Lecture 9-10

1.1. Stein's Lemma

The following lemma is a key ingredient to prove many results in random matrix theory. We will follow the book [**CGS11**]. We say that $X \sim \mathcal{N}(m, \sigma^2)$ if its probability density function (p.d.f) reads as

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}(x-m)^2\Big).$$

LEMMA 1.1. *Let $X \sim \mathcal{N}(0,1)$. Then*

$$\mathbb{E}\, f'(X) = \mathbb{E}\, X f(X) \tag{1.1}$$

holds for all absolutely continuous functions $f : \mathbb{R} \to \mathbb{R}$, such that $\mathbb{E}\,|f'(X)| < \infty$. Moreover, if (1.1) holds for all bounded, continuous, piecewise continuously differentiable functions f, such that $\mathbb{E}\,|f'(X)| < \infty$, then $X \sim \mathcal{N}(0,1)$.

PROOF. Let $X \sim \mathcal{N}(0,1)$ and f be absolutely continuous function with $\mathbb{E}\,|f'(X)| < \infty$. Then

$$\begin{aligned}
\mathbb{E}\, f'(X) &= \frac{1}{\sqrt{2\pi}} \int f'(x) e^{-\frac{x^2}{2}}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} f'(x) \int_{-\infty}^{x} -y e^{-\frac{y^2}{2}}\, dy\, dx + \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} f'(x) \int_{x}^{\infty} y e^{-\frac{y^2}{2}}\, dy\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} \int_{y}^{0} f'(x)\, dx\, (-y) e^{-\frac{y^2}{2}}\, dy + \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} \int_{0}^{y} f'(x)\, dx\, y e^{-\frac{y^2}{2}}\, dy \\
&= \frac{1}{\sqrt{2\pi}} \int (f(x) - f(0)) x e^{-\frac{x^2}{2}}\, dx = \mathbb{E}\, X f(X).
\end{aligned}$$

Let $z \in \mathbb{R}$. Denote $\Phi(z) = \mathbb{P}(Z \le z)$ d.f. of $Z \sim \mathcal{N}(0,1)$. Unique bounded solution $f(w) = f_z(w)$ of

$$f'(w) - w f(w) = \mathbb{1}[w \le z] - \Phi(z) \tag{1.2}$$

is given by

$$f_z(w) = \begin{cases} \sqrt{2\pi}\Phi(w)(1-\Phi(z)), & w \le z, \\ \sqrt{2\pi}\Phi(z)(1-\Phi(w)), & w > z. \end{cases} \tag{1.3}$$

Indeed,

$$(e^{-\frac{w^2}{2}} f(w))' = e^{-\frac{w^2}{2}} (\mathbb{1}[w \le z] - \Phi(z)).$$

Hence,

$$f_z(w) = -e^{\frac{w^2}{2}} \int_w^\infty (\mathbb{1}[w \le z] - \Phi(z)) e^{-\frac{x^2}{2}}\, dx + C e^{\frac{w^2}{2}}.$$

One may show that the first term in the previous equation is bounded (prove this fact!). Hence, $C = 0$. Moreover, one may check that $|f_z'(w)| \lesssim 1$ (prove this fact!). Let us take $f(w) = f_z(w)$, where $f_z(w)$ is given by (1.3). It is easy to check that this function is bounded, continuous, piecewise continuously differentiable functions f, such that $\mathbb{E}\,|f'(X)| < \infty$. It follows from (1.2) that

$$0 = \mathbb{E}(f_z'(X) - X f_z(X)) = \mathbb{E}(\mathbb{1}[X \le z] - \Phi(z)) = \mathbb{P}(X \le z) - \Phi(z).$$

$\square$

1.2. Gaussian processes

To estimate the operator norm of the random matrix we will often use the following lemma. To prove it we will follow [**vH15**] and [**Ben03**] choosing trigonometric parametrization.

