin{aligned}
&\Bigg|\mathbb{E}\Bigg[F\Bigg(p\int_{E_{1,1}}^{E_{1,0}} \Im m_{p,1}(y+i\eta)\mathrm{d}y, \cdots p\int_{E_{1,k_1}}^{E_{1,0}} \Im m_{p,1}(y+i\eta)\mathrm{d}y, \cdots,\\
&p\int_{E_{M,1}}^{E_{M,0}} \Im m_{p,M}(y+i\eta)\mathrm{d}y, \cdots p\int_{E_{M,k_M}}^{E_{M,0}} \Im m_{p,M}(y+i\eta)\mathrm{d}y\Bigg)\Bigg]\\
&-\mathbb{E}^{Gau}\Bigg[F\Bigg(p\int_{E_{1,1}}^{E_{1,0}} \Im m_{p,1}(y+i\eta)\mathrm{d}y, \cdots p\int_{E_{1,k_1}}^{E_{1,0}} \Im m_{p,1}(y+i\eta)\mathrm{d}y, \cdots,\\
&p\int_{E_{M,1}}^{E_{M,0}} \Im m_{p,M}(y+i\eta)\mathrm{d}y, \cdots p\int_{E_{M,k_M}}^{E_{M,0}} \Im m_{p,M}(y+i\eta)\mathrm{d}y\Bigg)\Bigg]\Bigg|\\
&\le Cp^{-\frac{1}{6}+C\epsilon},
\end{aligned}$$

for some constant $C > 0$, *where* $m_{p,m}(z)$ *is the Stieltjes transform* $\mathbf{\Sigma}_{2,m}^{-1/2}(\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}_{1,m})\mathbf{\Sigma}_{2,m}^{-1/2}$ *for* $m = 1, \cdots, M$. *Similar results hold if we replace* $E_{+,m}$ *by* $E_{-,m}$.

(ii) when $\gamma > 1$,

$$\begin{aligned}
&\Bigg|\mathbb{E}\Bigg[F\Bigg(n\int_{E_{1,1}}^{E_{1,0}} \Im m_{p,n,1}(y+i\eta)\mathrm{d}y, \cdots n\int_{E_{1,k_1}}^{E_{1,0}} \Im m_{p,n,1}(y+i\eta)\mathrm{d}y, \cdots,\\
&n\int_{E_{M,1}}^{E_{M,0}} \Im m_{p,n,M}(y+i\eta)\mathrm{d}y, \cdots n\int_{E_{M,k_M}}^{E_{M,0}} \Im m_{p,n,M}(y+i\eta)\mathrm{d}y\Bigg)\Bigg]\\
&-\mathbb{E}^{Gau}\Bigg[F\Bigg(n\int_{E_{1,1}}^{E_{1,0}} \Im m_{p,n,1}(y+i\eta)\mathrm{d}y, \cdots n\int_{E_{1,k_1}}^{E_{1,0}} \Im m_{p,n,1}(y+i\eta)\mathrm{d}y, \cdots,\\
&n\int_{E_{M,1}}^{E_{M,0}} \Im m_{p,n,M}(y+i\eta)\mathrm{d}y, \cdots n\int_{E_{M,k_M}}^{E_{M,0}} \Im m_{p,n,M}(y+i\eta)\mathrm{d}y\Bigg)\Bigg]\Bigg|\\
&\le Cn^{-\frac{1}{6}+C\epsilon},
\end{aligned}$$

for some constant $C > 0$, *where* $m_{p,n,m}(z)$ *is the Stieltjes transform of the first* n *largest eigenvalues of* $\mathbf{\Sigma}_{2,m}^{-1/2}(\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}_{1,m})\mathbf{\Sigma}_{2,m}^{-1/2}$ *for* $m = 1, \cdots, M$, *defined similar to (37).*

Proof. The proof is similar to Lemma 3.3. The main inputs are the averaged local law 7 for $m_{p,m}(z)$ and 3.2 for $m_{p,n,m}(z)$ for $m = 1, \cdots, M$. These averaged local laws still hold by definition. Details of the proof can be found similarly in Erdős et al. (2012). □

This immediately leads to the following lemma.

Lemma 7.2 (Generalized Gaussian approximation bound for the largest eigenvalue)**.** *Under Assumptions 1, 2, 3, 4, we have for any small $\epsilon > 0$, there exists a constant C for large enough $N = \min\{n, p\}$ and $t \in \mathbb{R}$,*

$$
\begin{aligned}
&\mathbb{P}\Big(N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}'_{n,1}) - E_{+,1}) \le t - N^{-\epsilon}, \cdots, N^{\frac{2}{3}}(\lambda_{k_1}(\boldsymbol{M}'_{n,1}) - E_{+,1}) \le t - N^{-\epsilon} \\
&\quad \cdots, N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}'_{n,M}) - E_{+,M}) \le t - N^{-\epsilon}, \cdots, N^{\frac{2}{3}}(\lambda_{k_M}(\boldsymbol{M}'_{n,M}) - E_{+,M}) \le t - N^{-\epsilon}\Big) - N^{-\frac{1}{6}+C\epsilon} \\
&\le \mathbb{P}^{Gau}\Big(N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}'_{n,1}) - E_{+,1}) \le t, \cdots, N^{\frac{2}{3}}(\lambda_{k_1}(\boldsymbol{M}'_{n,1}) - E_{+,1}) \le t \\
&\quad \cdots, N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}'_{n,M}) - E_{+,M}) \le t, \cdots, N^{\frac{2}{3}}(\lambda_{k_M}(\boldsymbol{M}'_{n,M}) - E_{+,M}) \le t\Big) \\
&\quad \le \mathbb{P}\Big(N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}'_{n,1}) - E_{+,1}) \le t + N^{-\epsilon}, \cdots, N^{\frac{2}{3}}(\lambda_{k_1}(\boldsymbol{M}'_{n,1}) - E_{+,1}) \le t + N^{-\epsilon} \\
&\qquad \cdots, N^{\frac{2}{3}}(\lambda_1(\boldsymbol{M}'_{n,M}) - E_{+,M}) \le t + N^{-\epsilon}, \cdots, N^{\frac{2}{3}}(\lambda_{k_M}(\boldsymbol{M}'_{n,M}) - E_{+,M}) \le t + n^{-\epsilon}\Big) - N^{-\frac{1}{6}+C\epsilon}.
\end{aligned}
$$

When $\gamma \le 1$, similar results also hold for $\lambda_{p-k}(\boldsymbol{M}'_{n,m})$ for $k = 1, \cdots, k_m$ and $m = 1, \cdots, M$.

Proof. The proof is similar to Theorem 3.1. □

To finish the proof, we proceed to first show the rigidity of $\boldsymbol{M}'_{n,1}, \cdots, \boldsymbol{M}'_{n,M}$, which can be derived using Theorem 3.2. This combined with the average local law leads to an anti-concentration inequality for $\boldsymbol{M}'_{n,1}, \cdots, \boldsymbol{M}'_{n,M}$. With these tools, we can prove Theorem 3.4 in the same way as the proof of Lemma 5.2

8 Proof of Theorem 3.5

In this subsection, we perform a power analysis under the generalized spike model setting across a family of statistics, with T and T^{Roy} as specific cases. Recall the following generalized spike structure

$$\boldsymbol{\Sigma} = \boldsymbol{U} \begin{bmatrix} \boldsymbol{D} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{V}_2 \end{bmatrix} \boldsymbol{U}^T, \ \boldsymbol{\Sigma}_0 = \boldsymbol{U} \begin{bmatrix} \boldsymbol{V}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{V}_2 \end{bmatrix} \boldsymbol{U}^T, \ \boldsymbol{\Sigma}_1 = \boldsymbol{U} \begin{bmatrix} \boldsymbol{R}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{R}_2 \end{bmatrix} \boldsymbol{U}^T, \tag{43}$$

where $\boldsymbol{D}, \boldsymbol{V}_1, \boldsymbol{R}_1 \in \mathbb{R}^{k\times k}$, $\boldsymbol{V}_2, \boldsymbol{R}_2 \in \mathbb{R}^{(p-k)\times(p-k)}$ are diagonal matrices, k is a fixed number, and $\boldsymbol{U}$ is orthogonal matrix. We consider the family of statistics $T(\boldsymbol{\Sigma}_1) = \left\| \boldsymbol{\Sigma}_1^{-\frac{1}{2}} \left(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0 \right) \boldsymbol{\Sigma}_1^{-\frac{1}{2}} \right\|_{\mathrm{op}}$, and reject H_0 if $T(\boldsymbol{\Sigma}_1) \geq \hat{q}^{\mathrm{sm},B}_{\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1^{-1/2}, \boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1^{-1/2}}(\alpha)$. Notably, $T = T(\boldsymbol{I}_p)$, $T^{\mathrm{Roy}} = T(\boldsymbol{\Sigma}_0)$ are both included in this family. Define $\boldsymbol{D}' = \boldsymbol{D}\boldsymbol{R}_1^{-1}$, $\boldsymbol{V}_1' = \boldsymbol{V}_1\boldsymbol{R}_1^{-1}$, $\boldsymbol{V}_2' = \boldsymbol{V}_2\boldsymbol{R}_2^{-1}$. We assume that $\|\boldsymbol{D}'\|_{\mathrm{op}}$, $\|\boldsymbol{V}_1'\|_{\mathrm{op}}$, $\|\boldsymbol{V}_2'\|_{\mathrm{op}}$ are bounded. Define $E_+(\boldsymbol{\Sigma}_1)$, $E_-(\boldsymbol{\Sigma}_1)$ and $\tilde{m}(z;\boldsymbol{\Sigma}_1)$ with $(\boldsymbol{\Sigma}, \boldsymbol{R}) = (\boldsymbol{V}_2', -\boldsymbol{V}_2')$ in $\boldsymbol{M}_n$. For simplicity, we only consider $\gamma = 1$.