Lemma 1.2. *Let $X \sim \mathcal{N}(0, \Sigma^X)), Y \sim \mathcal{N}(0, \Sigma^Y))$ and*

$$\mathbb{E}(X_i - X_j)^2 \le \mathbb{E}(Y_i - Y_j)^2, \quad 1 \le i, j \le N.$$

Then

$$\mathbb{E} \max_{1 \le i \le N} X_i \le \mathbb{E} \max_{1 \le i \le N} Y_i$$

Proof. Introduce r.v. $Z(\phi) \stackrel{\text{def}}{=} \cos\phi X + \sin\phi Y, 0 \le \phi \le \pi/2$. Then $Z(0) = X, Z(\pi/2) = Y$. Let g_1, g_2 be i.i.d $\mathcal{N}(0, \mathbf{I})$ r.v. Then $X = [\Sigma^X]^{\frac{1}{2}} g_1, Y = [\Sigma^Y]^{\frac{1}{2}} g_2$. Applying Lemma 1.1 we get

$$\begin{aligned}
\frac{\partial}{\partial\phi}\,\mathbb{E}\, f(Z) &= \mathbb{E}\left(\frac{\partial f(Z)}{\partial Z}, \frac{\partial Z}{\partial \phi}\right) = -\sin\phi\,\mathbb{E}\left(\frac{\partial f(Z)}{\partial Z}, X\right) + \cos\phi\,\mathbb{E}\left(\frac{\partial f(Z)}{\partial Z}, Y\right) \\
&= -\sin\phi\,\mathbb{E}\left([\Sigma^X]^{\frac{1}{2}} \frac{\partial f(Z)}{\partial Z}, g_1\right) + \cos\phi\,\mathbb{E}\left([\Sigma^Y]^{\frac{1}{2}} \frac{\partial f(Z)}{\partial Z}, g_2\right) \\
&= \sin\phi\cos\phi \sum_{i,j=1}^N (\Sigma_{ij}^Y - \Sigma_{ij}^X)\,\mathbb{E}\,\frac{\partial^2 f}{\partial x_i \partial x_j}(Z(\phi)).
\end{aligned}$$

We take

$$f(x) \stackrel{\text{def}}{=} f_\beta(x) \stackrel{\text{def}}{=} \frac{1}{\beta} \log\left(\sum_{i=1}^N e^{\beta x_i}\right), \quad \beta > 0.$$

It is straightforward to check that

$$\max x_i \le f_\beta(x) \le \max x_i + \frac{\log N}{\beta}.$$

Then $f_\beta(x) \to \max x_i$, $\beta \to \infty$. Moreover,

$$\frac{\partial f_\beta}{\partial x_i} = \frac{e^{\beta x_i}}{\sum_{j=1}^N e^{\beta x_j}} \stackrel{\mathsf{def}}{=} p_i(x), \quad \frac{\partial^2 f_\beta}{\partial x_i \partial x_j} = \beta(\delta_{ij} p_i(x) - p_i(x) p_j(x)),$$

$p_i(x) \geq 0, \sum p_i(x) = 1$. Then

$$\frac{\partial}{\partial \phi} \mathbb{E} f_\beta(Z) = \frac{\beta \sin\phi \cos\phi}{2} \sum_{i \neq j}^N (\mathbb{E}(Y_i - Y_j)^2 - \mathbb{E}(X_i - X_j)^2) \mathbb{E} p_i(Z) p_j(Z) \geq 0.$$

Hence,

$$\mathbb{E}(f_\beta(Z(\pi/2)) - f_\beta(Z(0))) = \int_0^{\pi/2} \frac{\partial}{\partial \phi} \mathbb{E} f_\beta(Z)\, d\phi \geq 0.$$

It remains to take $\beta \to \infty$. $\square$

1.3. Statistical applications: Covariance matrix estimation in the Gaussian case

Let $X, X_1, \ldots, X_n$ be independent identically distributed (i.i.d.) random vectors taking values in $\mathbb{R}^p$ with mean zero and $\mathbb{E}\|X\|^2 < \infty$. Denote by $\mathbf{\Sigma}$ its $p \times p$ symmetric covariance matrix defined as

$$\mathbf{\Sigma} \stackrel{\mathsf{def}}{=} \mathbb{E}(XX^\top).$$

We also consider the sample covariance matrix $\widehat{\mathbf{\Sigma}}$ of the observations $X_1, \ldots, X_n$ defined as the sum of $X_j X_j^\top$:

$$\widehat{\mathbf{\Sigma}} \stackrel{\mathsf{def}}{=} \frac{1}{n} \sum_{j=1}^n X_j X_j^\top = \frac{1}{n} \mathbf{X}\mathbf{X}^\top,$$

where $\mathbf{X} \stackrel{\mathsf{def}}{=} [X_1, \ldots, X_n] \in \mathbb{R}^{p \times n}$.