Theorem 8.1. *Under Assumption 1, define* $\boldsymbol{D}' = \mathrm{diag}\left((d_i')_{i=1}^k\right)$, $\boldsymbol{V}_1' = \mathrm{diag}\left((v_i')_{i=1}^k\right)$ *and*

$$\kappa = -\frac{\phi^{\frac{1}{2}}}{\tilde{m}(\phi^{-\frac{1}{2}}E_+(\boldsymbol{\Sigma}_1);\boldsymbol{\Sigma}_1)} - \frac{\phi^{\frac{1}{2}}v_i'}{E_+(\boldsymbol{\Sigma}_1)\tilde{m}(\phi^{-\frac{1}{2}}E_+(\boldsymbol{\Sigma}_1);\boldsymbol{\Sigma}_1)}. \tag{44}$$

(i) If $d_i' > \kappa$ for some $1 \leq i \leq k$, then the power goes to 1, i.e.

$$\mathbb{P}\left(T(\boldsymbol{\Sigma}_1) \geq \hat{q}^{sm,B}_{\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1^{-1/2}, \boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1^{-1/2}}(\alpha) \right) \to 1, \ \text{as } n\to\infty, \ B\to\infty.$$

(ii) If $d_i' < \kappa$ for all $1 \leq i \leq k$ and $T(\boldsymbol{\Sigma}_1) = \lambda_1\left(\boldsymbol{\Sigma}_1^{-1/2}\left(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\right)\boldsymbol{\Sigma}_1^{-1/2}\right)$, then

$$\mathbb{P}\left(T(\boldsymbol{\Sigma}_1) \geq \hat{q}^{sm,B}_{\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1^{-1/2}, \boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1^{-1/2}}(\alpha) \right) \to \alpha, \ \text{as } n\to\infty, \ B\to\infty,$$

i.e. there is no power under this setting.

Theorem 8.1 provides the power analysis of the test based on the statistic $T(\boldsymbol{\Sigma}_1)$. Specifically, it identifies the sharp phase transition point κ in (45), which closely resembles the well-known Baik-Ben Arous-Péché (BBP) phase transition for the largest eigenvalues of the sample covariance matrix, see Baik et al. (2005), Baik & Silverstein (2006), Paul (2007). When the spike eigenvalues exceed this phase transition point, they lie outside the bulk eigenvalue support, yielding full test power. Conversely, if the spike eigenvalues d_i' fall below this point, distinguishing them from the null case using $T(\boldsymbol{\Sigma}_1)$ becomes infeasible. Theorem 8.1 discusses the phase transition for the spikes of the right endpoint. The same analysis extends to more general spiked covariance models with multiple bulk components, as studied in Ding (2021). In such settings, analogous phase transitions occur for spikes associated with different bulks. Corresponding results and proofs are provided in Theorem 8.2. Notice also that Theorem 8.1 directly gives the results of Theorem 3.5 in the main text, so we focus on the proof of Theorem 8.1 from now on.

We define $\boldsymbol{T} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}_1^{-1/2}$, $\boldsymbol{\Sigma}_0' = \boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1^{-1/2}$, and

$$\boldsymbol{M}_n = \frac{1}{n}\boldsymbol{T}^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{T} - \boldsymbol{\Sigma}_0',$$

so that $T(\boldsymbol{\Sigma}_1) = \|\boldsymbol{M}_n\|_{\mathrm{op}}$. Under this notation, we have

$$\boldsymbol{T}^T\boldsymbol{T} = \boldsymbol{U}\begin{bmatrix} \boldsymbol{D}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}_2' \end{bmatrix}\boldsymbol{U}^T,\ \boldsymbol{T} = \boldsymbol{W}\begin{bmatrix} (\boldsymbol{D}')^{\frac{1}{2}} & \boldsymbol{0} \\ \boldsymbol{0} & (\boldsymbol{V}_2')^{\frac{1}{2}} \end{bmatrix}\boldsymbol{U}^T,\ \boldsymbol{\Sigma}_0' = \boldsymbol{U}\begin{bmatrix} \boldsymbol{V}_1' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}_2' \end{bmatrix}\boldsymbol{U}^T,$$

$\boldsymbol{U}$ and $\boldsymbol{W}$ are orthogonal matrices. Accordingly, we define

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{W}_1, & \boldsymbol{W}_2 \end{bmatrix}$$

where $\boldsymbol{W}_1 \in \mathbb{R}^{p\times k}$, $\boldsymbol{W}_2 \in \mathbb{R}^{p\times(n-k)}$.

Notice that the eigenvalues of $\boldsymbol{M}_n$ satisfy the equation

$$\det(\lambda\boldsymbol{I}_p + \boldsymbol{\Sigma}_0' - \frac{1}{n}\boldsymbol{T}^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{T}) = 0.$$

We seek for eigenvalues λ larger than E_+ or smaller than E_-. When λ is larger than E_+ or smaller than E_-, we have $\det(\lambda \boldsymbol{I}_p + \boldsymbol{\Sigma}_0') \neq 0$. We thus have

$$\det(\boldsymbol{I}_p - \frac{1}{n}(\lambda \boldsymbol{I}_p + \boldsymbol{\Sigma}_0')^{-\frac{1}{2}} \boldsymbol{T}^T \boldsymbol{Z}^T \boldsymbol{Z} \boldsymbol{T} (\lambda \boldsymbol{I}_p + \boldsymbol{\Sigma}_0')^{-\frac{1}{2}}) = 0.$$

Notice that

$$\begin{aligned}
&\det(\boldsymbol{I}_p - \frac{1}{n}(\lambda \boldsymbol{I}_p + \boldsymbol{\Sigma}_0')^{-\frac{1}{2}} \boldsymbol{T}^T \boldsymbol{Z}^T \boldsymbol{Z} \boldsymbol{T} (\lambda \boldsymbol{I}_p + \boldsymbol{\Sigma}_0')^{-\frac{1}{2}}) \\
=&\det(\boldsymbol{I}_p - \frac{1}{n}\boldsymbol{U} \begin{bmatrix} ((\lambda \boldsymbol{I}_p + \boldsymbol{V}_1')^{-1}\boldsymbol{D}')^{\frac{1}{2}} & \boldsymbol{0} \\ \boldsymbol{0} & ((\lambda \boldsymbol{I}_p + \boldsymbol{V}_2')^{-1}\boldsymbol{V}_2')^{\frac{1}{2}} \end{bmatrix} \boldsymbol{W}^T \boldsymbol{Z}^T \boldsymbol{Z} \boldsymbol{W} \\
&\cdot \begin{bmatrix} ((\lambda \boldsymbol{I}_p + \boldsymbol{V}_1')^{-1}\boldsymbol{D}')^{\frac{1}{2}} & \boldsymbol{0} \\ \boldsymbol{0} & ((\lambda \boldsymbol{I}_p + \boldsymbol{V}_2')^{-1}\boldsymbol{V}_2')^{\frac{1}{2}} \end{bmatrix} \boldsymbol{U}^T) \\
=&\det(\boldsymbol{I}_p - \frac{1}{n} \begin{bmatrix} ((\lambda \boldsymbol{I}_p + \boldsymbol{V}_1')^{-1}\boldsymbol{D}')^{\frac{1}{2}} & \boldsymbol{0} \\ \boldsymbol{0} & ((\lambda \boldsymbol{I}_p + \boldsymbol{V}_2')^{-1}\boldsymbol{V}_2')^{\frac{1}{2}} \end{bmatrix} \boldsymbol{W}^T \boldsymbol{Z}^T \boldsymbol{Z} \boldsymbol{W} \\
&\cdot \begin{bmatrix} ((\lambda \boldsymbol{I}_p + \boldsymbol{V}_1')^{-1}\boldsymbol{D}')^{\frac{1}{2}} & \boldsymbol{0} \\ \boldsymbol{0} & ((\lambda \boldsymbol{I}_p + \boldsymbol{V}_2')^{-1}\boldsymbol{V}_2')^{\frac{1}{2}} \end{bmatrix}) \\
=&\det(\begin{bmatrix} \boldsymbol{I}_k - \frac{1}{n}\boldsymbol{A}^{\frac{1}{2}} \boldsymbol{W}_1^T \boldsymbol{Z}^T \boldsymbol{Z} \boldsymbol{W}_1 \boldsymbol{A}^{\frac{1}{2}} & -\frac{1}{n}\boldsymbol{A}^{\frac{1}{2}} \boldsymbol{W}_1^T \boldsymbol{Z}^T \boldsymbol{Z} \boldsymbol{W}_2 \boldsymbol{B}^{\frac{1}{2}} \\ -\frac{1}{n}\boldsymbol{B}^{\frac{1}{2}} \boldsymbol{W}_2^T \boldsymbol{Z}^T \boldsymbol{Z} \boldsymbol{W}_1 \boldsymbol{A}^{\frac{1}{2}} & \boldsymbol{I}_{p-k} - \frac{1}{n}\boldsymbol{B}^{\frac{1}{2}} \boldsymbol{W}_2^T \boldsymbol{Z}^T \boldsymbol{Z} \boldsymbol{W}_2 \boldsymbol{B}^{\frac{1}{2}} \end{bmatrix}),
\end{aligned}$$

where

$$\boldsymbol{A} = (\lambda \boldsymbol{I}_p + \boldsymbol{V}_1')^{-1}\boldsymbol{D}', \ \boldsymbol{B} = (\lambda \boldsymbol{I}_p + \boldsymbol{V}_2')^{-1}\boldsymbol{V}_2',$$

Since λ is larger than E_+ or smaller than E_-, we have

$$\det(\boldsymbol{I}_{p-k} - \frac{1}{n}\boldsymbol{B}^{\frac{1}{2}} \boldsymbol{W}_2^T \boldsymbol{Z}^T \boldsymbol{Z} \boldsymbol{W}_2 \boldsymbol{B}^{\frac{1}{2}}) \neq 0.$$

This combined with Schur's complement formula gives

$$\det(\boldsymbol{I}_k - \frac{1}{n}\boldsymbol{A}^{\frac{1}{2}} \boldsymbol{W}_1^T \boldsymbol{Z}^T \boldsymbol{Z} \boldsymbol{W}_1 \boldsymbol{A}^{\frac{1}{2}} - \frac{1}{n}\boldsymbol{A}^{\frac{1}{2}} \boldsymbol{W}_1^T \boldsymbol{Z}^T \boldsymbol{Z} \boldsymbol{W}_2 \boldsymbol{B}^{\frac{1}{2}} \cdot \boldsymbol{M}_{n-k}^{-1} \cdot \frac{1}{n}\boldsymbol{B}^{\frac{1}{2}} \boldsymbol{W}_2^T \boldsymbol{Z}^T \boldsymbol{Z} \boldsymbol{W}_1 \boldsymbol{A}^{\frac{1}{2}}) = 0,$$

where

$$\boldsymbol{M}_{n-k} = \boldsymbol{I}_{p-k} - \frac{1}{n}\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{W}_2^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{W}_2\boldsymbol{B}^{\frac{1}{2}}.$$

Using

$$\frac{1}{\sqrt{n}}\boldsymbol{Z}\boldsymbol{W}_2\boldsymbol{B}^{\frac{1}{2}} \cdot \boldsymbol{M}_{n-k}^{-1} \cdot \frac{1}{\sqrt{n}}\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{W}_2^T\boldsymbol{Z}^T = (\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{Z}\boldsymbol{W}_2\boldsymbol{B}\boldsymbol{W}_2^T\boldsymbol{Z}^T)^{-1} - \boldsymbol{I}_n,$$

we obtain

$$\det(\boldsymbol{I}_k - \frac{1}{n}\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{W}_1^T\boldsymbol{Z}^T(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{Z}\boldsymbol{W}_2\boldsymbol{B}\boldsymbol{W}_2^T\boldsymbol{Z}^T)^{-1}\boldsymbol{Z}\boldsymbol{W}_1\boldsymbol{A}^{\frac{1}{2}}) = 0.$$

Following Jiang & Bai (2021), we define

$$\Omega(\lambda) = \frac{1}{n}\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{W}_1^T\boldsymbol{Z}^T(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{Z}\boldsymbol{W}_2\boldsymbol{B}\boldsymbol{W}_2^T\boldsymbol{Z}^T)^{-1}\boldsymbol{Z}\boldsymbol{W}_1\boldsymbol{A}^{\frac{1}{2}} - \frac{1}{n}\mathrm{tr}(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{Z}\boldsymbol{W}_2\boldsymbol{B}\boldsymbol{W}_2^T\boldsymbol{Z}^T)^{-1}\boldsymbol{A},$$

we have

$$\det(\boldsymbol{I}_k - \frac{1}{n}\mathrm{tr}(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{Z}\boldsymbol{W}_2\boldsymbol{B}\boldsymbol{W}_2^T\boldsymbol{Z}^T)^{-1}\boldsymbol{A} - \Omega(\lambda)) = 0.$$

Following results of Jiang & Bai (2021), we can show that $\Omega(\lambda) \to 0$ in probability. This gives

$$\det(\boldsymbol{I}_k - \frac{1}{n}\mathrm{tr}(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{Z}\boldsymbol{W}_2\boldsymbol{B}\boldsymbol{W}_2^T\boldsymbol{Z}^T)^{-1}\boldsymbol{A}) = o_P(1).$$

Notice that

$$\frac{1}{n}\mathrm{tr}(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{Z}\boldsymbol{W}_2\boldsymbol{B}\boldsymbol{W}_2^T\boldsymbol{Z}^T)^{-1} = -\phi^{-\frac{1}{2}}\lambda\tilde{m}(\phi^{-\frac{1}{2}}\lambda; \boldsymbol{\Sigma}_1) + o_P(1),$$

the definition of the diagonal matrix $\boldsymbol{A}$ gives

$$1 + \phi^{-\frac{1}{2}}\lambda_i\tilde{m}(\phi^{-\frac{1}{2}}\lambda_i; \boldsymbol{\Sigma}_1)\frac{d_i'}{\lambda_i + v_i'} = 0,$$

we thus have

$$d_i' = -\frac{\lambda_i + v_i'}{\phi^{-\frac{1}{2}}\lambda_i\tilde{m}(\phi^{-\frac{1}{2}}\lambda_i; \boldsymbol{\Sigma}_1)}.$$

Notice that $\lambda\tilde{m}(\lambda)$ and $\tilde{m}(\lambda)$ are increasing in λ outside the support, the existence of the above equation is that

$$d_i > -\frac{E_+ + v_i}{\phi^{-\frac{1}{2}}E_+\tilde{m}(\phi^{-\frac{1}{2}}E_+)} = -\frac{\phi^{\frac{1}{2}}}{\tilde{m}(\phi^{-\frac{1}{2}}E_+;\boldsymbol{\Sigma}_1)} - \frac{\phi^{\frac{1}{2}}v_i}{E_+\tilde{m}(\phi^{-\frac{1}{2}}E_+;\boldsymbol{\Sigma}_1)}.$$

When $|E_+(\boldsymbol{\Sigma}_1)| > |E_-(\boldsymbol{\Sigma}_1)|$, only the largest eigenvalue can spike. If the above inequality does not hold, we have

$$\lambda_1(\boldsymbol{M}_n) \to E_+,$$

and thus the test does not have power. This finishes the proof.

Remark. The above theorem focuses on the largest spike case. We can also show that a similar result also holds for spikes between bulks when multiple bulks exist. We define the matrix $\boldsymbol{M}_{n,0}$ as $\boldsymbol{M}_n$ under H_0, i.e.

$$\boldsymbol{M}_{n,0} = \frac{1}{n}\boldsymbol{T}_0^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{T}_0 - \boldsymbol{\Sigma}_0',$$

where $\boldsymbol{T}_0 = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_1^{-1/2}$. The limiting spectral distribution ρ of $\boldsymbol{M}_{n,0}$ exists, as shown by Couillet et al. (2011). We assume the support of ρ to be the union of J non-overlapping intervals, i.e. $\mathrm{supp}(\rho) = \bigcup_{j=1}^J I_j$ where $I_j = [a_j, b_j]$ with $a_j \le b_j$. Without loss of generality, we assume $b_j < a_{j+1}$ for $j = 1, \cdots, J-1$. Especially, we have $a_1 = E_-$ and $b_J = E_+$. We define a spike d_i to be the spike of the j-th bulk if the limit eigenvalue of $\boldsymbol{M}_n$ associated with d_i lies between $[b_j, a_{j+1})$. We have the following result.