In statistical applications, the true covariance matrix $\mathbf{\Sigma}$ is typically unknown and one often uses the sample covariance matrix $\widehat{\mathbf{\Sigma}}$ as its estimator. The accuracy $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|$ of estimation of $\mathbf{\Sigma}$ by $\widehat{\mathbf{\Sigma}}$, in particular, for p much larger than n, has been actively studied in the literature. We refer to [**Tro12**] for an overview of the recent results based on the matrix Bernstein inequality; see also [**Ver12**] and [**vH15**]. A bound in term of the effective rank $\mathtt{r}(\mathbf{\Sigma}) \stackrel{\mathsf{def}}{=} \mathrm{Tr}(\mathbf{\Sigma})/\|\mathbf{\Sigma}\|$ can be found in [**KL15**]. This or similar bound on the spectral norm $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|$ can be effectively applied to relate the eigenvalues of $\mathbf{\Sigma}$ and of $\widehat{\mathbf{\Sigma}}$ under the spectral gap condition.

LEMMA 1.3. *Let $X_1, \ldots, X_n$ be i.i.d. $\mathcal{N}(0, \mathbf{\Sigma})$ random vectors in $\mathbb{R}^p$. Then*

$$\mathbb{E}\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\| \lesssim \|\mathbf{\Sigma}\| \left(\sqrt{\frac{\mathtt{r}(\mathbf{\Sigma})}{n}} + \frac{\mathtt{r}(\mathbf{\Sigma})}{n} \right).$$

PROOF. We will follow Ramon Van Handel. It is easy to see that

$$\begin{aligned}\mathbb{E}\,\|\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\| &= \mathbb{E}\sup_{u\in B_2}\left|((\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma})u,u)\right| \\ &= \mathbb{E}\sup_{u\in B_2}\left|\frac{1}{n}\sum_{j=1}^{n}(X_jX_j^{\mathsf{T}}u,u)-(\mathbf{\Sigma}u,u)\right| \\ &= \mathbb{E}\sup_{u\in B_2}\left|\frac{1}{n}\sum_{j=1}^{n}(X_j,u)^2-\mathbb{E}(u,X)^2\right|\end{aligned}$$

We may use ε-net argument + Bernstein's inequality.

DEFINITION 1.4. Consider a subset K of $\mathbb{R}^n$ and let $\varepsilon>0$. A subset $\mathcal{N}\subset K$ is called an ε-net of K if every point in K is within distance ε of some point of $\mathcal{N}$, i.e.

$$\forall x\in K, \exists x_0\in\mathcal{N}: \|x-x_0\|\le\varepsilon.$$

DEFINITION 1.5. The smallest cardinality of an ε-net of K is called the covering number of K and is denoted $N(K;\varepsilon)$. Equivalently, the $N(K;\varepsilon)$ is the smallest number of closed balls with centers in K and radius ε whose union covers K.

LEMMA 1.6. *The covering numbers of the unit Euclidean ball B_2^n satisfy the following for any $\varepsilon>0$:*

$$\left(\frac{1}{\varepsilon}\right)^n \le N(B_2^n;\varepsilon)\le\left(\frac{2}{\varepsilon}+1\right)^n$$

The same upper bound is true for the unit Euclidean sphere S^{n-1}.

LEMMA 1.7. *Let $\mathbf{A}\in\mathbb{R}^{m\times n}$. Then for any ε-net $\mathcal{N}$ of S^{n-1} and any ε-net $\mathcal{M}$ of S^{m-1} we get*

$$\sup_{u\in\mathcal{N},v\in\mathcal{M}}(Au,v)\le\|\mathbf{A}\|\le\frac{1}{1-2\varepsilon}\sup_{u\in\mathcal{N},v\in\mathcal{M}}(Au,v)$$

EXERCISE 1.8. Prove Lemma 1.7.

Let $\widetilde{\mathbf{X}}$ be independent copy of $\mathbf{X}$. Then

$$\mathbb{E}\,\|\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\| = \frac{1}{n}\mathbb{E}_{\mathbf{X}}\,\|\mathbb{E}_{\widetilde{X}}(\mathbf{X}-\widetilde{\mathbf{X}})(\mathbf{X}+\widetilde{\mathbf{X}})\|\le\frac{1}{n}\mathbb{E}\,\|(\mathbf{X}-\widetilde{\mathbf{X}})(\mathbf{X}+\widetilde{\mathbf{X}})\| = \frac{2}{n}\mathbb{E}\,\|\mathbf{X}\widetilde{\mathbf{X}}^{\mathsf{T}}\| \quad (1.4)$$

Hence,

$$\mathbb{E}\,\|\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\|\le\frac{2}{n}\mathbb{E}\,\|\mathbf{X}\widetilde{\mathbf{X}}^{\mathsf{T}}\| = \frac{2}{n}\mathbb{E}\sup_{u,v\in B_2}\left|\sum_{j=1}^{n}(\widetilde{X}_j,u)(X_j,v)\right|$$

Let us fix $u,v\in B_2$. Denote

$$Z_{u,v}\overset{\mathsf{def}}{=}\sum_{j=1}^{n}(\widetilde{X}_j,u)(X_j,v).$$

Fix $\widetilde{\mathbf{X}}$. Then

$$\begin{aligned}
\mathbb{E}_{\widetilde{\mathbf{X}}}(Z_{u,v} - Z_{u',v'})^2 &\le 2(u-u', \mathbf{\Sigma}(u-u'))\sum_{j=1}^{n}(v, \widetilde{X}_j)^2 + 2(u', \mathbf{\Sigma}u')\sum_{j=1}^{n}(v-v', \widetilde{X}_j)^2 \\
&\le 2\|\widetilde{\mathbf{X}}\|^2\|\mathbf{\Sigma}^{\frac{1}{2}}(u-u')\| + 2\|\mathbf{\Sigma}\|\|\widetilde{\mathbf{X}}^{\mathsf{T}}(v-v')\|^2 \\
&= \mathbb{E}_{\widetilde{X}}(Y_{u,v} - Y_{u',v'})^2,
\end{aligned}$$

where

$$Y_{u,v} \overset{\mathsf{def}}{=} \sqrt{2}\|\widetilde{\mathbf{X}}\|(u, \mathbf{\Sigma}^{\frac{1}{2}}g) + (2\|\mathbf{\Sigma}\|)^{\frac{1}{2}}(v, \widetilde{X}g').$$

By Lemma 1.2

$$\begin{aligned}
\mathbb{E}_{\widetilde{\mathbf{X}}}\sup_{u,v\in B_2} Z_{u,v} \le \mathbb{E}_{\widetilde{\mathbf{X}}}\sup_{u,v\in B_2} Y_{u,v} &\lesssim \|\widetilde{\mathbf{X}}\|\,\mathbb{E}\,\|\mathbf{\Sigma}^{\frac{1}{2}}g\| + \|\mathbf{\Sigma}\|^{\frac{1}{2}}\,\mathbb{E}\,\|\widetilde{X}g'\| \\
&\le \|\widetilde{\mathbf{X}}\|\sqrt{\operatorname{Tr}\mathbf{\Sigma}} + \|\mathbf{\Sigma}\|^{\frac{1}{2}}\operatorname{Tr}^{\frac{1}{2}}(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathsf{T}}).
\end{aligned}$$

Hence,

$$\mathbb{E}\,\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\| \le \frac{1}{n}\,\mathbb{E}\,\|\mathbf{X}\|\sqrt{\operatorname{Tr}\mathbf{\Sigma}} + \frac{1}{\sqrt{n}}\|\mathbf{\Sigma}\|^{\frac{1}{2}}\sqrt{\operatorname{Tr}\mathbf{\Sigma}}.$$

It remains to show that

$$\mathbb{E}\,\|\mathbf{X}\| \lesssim \sqrt{\operatorname{Tr}\mathbf{\Sigma}} + \sqrt{n\|\mathbf{\Sigma}\|} \tag{1.5}$$

Similarly,

$$\begin{aligned}
\mathbb{E}((u, \mathbf{X}v) - (u', \mathbf{X}v'))^2 &\le 2\,\mathbb{E}(u-u', \mathbf{X}w)^2 + 2\,\mathbb{E}(u', \mathbf{X}(v-v'))^2 \\
&\le 2\|\mathbf{\Sigma}^{\frac{1}{2}}(u-u')\|^2 + 2\|\mathbf{\Sigma}^{\frac{1}{2}}v'\|^2\|v-v'\|^2 \\
&= \mathbb{E}(Y_{u,v} - Y_{u',v'})^2,
\end{aligned}$$

where

$$Y_{u,v} = \sqrt{2}(u, \mathbf{\Sigma}^{\frac{1}{2}}g) + \sqrt{2}\|\mathbf{\Sigma}\|^{\frac{1}{2}}(v, g').$$

Hence, by Lemma 1.2

$$\mathbb{E}\,\|\mathbf{X}\| \le \mathbb{E}\sup_{u,v\in B_2} Y_{u,v} \le \mathbb{E}\,\|\mathbf{\Sigma}^{\frac{1}{2}}g\| + \|\mathbf{\Sigma}\|^{\frac{1}{2}}\,\mathbb{E}\,\|g\|.$$

□

Lecture 11

The aim of this section is to introduce some example of classical invariant Gaussian ensembles of random matrices.