Theorem 8.2. *Under the assumptions in Theorem 4.1 in the main text. Define the j-th threshold*

$$\kappa_j = -\frac{\phi^{\frac{1}{2}}}{\tilde{m}(\phi^{-\frac{1}{2}}b_j;\boldsymbol{\Sigma}_1)} - \frac{\phi^{\frac{1}{2}}v_i'}{b_j\tilde{m}(\phi^{-\frac{1}{2}}b_j;\boldsymbol{\Sigma}_1)}. \tag{45}$$

Suppose d_i is the spike of the j-th bulk, and denote the eigenvalues associated with d_i as $\lambda_i(\boldsymbol{M}_n)$. Then *(i) If $d_i' > \kappa_j$, we have*

$$\lim_{n\to\infty} \lambda_i(\boldsymbol{M}_n) > b_j.$$

(ii) If $d_i' \leq \kappa_j$, we have

$$\lim_{n\to\infty} \lambda_i(\boldsymbol{M}_n) = b_j.$$

This demonstrates a phase transition phenomenon for spikes among bulks, extending the results for the largest bulk as in Theorem 4.1.

Proof. The proof follows similarly to Theorem 4.1. Denote $\lambda_i = \lambda_i(\boldsymbol{M}_n)$. When $\lim_{n\to\infty} \lambda_i > b_j$, following the proof of Theorem 4.1, we still have

$$d_i' = -\frac{\lambda_i + v_i'}{\phi^{-\frac{1}{2}}\lambda_i \tilde{m}(\phi^{-\frac{1}{2}}\lambda_i; \boldsymbol{\Sigma}_1)}.$$

Using $\lambda\tilde{m}(\lambda)$ and $\tilde{m}(\lambda)$ are increasing in λ outside the support, we have

$$d_i > -\frac{b_j + v_i}{\phi^{-\frac{1}{2}} b_j \tilde{m}(\phi^{-\frac{1}{2}} b_j)} = -\frac{\phi^{\frac{1}{2}}}{\tilde{m}(\phi^{-\frac{1}{2}} b_j; \boldsymbol{\Sigma}_1)} - \frac{\phi^{\frac{1}{2}} v_i}{b_j \tilde{m}(\phi^{-\frac{1}{2}} b_j; \boldsymbol{\Sigma}_1)} = \kappa_j.$$

The monotonicity also provides the other side. This proves the claim. $\square$

9 Proof of Theorem 3.6

We use the notation in Theorem 8.1. Under null, the combined statistics T^{Com} converges to $\frac{E_+^2(\boldsymbol{I}_p)}{\mathrm{tr}^2(\boldsymbol{\Sigma}_0)} + E_+^2(\boldsymbol{\Sigma}_0)$. Without loss of generality, we assume that under H_1, $d_i - v_i > -\frac{\phi^{1/2}}{\tilde{m}(\phi^{-1/2}E_+)} - v_i\left(1 + \frac{\phi^{1/2}}{E_+\tilde{m}(\phi^{-1/2}E_+)}\right)$. In this case, Theorem 8.1 shows that the limit of T is larger than $E_+(\boldsymbol{I}_p)$ while the limit of T^{Roy} is no less then $E_+(\boldsymbol{\Sigma}_0)$. This shows that T^{Com} converges to point larger than $\frac{E_+^2(\boldsymbol{I}_p)}{\mathrm{tr}^2(\boldsymbol{\Sigma}_0)} + E_+^2(\boldsymbol{\Sigma}_0)$, so that the power of T^{Com} goes to 1.

10 Adapting to the general mean vector case

In this section, we extend our analysis to the general mean case, i.e. we allow $\mathbb{E}[X_i] = \boldsymbol{\mu}$ for general $\boldsymbol{\mu}$. In this case, we consider the empirical covariance matrix as

$$\hat{\boldsymbol{\Sigma}}^* = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i^T - \bar{\boldsymbol{X}}\bar{\boldsymbol{X}}^T,$$

where $\bar{\boldsymbol{X}} = \sum_{i=1}^n \boldsymbol{X}_i/n$. By studying $\hat{\boldsymbol{\Sigma}}^*$, we can assume without loss of generality that $\boldsymbol{\mu} = \boldsymbol{0}$ and $\boldsymbol{X}_i = \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}_i$ with $\boldsymbol{Z}_i$ satisfying Assumption 1 in the main text. We accordingly denote $*$ on any quantity depending on $\boldsymbol{Z} = (\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_n)^T$ means that $\boldsymbol{Z}$ has been replaced by $\boldsymbol{Z}^* = (\boldsymbol{I}_n - \boldsymbol{1}_n\boldsymbol{1}_n^T/n)\boldsymbol{Z}$ in its definition, where $\boldsymbol{1}_n$ represents the vector of 1 of length n. For example,

$$\boldsymbol{H}^*(z) = \begin{bmatrix} -\boldsymbol{\Sigma}^{-1} & (\boldsymbol{N}^*)^T \\ \boldsymbol{N}^* & -z\boldsymbol{I}_{n+p} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{\Sigma}^{-1} & (\boldsymbol{R}'')^{\frac{1}{2}} & \frac{1}{(np)^{\frac{1}{4}}}(\boldsymbol{Z}^*)^T \\ (\boldsymbol{R}'')^{\frac{1}{2}} & -z\boldsymbol{I}_p & \boldsymbol{0} \\ \frac{1}{(np)^{\frac{1}{4}}}\boldsymbol{Z}^* & \boldsymbol{0} & -z\boldsymbol{I}_n \end{bmatrix}, \ \boldsymbol{G}^*(z) = (\boldsymbol{H}^*)^{-1}(z).$$

First, notice we have

$$\boldsymbol{H}_p^*(z) = \boldsymbol{H}_p(z) - \frac{1}{\sqrt{np}}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Z}^T(\boldsymbol{1}_n\boldsymbol{1}_n^T/n)\boldsymbol{Z}\boldsymbol{\Sigma}^{\frac{1}{2}},$$

and thus

$$\boldsymbol{Q}_p^*(z) = \boldsymbol{Q}_p(z) + \frac{\boldsymbol{Q}_p(z)\frac{1}{\sqrt{np}}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Z}^T(\boldsymbol{1}_n\boldsymbol{1}_n^T/n)\boldsymbol{Z}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Q}_p(z)}{1 - \frac{1}{\sqrt{n^3p}}\boldsymbol{1}_n^T\boldsymbol{Z}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Q}_p(z)\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Z}^T\boldsymbol{1}_n},$$

i.e.

$$\boldsymbol{G}_p^*(z) = \boldsymbol{G}_p(z) + \frac{\boldsymbol{G}_p(z)\frac{1}{\sqrt{np}}\boldsymbol{Z}^T(\boldsymbol{1}_n\boldsymbol{1}_n^T/n)\boldsymbol{Z}\boldsymbol{G}_p(z)}{z - \frac{1}{\sqrt{n^3p}}\boldsymbol{1}_n^T\boldsymbol{Z}\boldsymbol{G}_p(z)\boldsymbol{Z}^T\boldsymbol{1}_n} = \boldsymbol{G}_p(z) - \frac{\boldsymbol{G}_p(z)\frac{1}{\sqrt{np}}\boldsymbol{Z}^T\boldsymbol{1}_n\boldsymbol{1}_n^T\boldsymbol{Z}\boldsymbol{G}_p(z)}{z^2\boldsymbol{1}_n^T\boldsymbol{G}_{n+p,n}(z)\boldsymbol{1}_n}.$$

Next, the upper-right block of $\boldsymbol{G}^*(z)$ is given by

$$\frac{1}{z}\boldsymbol{G}_p^*(z)(\boldsymbol{N}^*)^T = \frac{1}{z}\boldsymbol{G}_p(z)(\boldsymbol{N}^*)^T - \frac{\boldsymbol{G}_p(z)\frac{1}{\sqrt{np}}\boldsymbol{Z}^T\boldsymbol{1}_n\boldsymbol{1}_n^T\boldsymbol{Z}\boldsymbol{G}_p(z)(\boldsymbol{N}^*)^T}{z^3\boldsymbol{1}_n^T\boldsymbol{G}_{n+p,n}(z)\boldsymbol{1}_n}.$$

Define $\boldsymbol{P} = \operatorname{diag}\{\boldsymbol{I}_p, (\boldsymbol{I}_n - \boldsymbol{1}_n\boldsymbol{1}_n^T/n)\} \in \mathbb{R}^{(n+p)\times(n+p)}$, we have

$$\boldsymbol{N}^* = \begin{bmatrix} (\boldsymbol{R}'')^{1/2} \\ \boldsymbol{Z}^*/(np)^{1/4} \end{bmatrix} = \begin{bmatrix} (\boldsymbol{R}'')^{1/2} \\ (\boldsymbol{I}_n - \boldsymbol{1}_n\boldsymbol{1}_n^T/n)\boldsymbol{Z}/(np)^{1/4} \end{bmatrix} = \boldsymbol{P}\boldsymbol{N},$$

so that

$$\frac{1}{z}\boldsymbol{G}_p^*(z)(\boldsymbol{N}^*)^T = \frac{1}{z}\boldsymbol{G}_p(z)\boldsymbol{N}^T\boldsymbol{P} - \frac{\boldsymbol{G}_p(z)\frac{1}{\sqrt{np}}\boldsymbol{Z}^T\boldsymbol{1}_n\boldsymbol{1}_n^T\boldsymbol{Z}\boldsymbol{G}_p(z)\boldsymbol{N}^T\boldsymbol{P}}{z^3\boldsymbol{1}_n^T\boldsymbol{G}_{n+p,n}(z)\boldsymbol{1}_n}.$$