2.1. Gaussian Orthogonal and Unitary Ensembles

DEFINITION 2.1 (GOE). We say that a random matrix $\mathbf{X} \in \mathbb{R}^{n\times n}$ belongs to the *Gaussian orthogonal ensemble (GOE)* if: $X_{jk}, 1 \le j \le k \le n$, are independent $\mathcal{N}(0, \sigma_{jk}^2)$ r.v. with $\sigma_{jk}^2 = 1 + \delta_{jk}$, where δ_{jk} is Kronecker delta ($\delta_{jk} = 1$ if $j = k$, and 0 otherwise).

Let H_{jk} denotes the entry of $\mathbf{H} \in \mathbb{R}^{n\times n}, \mathbf{H}^\mathsf{T} = \mathbf{H}$ and $\mathbb{P}_n^{(1)}$ be the law of $\mathbf{X}$. Then

$$
\begin{aligned}
d\,\mathbb{P}_n^{(1)}(\mathbf{H}) &= \prod_{j<k} \frac{1}{\sqrt{2\pi}} e^{-\frac{H_{jk}^2}{2}} \prod_{j=1}^n \frac{1}{\sqrt{4\pi}} e^{-\frac{H_{jj}^2}{4}} d\mathbf{H} \\
&= \frac{1}{2^{n/2}(2\pi)^{n(n+1)/4}} \exp(-\operatorname{Tr}\mathbf{H}^2/4) d\mathbf{H}, \qquad (2.1)
\end{aligned}
$$

where $d\mathbf{H} = \prod_{j<k} dH_{jk} \prod_{j=1}^n dH_{jj}$. It is easy to see that $d\,\mathbb{P}_n^{(1)}(\mathbf{H}) = d\,\mathbb{P}_n^{(1)}(\mathbf{U}\mathbf{H}\mathbf{U}^\mathsf{T})$ for all $\mathbf{U} \in \mathbb{O}(n)$ (invariance with respect to the orthogonal group transformations).

LEMMA 2.2. *Let $\mathbf{X}$ belongs to GOE. The joint density function of $\lambda_1(\mathbf{X}) \le \lambda_2(\mathbf{X}) \le \ldots \le \lambda_n(\mathbf{X})$ is given by*

$$
p_n^{(1)}(x_1, ..., x_n) = n! C_n^{(1)} \mathbb{I}[x_1 \le ... \le x_n] |\Delta(x)| \prod_{i=1}^n e^{-x_i^2/4},
$$

where

$$
\Delta(x) \overset{\mathsf{def}}{=} \prod_{i<j} (x_j - x_i)
$$

and $C_n^{(1)}$ is some constant.

PROOF. We will follow [Meh91][Chapter 3] and prove this lemma on the physical level of rigour. It is easy to see that (2.1) is defined by $N = n(n+1)/2$ parameters, $H_{jk}, 1 \le j \le k \le n$. Since the number of eigenvalues $\lambda_1, \ldots, \lambda_n$ is n we need to find additional $L = N - n = n(n-1)/2$ parameters $p_1, \ldots, p_L$. We may rewrite (2.1) as follows

$$
d\,\mathbb{P}_n^{(1)}(\mathbf{H}) = \exp\Big(-\frac{1}{4}\sum_{j=1}^n \lambda_j^2 + c_n\Big) d\mathbf{H},
$$

where $c_n \stackrel{\mathsf{def}}{=} \log(2^{-n/2}(2\pi)^{-n(n+1)/4})$.

Let $y = \varphi(x)$, $\varphi : \mathbb{R}^N \to \mathbb{R}^N$, and denote the Jacobian matrix of φ by

$$\mathbf{J}_\varphi(x) \stackrel{\mathsf{def}}{=} \left[\frac{\partial \varphi_j}{\partial x_k}\right], 1 \le j, k \le N.$$

Then the density of $\eta = \varphi(\xi)$ is given by

$$f_\eta(y) = f_\xi(\varphi^{-1}(y))|\det \mathbf{J}_{\varphi^{-1}}(y)|. \tag{2.2}$$