Finally, for $\boldsymbol{G}_{n+p}^*(z)$ we have

$$\begin{aligned}
\boldsymbol{G}_{n+p}^*(z) &= \frac{1}{z^2}\boldsymbol{N}^*\boldsymbol{G}_p^*(z)(\boldsymbol{N}^*)^T - \frac{1}{z} \\
&= \frac{1}{z^2}\boldsymbol{P}\boldsymbol{N}\boldsymbol{G}_p(z)\boldsymbol{N}^T\boldsymbol{P} - \frac{1}{z} - \frac{1}{z^2}\frac{\boldsymbol{P}\boldsymbol{N}\boldsymbol{G}_p(z)\frac{1}{\sqrt{np}}\boldsymbol{Z}^T\boldsymbol{1}_n\boldsymbol{1}_n^T\boldsymbol{Z}\boldsymbol{G}_p(z)\boldsymbol{N}^T\boldsymbol{P}}{z^2\boldsymbol{1}_n^T\boldsymbol{G}_{n+p,n}(z)\boldsymbol{1}_n} \\
&= \boldsymbol{P}\boldsymbol{G}_{n+p}(z)\boldsymbol{P} - \frac{1}{z}(\boldsymbol{I}_{n+p} - \boldsymbol{P}) - \frac{\boldsymbol{P}\boldsymbol{G}_{n+p}(z)(\boldsymbol{I}_{n+p} - \boldsymbol{P})\boldsymbol{G}_{n+p}(z)\boldsymbol{P}}{\boldsymbol{1}_n^T\boldsymbol{G}_{n+p,n}(z)\boldsymbol{1}_n/n}.
\end{aligned}$$

This together with the anisotropic local law for $\boldsymbol{G}(z)$ in Theorem 2.1 means uniformly in vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{n+p}$ and $z \in \mathcal{D}$, we have

$$\langle \boldsymbol{u}, \boldsymbol{G}_{n+p}^*(z)\boldsymbol{v}\rangle = \langle \boldsymbol{u}, (\boldsymbol{P}\boldsymbol{G}_{n+p}(z)\boldsymbol{P} - \frac{1}{z}(\boldsymbol{I}_{n+p} - \boldsymbol{P}))\boldsymbol{v}\rangle + O_\prec(\Psi^2(z)\|\boldsymbol{u}\|\|\boldsymbol{v}\|). \tag{46}$$

We thus define the deterministic equivalence matrix as

$$\bar{\boldsymbol{G}}^*(z) = \begin{bmatrix} z\boldsymbol{\Sigma}^{\frac12}\bar{\boldsymbol{Q}}_p(z)\boldsymbol{\Sigma}^{\frac12} & \boldsymbol{\Sigma}^{\frac12}\bar{\boldsymbol{Q}}_p(z)\boldsymbol{\Sigma}^{\frac12}(\boldsymbol{R}'')^{\frac12} & \boldsymbol{0} \\ (\boldsymbol{R}'')^{\frac12}\boldsymbol{\Sigma}^{\frac12}\bar{\boldsymbol{Q}}_p(z)\boldsymbol{\Sigma}^{\frac12} & \frac{1}{z}\left((\boldsymbol{R}'')^{\frac12}\boldsymbol{\Sigma}^{\frac12}\bar{\boldsymbol{Q}}_p(z)\boldsymbol{\Sigma}^{\frac12}(\boldsymbol{R}'')^{\frac12} - \boldsymbol{I}_p\right) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \tilde{m}(z)(\boldsymbol{I}_n - \boldsymbol{1}_n\boldsymbol{1}_n^T/n) - \frac{1}{z}\boldsymbol{1}_n\boldsymbol{1}_n^T/n \end{bmatrix}.$$

By our anisotropic local law for $\boldsymbol{G}(z)$ in Theorem 2.1 and the expansion above, we can prove the anisotropic local law for $\boldsymbol{G}^*(z)$

$$\boldsymbol{u}^T\boldsymbol{\Phi}\bar{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{G}^*(z) - \bar{\boldsymbol{G}}^*(z))\bar{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Phi}\boldsymbol{v} = O_\prec(\|\boldsymbol{u}\|\|\boldsymbol{v}\|\Psi(z)),$$

uniformly in vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{n+2p}$ and $z \in \mathcal{D}$. With this, the average local law can be derived accordingly using the anisotropic local law and expansion (46). All of our results for $\hat{\mathbf{\Sigma}} + \boldsymbol{R}$, like the universality of the largest eigenvalues, can then be proved for $\hat{\mathbf{\Sigma}}^* + \boldsymbol{R}$ using the same proof strategy with the anisotropic local law and the average local law as inputs.

11 Additional numerical results

11.1 Simulation

We first present the exact formulations of the seven statistics used in the numerical experiments. Given the zero mean data $\boldsymbol{X}_i = \mathbf{\Sigma}^{1/2}\boldsymbol{Z}_i$ for $i = 1 \cdots, n$, we define the sample covariance matrix as $\hat{\mathbf{\Sigma}} = \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i^T / n$. To test $\mathbf{\Sigma} = \mathbf{\Sigma}_0$, we consider the following statistics. The supremum norm-based statistic T^{sup} (Supn) is defined as

$$T^{\text{sup}} = \max_{i,j} \frac{\theta_{ij}^2}{\sigma_{ij}^2/n},$$

where

$$\theta_{ij} = \frac{1}{n}\sum_{k=1}^n \Big(Z_{ki}Z_{kj} - \delta_{ij} \Big), \ \sigma_{ij} = \frac{1}{n}\sum_{k=1}^n \Big((Z_{ki}Z_{kj} - \delta_{ij}) - \theta_{ij} \Big)^2.$$

The Frobenius norm-based statistic $T^{\text{F},1}$ (Lfn) is defined as

$$T^{\text{F},1} = nW - p - (\nu_4 - 2),$$

where

$$W = \frac{1}{p} tr(\mathbf{\Sigma}_0^{-\frac{1}{2}} \hat{\mathbf{\Sigma}} \mathbf{\Sigma}_0^{-\frac{1}{2}} - \boldsymbol{I}_p)^2 - \frac{1}{np}(tr(\mathbf{\Sigma}_0^{-\frac{1}{2}} \hat{\mathbf{\Sigma}} \mathbf{\Sigma}_0^{-\frac{1}{2}}))^2 + \frac{p}{n}, \ \nu_4 = \frac{1}{np}\sum_{i=1}^n \sum_{j=1}^p Z_{ij}^4.$$

The debiased Frobenius norm-based U statistic $T^{\text{F},2}$ (Ufn) is defined as

$$T^{\text{F},2} = \frac{1}{n(n-1)} \sum_{i \neq j} h(\mathbf{\Sigma}_0^{-\frac{1}{2}} \boldsymbol{X}_i, \mathbf{\Sigma}_0^{-\frac{1}{2}} \boldsymbol{X}_j),$$

where

$$h(\boldsymbol{Z}_i, \boldsymbol{Z}_j) = (\boldsymbol{Z}_i^T \boldsymbol{Z}_j)^2 - (\boldsymbol{Z}_i^T \boldsymbol{Z}_i + \boldsymbol{Z}_j^T \boldsymbol{Z}_j) + p.$$

The proposed operator norm-based statistics T (Opn) is

$$T = \|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\|_{op} = \sigma_1(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0).$$

The normalized statistic T^{Roy} (Roy) is defined as

$$T^{\mathrm{Roy}} = \|\boldsymbol{\Sigma}_0^{-\frac{1}{2}} \hat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}_0^{-\frac{1}{2}} - \boldsymbol{I}_p\|_{op}.$$

The combined statistic T^{Com} (Com)is defined as

$$T^{\mathrm{Com}} = \left(\frac{T}{tr(\boldsymbol{\Sigma}_0)}\right)^2 + (T^{\mathrm{Roy}})^2 = \left(\frac{\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\|_{op}}{tr(\boldsymbol{\Sigma}_0)}\right)^2 + \|\boldsymbol{\Sigma}_0^{-\frac{1}{2}} \hat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}_0^{-\frac{1}{2}} - \boldsymbol{I}_p\|_{op}^2.$$

To evaluate the robustness of our method, we generate n independent, non-identically distributed samples of p-dimensional random vectors with zero means and covariance matrix $\boldsymbol{\Sigma}$. We consider three different structures for the null covariance matrix $\boldsymbol{\Sigma}_0 = (\sigma_{0,ij})_{p\times p}$:

(a). *Exponential decay.* The elements $\sigma_{0,ij} = 0.6^{|i-j|}$.

(b). *Block diagonal.* Let $K = \lfloor p/10 \rfloor$ be the number of blocks, with each block being 1 on the diagonal and 0.55 on the off-diagonal.

(c). *Signed sub-exponential decay.* The elements $\sigma_{0,ij} = (-1)^{i+j} 0.4^{|i-j|^{0.5}}$.

To evaluate the empirical power of these statistics, we consider the setting where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 + \boldsymbol{\Delta}$, with $\boldsymbol{\Delta}$ having a low rank. This setup corresponds to the power analysis presented in Theorem 8.1. As shown in Theorem 8.1, the statistic T outperforms T^{Roy} when the eigenspaces of $\boldsymbol{\Delta}$ align with the eigenvectors of $\boldsymbol{\Sigma}_0$ corresponding to larger eigenvalues. In contrast, T^{Roy} performs better when $\boldsymbol{\Delta}$ aligns with the smaller eigenvectors of $\boldsymbol{\Sigma}_0$. For a fair comparison, we define $\boldsymbol{\Delta} = \sigma \boldsymbol{u}\boldsymbol{u}^T/2 + \sigma \boldsymbol{v}\boldsymbol{v}^T/4$, where $\boldsymbol{u}$ is the fifth-largest eigenvector of $\boldsymbol{\Sigma}_0$, $\boldsymbol{v}$ is uniformly sampled on the unit sphere in $\mathbb{R}^p$, and σ is the signal strength.

We term this configuration the *spike setting*. For completeness, we also conduct a power analysis where $\mathbf{\Delta}$ has full rank by setting $\mathbf{\Delta} = \sigma \boldsymbol{I}_p$, where σ represents the signal level. This configuration, termed the *white noise setting*, represents the covariance structure after adding white noise to the null covariance. Together, the spike and white noise settings cover both low- and full-rank deviations from the null covariance.

We present the table 1-4 of the empirical power of the supremum, the Frobenius, and the operator norm tests at the significance level 0.05 under different sample size n, dimension p, covariance $\mathbf{\Sigma}_0$. Based on our universality results, we only consider the entries with Gaussian distributions. We consider the spike setting as the main text $\mathbf{\Sigma} = \mathbf{\Sigma}_0 + \mathbf{\Delta}$ with $\mathbf{\Delta} = \sigma \boldsymbol{u}\boldsymbol{u}^T/2 + \sigma \boldsymbol{v}\boldsymbol{v}^T/4$, where $\boldsymbol{u}$ is the largest fifth eigenvector of $\mathbf{\Sigma}_0$, $\boldsymbol{v}$ is sampled from a uniform distribution on unit sphere in $\mathbb{R}^p$ and σ is the signal strength. We highlight the largest power in each column.

Table 1: The empirical power of the supremum, the Frobenius, and the operator norm tests with a significance level 0.05, the sample size $n = 100$ for data with Gaussian distribution under the spike alternative and the combinations of the dimension p, the covariance matrix $\mathbf{\Sigma}_0$ and the alternative signal $\sigma = 2$. The largest power is highlighted.

σ	$\mathbf{\Sigma}_0$	**Exp. decay(0.6)**					**Block diagonal**					**Signed subExp. decay(0.4)**				
	p	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**
								n=100								
	Supn	0.169	0.213	0.304	**0.395**	**0.482**	0.182	0.224	0.304	0.400	0.482	0.172	0.224	0.303	0.400	0.477
	Lfn	0.391	0.143	0.071	0.062	0.064	0.293	0.140	0.068	0.061	0.063	0.187	0.107	0.064	0.060	0.060
	Ufn	0.393	0.151	0.070	0.063	0.065	0.297	0.145	0.068	0.061	0.064	0.188	0.110	0.063	0.057	0.060
2	Opn	0.910	0.785	0.456	0.221	0.096	0.940	**0.814**	0.458	0.209	0.105	0.849	0.879	**0.792**	**0.568**	**0.301**
	Roy	0.634	0.212	0.087	0.070	0.059	0.470	0.198	0.085	0.073	0.058	0.291	0.148	0.075	0.066	0.056
	Com	**0.932**	**0.786**	0.439	0.197	0.097	**0.948**	**0.830**	**0.463**	0.220	0.105	0.839	0.867	0.745	0.532	0.267

Table 2: The empirical power of the supremum, the Frobenius, and the operator norm tests with a significance level 0.05, the sample size $n = 100$ for data with Gaussian distribution under the spike alternative and the combinations of the dimension p, the covariance matrix $\mathbf{\Sigma}_0$ and the alternative signal $\sigma = 3$. The largest power is highlighted.

σ	$\mathbf{\Sigma}_0$	**Exp. decay(0.6)**					**Block diagonal**					**Signed subExp. decay(0.4)**				
	p	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**
									n=100							
	Supn	0.166	0.210	0.298	0.393	**0.478**	0.194	0.226	0.309	0.400	0.479	0.168	0.218	0.299	0.401	0.473
	Lfn	0.673	0.342	0.108	0.078	0.069	0.675	0.319	0.112	0.075	0.072	0.477	0.232	0.089	0.065	0.067
	Ufn	0.685	0.338	0.110	0.081	0.073	0.687	0.317	0.113	0.078	0.074	0.478	0.231	0.092	0.068	0.067
3	Opn	0.996	0.979	**0.859**	**0.548**	0.235	0.999	0.993	0.855	0.550	0.247	0.992	**0.996**	**0.976**	**0.906**	**0.691**
	Roy	0.893	0.559	0.159	0.099	0.071	0.884	0.516	0.153	0.091	0.070	0.742	0.363	0.125	0.078	0.065
	Com	**0.998**	**0.982**	0.850	0.528	0.216	**1.000**	**0.995**	**0.867**	**0.575**	**0.254**	**0.993**	0.995	0.970	0.887	0.633

Table 3: The empirical power of the supremum, the Frobenius, and the operator norm tests with a significance level 0.05 for data with Gaussian distribution under the spike alternative and the combinations of the sample size n, the dimension p, the covariance matrix $\mathbf{\Sigma}_0$ and the alternative signal $\sigma = 1$. The largest power is highlighted.

σ	$\mathbf{\Sigma}_0$	**Exp. decay(0.6)**					**Block diagonal**					**Signed subExp. decay(0.4)**				
	p	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**
								n=300								
	Supn	0.068	0.066	0.070	0.071	0.074	0.076	0.066	0.069	0.070	0.075	0.068	0.066	0.070	0.070	0.076
	Lfn	0.270	0.109	0.061	0.055	0.061	0.206	0.104	0.064	0.053	0.058	0.138	0.086	0.059	0.052	0.056
	Ufn	0.269	0.110	0.068	0.053	0.060	0.209	0.105	0.067	0.052	0.060	0.138	0.087	0.058	0.053	0.060
1	Opn	0.877	**0.756**	**0.416**	0.165	0.084	0.928	0.768	**0.382**	0.148	0.079	0.774	0.843	**0.765**	**0.527**	0.249
	Roy	0.435	0.146	0.075	0.067	0.057	0.297	0.133	0.069	0.063	0.058	0.189	0.100	0.070	0.063	0.057
	Com	**0.897**	0.748	0.375	0.159	0.085	**0.936**	**0.774**	0.376	0.147	0.083	0.775	0.840	0.750	0.483	0.215

Table 4: The empirical power of the supremum, the Frobenius, and the operator norm tests with a significance level 0.05, the sample size $n = 300$ for data with Gaussian distribution under the spike alternative and the combinations of the dimension p, the covariance matrix $\mathbf{\Sigma}_0$ and the alternative signal $\sigma = 2$. The largest power is highlighted.

σ	$\mathbf{\Sigma}_0$	**Exp. decay(0.6)**					**Block diagonal**					**Signed subExp. decay(0.4)**				
	p	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**
								n=300								
	Supn	0.090	0.067	0.071	0.066	0.075	0.145	0.093	0.072	0.069	0.076	0.081	0.064	0.071	0.067	0.076
	Lfn	0.882	0.483	0.149	0.087	0.074	0.874	0.439	0.147	0.087	0.075	0.679	0.303	0.111	0.071	0.068
	Ufn	0.889	0.482	0.151	0.086	0.075	0.875	0.436	0.153	0.086	0.073	0.682	0.308	0.114	0.070	0.066
2	Opn	**1.000**	0.999	**0.991**	**0.854**	0.423	**1.000**	**0.999**	**0.987**	0.850	0.425	**1.000**	**1.000**	**0.999**	0.993	**0.938**
	Roy	0.994	0.790	0.200	0.102	0.070	0.990	0.733	0.186	0.101	0.067	0.939	0.550	0.132	0.083	0.063
	Com	**1.000**	**1.000**	0.989	0.843	0.397	**1.000**	**0.999**	0.985	**0.851**	0.423	**1.000**	**1.000**	**0.999**	0.993	0.929

We also present Table 5-10 as the empirical sizes of the supremum, the Frobenius, and the operator norm tests at the significance level 0.05 under different sample size n, dimension p, covariance $\mathbf{\Sigma}_0$ and data generation with the Gaussian and uniform distribution and the Gaussian and t distribution. For each setting, the average number of rejections is calculated based on 2000 replications, and the spectrum matching is conducted with 1000 resampling. The significance level is set to be 0.05 for all the statistics.

Table 5: The empirical sizes of the supremum, the Frobenius, and the operator norm tests with a significance level 0.05, the sample size $n = 100$ for data with Gaussian distribution and the combinations of the dimension p, and the covariance matrix $\mathbf{\Sigma}_0$.

$\mathbf{\Sigma}_0$	**Exp. decay(0.6)**					**Block diagonal**					**Signed subExp. decay(0.4)**				
p	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**
								Gaussian, n=100							
Supn	0.203	0.240	0.316	0.406	0.487	0.203	0.240	0.316	0.406	0.487	0.203	0.240	0.316	0.406	0.487
Lfn	0.049	0.050	0.049	0.050	0.055	0.049	0.050	0.049	0.050	0.055	0.049	0.050	0.049	0.050	0.055
Ufn	0.048	0.056	0.046	0.049	0.056	0.048	0.056	0.046	0.049	0.056	0.048	0.056	0.046	0.049	0.056
Roy	0.050	0.050	0.045	0.053	0.049	0.050	0.050	0.045	0.053	0.049	0.050	0.050	0.045	0.053	0.049
Opn	0.051	0.046	0.048	0.044	0.049	0.048	0.047	0.048	0.047	0.052	0.053	0.055	0.049	0.049	0.052
Com	0.050	0.046	0.046	0.047	0.049	0.050	0.046	0.047	0.045	0.050	0.055	0.054	0.048	0.052	0.051

Table 6: The empirical sizes of the supremum, the Frobenius, and the operator norm tests with a significance level 0.05, the sample size $n = 100$ for data with the t and uniform distribution and the combinations of the dimension p, and the covariance matrix $\mathbf{\Sigma}_0$.

$\mathbf{\Sigma}_0$	**Exp. decay(0.6)**					**Block diagonal**					**Signed subExp. decay(0.4)**				
p	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**
								uniform and t(df=12), n=100							
Supn	0.251	0.293	0.391	0.492	0.613	0.251	0.293	0.391	0.492	0.613	0.251	0.293	0.391	0.492	0.613
Lfn	0.057	0.050	0.053	0.052	0.062	0.057	0.050	0.053	0.052	0.062	0.057	0.050	0.053	0.052	0.062
Ufn	0.056	0.051	0.048	0.051	0.064	0.056	0.051	0.048	0.051	0.064	0.056	0.051	0.048	0.051	0.064
Roy	0.050	0.045	0.051	0.050	0.050	0.050	0.045	0.051	0.050	0.050	0.050	0.051	0.051	0.050	0.050
Opn	0.045	0.033	0.043	0.041	0.047	0.052	0.043	0.042	0.051	0.043	0.047	0.047	0.051	0.060	0.049
Com	0.045	0.033	0.044	0.041	0.050	0.051	0.044	0.041	0.055	0.046	0.049	0.047	0.046	0.063	0.055

Table 7: The empirical sizes of the supremum, the Frobenius, and the operator norm tests with a significance level 0.05, the sample size $n = 100$ for data with the Gaussian and uniform distribution and the combinations of the dimension p, and the covariance matrix $\mathbf{\Sigma}_0$.

$\mathbf{\Sigma}_0$	**Exp. decay(0.6)**					**Block diagonal**					**Signed subExp. decay(0.4)**				
p	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**
							Gaussian and uniform,n=100								
Supn	0.169	0.197	0.258	0.329	0.403	0.169	0.197	0.258	0.329	0.403	0.169	0.197	0.258	0.329	0.403
Lfn	0.059	0.050	0.051	0.055	0.052	0.059	0.050	0.051	0.055	0.052	0.059	0.050	0.051	0.055	0.052
Ufn	0.056	0.048	0.051	0.053	0.053	0.056	0.048	0.051	0.053	0.053	0.056	0.048	0.051	0.053	0.053
Opn	0.040	0.032	0.039	0.047	0.041	0.045	0.038	0.037	0.050	0.045	0.055	0.043	0.050	0.046	0.048
Roy	0.044	0.038	0.031	0.032	0.038	0.044	0.038	0.031	0.032	0.038	0.044	0.038	0.031	0.032	0.038
Com	0.039	0.028	0.035	0.038	0.036	0.046	0.039	0.037	0.054	0.044	0.049	0.040	0.045	0.038	0.045

Table 8: The empirical sizes of the supremum, the Frobenius, and the operator norm tests with a significance level 0.05, the sample size $n = 300$ for data with the Gaussian and uniform distribution and the combinations of the dimension p, and the covariance matrix $\mathbf{\Sigma}_0$.

$\mathbf{\Sigma}_0$	**Exp. decay(0.6)**					**Block diagonal**					**Signed subExp. decay(0.4)**				
p	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**
						Gaussian and uniform,n=300									
Supn	0.072	0.067	0.080	0.092	0.090	0.072	0.067	0.080	0.092	0.090	0.072	0.067	0.080	0.092	0.090
Lfn	0.047	0.045	0.050	0.050	0.043	0.047	0.045	0.050	0.050	0.043	0.047	0.045	0.050	0.050	0.043
Ufn	0.047	0.045	0.052	0.050	0.044	0.047	0.045	0.052	0.050	0.044	0.047	0.045	0.052	0.050	0.044
Opn	0.043	0.048	0.051	0.042	0.039	0.038	0.055	0.054	0.046	0.053	0.052	0.046	0.042	0.047	0.046
Roy	0.036	0.050	0.039	0.042	0.045	0.036	0.050	0.039	0.042	0.045	0.036	0.050	0.039	0.042	0.046
Com	0.042	0.043	0.045	0.035	0.037	0.039	0.056	0.058	0.045	0.046	0.056	0.050	0.043	0.039	0.040

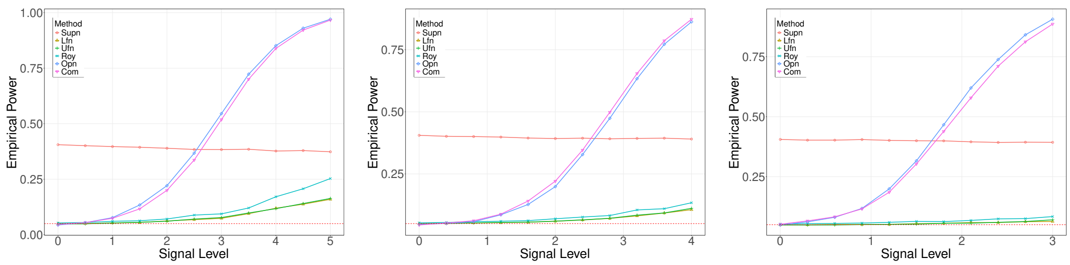
Table 9: The empirical sizes of the supremum, the Frobenius, and the operator norm tests with a significance level 0.05, the sample size $n = 100$ for data with the Gaussian and t distribution and the combinations of the dimension p, and the covariance matrix $\mathbf{\Sigma}_0$.

$\mathbf{\Sigma}_0$	**Exp. decay(0.6)**					**Block diagonal**					**Signed subExp. decay(0.4)**				
p	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**
								Gaussian and t(df=12),n=100							
Supn	0.285	0.324	0.435	0.562	0.676	0.285	0.324	0.435	0.562	0.676	0.285	0.324	0.435	0.562	0.676
Lfn	0.071	0.054	0.052	0.051	0.053	0.071	0.054	0.052	0.051	0.053	0.071	0.054	0.052	0.051	0.053
Ufn	0.066	0.053	0.053	0.050	0.053	0.066	0.053	0.053	0.050	0.053	0.066	0.053	0.053	0.050	0.053
Opn	0.058	0.046	0.055	0.049	0.045	0.062	0.040	0.049	0.057	0.040	0.066	0.050	0.054	0.052	0.062
Roy	0.066	0.060	0.050	0.062	0.052	0.066	0.060	0.050	0.062	0.052	0.066	0.060	0.050	0.062	0.052
Com	0.057	0.046	0.056	0.052	0.058	0.067	0.044	0.054	0.061	0.045	0.066	0.056	0.053	0.056	0.059

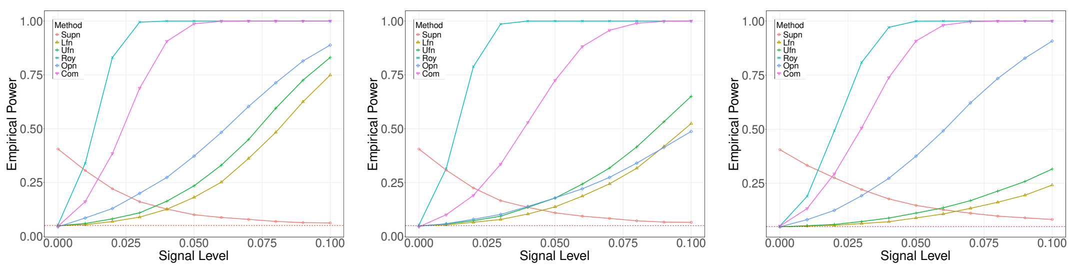
Table 10: The empirical sizes of the supremum, the Frobenius, and the operator norm tests with a significance level 0.05, the sample size $n = 300$ for data with the Gaussian and t distribution and the combinations of the dimension p, and the covariance matrix $\mathbf{\Sigma}_0$.

$\mathbf{\Sigma}_0$	**Exp. decay(0.6)**					**Block diagonal**					**Signed subExp. decay(0.4)**				
p	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**	**100**	**200**	**500**	**1000**	**2000**
						Gaussian and t(df=12),n=300									
SUpn	0.095	0.092	0.097	0.118	0.115	0.095	0.092	0.097	0.118	0.115	0.095	0.092	0.097	0.118	0.115
Lfn	0.047	0.057	0.046	0.046	0.050	0.047	0.057	0.046	0.046	0.050	0.047	0.057	0.046	0.046	0.050
Ufn	0.048	0.055	0.046	0.047	0.055	0.048	0.055	0.046	0.047	0.055	0.048	0.055	0.046	0.047	0.055
Opn	0.048	0.054	0.061	0.047	0.042	0.050	0.053	0.060	0.052	0.052	0.054	0.050	0.048	0.053	0.046
Roy	0.065	0.063	0.056	0.071	0.058	0.065	0.063	0.056	0.071	0.058	0.065	0.063	0.056	0.071	0.058
Com	0.053	0.057	0.057	0.054	0.041	0.052	0.054	0.062	0.053	0.052	0.039	0.054	0.054	0.047	0.050

The following Figure 1 represents empirical powers of the supremum, the Frobenius, and the proposed operator norm tests with respect to the signal level of the spike setting (above) and white noise setting (below) alternatives under three covariance structures, the sample size $n = 100$, dimension $p = 1000$, and the Gaussian data with 2000 replications.



(a) Exp. decay, n=100 (b) Block diagonal, n=100 (c) Signed subExp, n=100



(d) Exp. decay, n=100 (e) Block diagonal, n=100 (f) Signed subExp, n=100

Figure 1: Empirical powers of the supremum, the Frobenius, and the proposed operator norm tests with respect to the signal level of the spike setting (above) and white noise setting (below) alternatives under three covariance structures, the sample size $n = 100$, dimension $p = 1000$, and the Gaussian data with 2000 replications.

11.2 Real data

The HadCRUT4 dataset (Morice et al. 2012) and the Coupled Model Intercomparison Project Phase 5 (CMIP5, Taylor et al. 2012) are provided in `https://figshare.com/articles/dataset/Uncertainty_in_Optimal_Fingerprinting_is_Underestimated/14981241/3?file=28924416`, Li et al. (2021). Following Li et al. (2021), we consider the high-dimensional linear regression

$$\boldsymbol{Y}_k = \sum_{i=1}^{2} \boldsymbol{X}_{k,i}\beta_{k,i} + \boldsymbol{\epsilon}_k,\ \ k = 1, \cdots, 5,$$

where $\boldsymbol{Y}_k \in \mathbb{R}^L$ is the vectorized observed mean temperature across spatial and temporal dimensions for the k-th decades after 1960. The climate models provide N independent simulations $\hat{\boldsymbol{\epsilon}}_{k,i},\ k = 1, \cdots, 5;\ i = 1, \cdots, N$ with $N = 223$, which are assumed to be independent with and follow the same distribution as the unobserved noise $\boldsymbol{\epsilon}_k$. We thus take the Monte Carlo average of the climate model simulation to calculate the covariance $\boldsymbol{\Sigma}_{0,k} = \sum_{i=1}^{N} \hat{\boldsymbol{\epsilon}}_{k,i}\hat{\boldsymbol{\epsilon}}_{k,i}^T/N$ of the unobserved noise $\boldsymbol{\epsilon}_k$.

We construct the data matrix $\tilde{\boldsymbol{X}}_k \in \mathbb{R}^{n\times p}$ by combining quantities $\tilde{\boldsymbol{X}}_{k,1,j} - \bar{\tilde{\boldsymbol{X}}}_{k,1}$ for $j = 1, \ldots, n_1$ and $\tilde{\boldsymbol{X}}_{k,2,j} - \bar{\tilde{\boldsymbol{X}}}_{k,2}$ for $j = 1, \ldots, n_2$, where $n = n_1 + n_2 = 81$ and $p = L = 108$. Here $\bar{\tilde{\boldsymbol{X}}}_{k,i} = \sum_{j=1}^{n_i} \tilde{\boldsymbol{X}}_{k,i,j}/n_i$. We apply several statistics to the data matrix $\tilde{\boldsymbol{X}}_k$ with the hypothesized matrix $\boldsymbol{\Sigma}_{0,k}$: the proposed operator norm-based statistics T (Opn) and T^{Com} (Com), the operator norm-based statistics T^{Roy} (Roy), the Frobenius norm-based statistics $T^{\mathrm{F},1}$ (Lfn) and $T^{\mathrm{F},2}$ (Ufn), and the supremum norm-based statistic $T^{\sup}$ (Supn).

References

Alex, B., Erdős, L., Knowles, A., Yau, H.-T. & Yin, J. (2014), 'Isotropic local laws for sample covariance and generalized Wigner matrices', *Electronic Journal of Probability* **19**(none), 1 – 53.

Baik, J., Arous, G. B. & Péché, S. (2005), 'Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices', *The Annals of Probability* **33**(5), 1643 – 1697.

Baik, J. & Silverstein, J. W. (2006), 'Eigenvalues of large sample covariance matrices of spiked population models', *Journal of multivariate analysis* **97**(6), 1382–1408.

Couillet, R., Debbah, M. & Silverstein, J. W. (2011), 'A deterministic equivalent for the analysis of correlated mimo multiple access channels', *IEEE Transactions on Information Theory* **57**(6), 3493–3514.

Ding, X. (2021), 'Spiked sample covariance matrices with possibly multiple bulk components', *Random Matrices: Theory and Applications* **10**(01), 2150014.

Ding, X., Hu, Y. & Wang, Z. (2024), 'Two sample test for covariance matrices in ultra-high dimension', *Journal of the American Statistical Association* pp. 1–12.

Ding, X. & Wang, Z. (2025), 'Global and local clts for linear spectral statistics of general sample covariance matrices when the dimension is much larger than the sample size with applications', *IEEE Transactions on Information Theory* .

Erdős, L., Knowles, A. & Yau, H.-T. (2013), Averaging fluctuations in resolvents of random band matrices, *in* 'Annales Henri Poincaré', Vol. 14, Springer, pp. 1837–1926.

Erdős, L., Yau, H.-T. & Yin, J. (2012), 'Rigidity of eigenvalues of generalized wigner matrices', *Advances in Mathematics* **229**(3), 1435–1515.

Jiang, D. & Bai, Z. (2021), 'Generalized four moment theorem and an application to CLT for spiked eigenvalues of high-dimensional covariance matrices', *Bernoulli* **27**(1), 274 – 294.

Karoui, N. E. (2003), 'On the largest eigenvalue of wishart matrices with identity covariance when n, p and p/n tend to infinity', *arXiv preprint math/0309355* .

Knowles, A. & Yin, J. (2017), 'Anisotropic local laws for random matrices', *Probability Theory and Related Fields* **169**, 257–352.

Li, Y., Chen, K., Yan, J. & Zhang, X. (2021), 'Uncertainty in optimal fingerprinting is underestimated', *Environmental Research Letters* **16**(8), 084043.

Morice, C. P., Kennedy, J. J., Rayner, N. A. & Jones, P. D. (2012), 'Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The hadcrut4 data set', *Journal of Geophysical Research: Atmospheres* **117**(D8).

Paul, D. (2007), 'Asymptotics of sample eigenstructure for a large dimensional spiked covariance model', *Statistica Sinica* pp. 1617–1642.

Taylor, K. E., Stouffer, R. J. & Meehl, G. A. (2012), 'An overview of cmip5 and the experiment design', *Bulletin of the American meteorological Society* **93**(4), 485–498.