Applying (2.2) we get

$$f(\lambda_1, \dots, \lambda_n, p_1, \dots, p_L) = \exp\Big(-\frac{1}{4}\sum_{j=1}^n \lambda_j^2 + c_n\Big)|\det \mathbf{J}(\lambda, p)|,$$

where

$$\mathbf{J}(\lambda, p) = \Big[\frac{\partial(H_{11}, \dots, H_{nn})}{\partial(\lambda_1, \dots, \lambda_n, p_1, \dots, p_L)}\Big].$$

We rewrite $\mathbf{H}$ as follows

$$\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}, \tag{2.3}$$

where $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ and

$$\mathbf{U}\mathbf{U}^\mathsf{T} = \mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{I}, \tag{2.4}$$

We assume that all λ_j are distinct (in general case see [**AGZ10**]). We choose $\mathbf{U}$ such that the first non-zero coordinates of each column are positive. Then we may choose $p_j, j = 1, \dots, L$ equal to $U_{kl}, k > l$. Differentiating (2.4) we obtain

$$\frac{\partial \mathbf{U}^\mathsf{T}}{\partial p_l}\mathbf{U} + \mathbf{U}^\mathsf{T}\frac{\partial \mathbf{U}}{\partial p_l}\mathbf{U} = 0, \quad l = 1, \dots, L.$$

Let us define the following matrix:

$$\mathbf{S}_l \stackrel{\mathsf{def}}{=} \mathbf{U}^\mathsf{T}\frac{\partial \mathbf{U}}{\partial p_l} = -\frac{\partial \mathbf{U}^\mathsf{T}}{\partial p_l}\mathbf{U}.$$

From (2.3) we get

$$\frac{\partial \mathbf{H}}{\partial p_l} = \frac{\partial \mathbf{U}}{\partial p_l}\mathbf{\Lambda}\mathbf{U}^\mathsf{T} + \mathbf{U}\mathbf{\Lambda}\frac{\partial \mathbf{U}^\mathsf{T}}{\partial p_l}.$$

Multiplying the last equation by $\mathbf{U}^\mathsf{T}$ and $\mathbf{U}$ we write

$$\mathbf{U}^\mathsf{T}\frac{\partial \mathbf{H}}{\partial p_l}\mathbf{U} = \mathbf{S}_l\mathbf{\Lambda} - \mathbf{\Lambda}\mathbf{S}_l,$$

or

$$\sum_{j,k}\frac{\partial \mathbf{H}_{jk}}{\partial p_l}U_{j\alpha}U_{k\beta} = S^{(l)}_{\alpha\beta}(\lambda_\beta - \lambda_\alpha).$$

Similarly,

$$\sum_{j,k}\frac{\partial \mathbf{H}_{jk}}{\partial \lambda_\gamma}U_{j\alpha}U_{k\beta} = \delta_{\alpha\beta}\delta_{\alpha\gamma}.$$

We may rewrite $\mathbf{J}(\lambda, p)$ in the block form

$$\mathbf{J}(\lambda, p) = \begin{pmatrix} \frac{\partial \mathbf{H}_{jj}}{\partial \lambda_\gamma} & \frac{\partial \mathbf{H}_{jk}}{\partial \lambda_\gamma} \\ \frac{\partial \mathbf{H}_{jj}}{\partial p_l} & \frac{\partial \mathbf{H}_{jk}}{\partial p_l} \end{pmatrix}, \quad \gamma = 1, \ldots, n, l = 1, \ldots, L.$$

Let

$$\mathbf{V} = \begin{pmatrix} (U_{j\alpha} U_{j\beta}) \\ (2U_{j\alpha} U_{k\beta}) \end{pmatrix}, \quad 1 \le j < k \le n, 1 \le \alpha \le \beta \le n.$$

Then

$$\mathbf{J}(\lambda, p)\mathbf{V} = \begin{pmatrix} \delta_{\alpha\beta}\delta_{\alpha\gamma} \\ S^{(l)}_{\alpha\beta}(\lambda_\beta - \lambda_\alpha) \end{pmatrix}.$$

Hence,

$$|\det \mathbf{J}(\lambda, p)||\det \mathbf{V}| = \prod_{\alpha<\beta} |\lambda_\alpha - \lambda_\beta||\det \begin{pmatrix} \delta_{\alpha\beta}\delta_{\alpha\gamma} \\ S^{(l)}_{\alpha\beta} \end{pmatrix}|,$$

or

$$|\det \mathbf{J}(\lambda, p)| = \prod_{\alpha<\beta} |\lambda_\alpha - \lambda_\beta| f(p),$$

where $f(p)$ is independent of $\lambda_j, j = 1, \ldots, n$. Integrating the last equation with respect to $p_l, l = 1, \ldots, L$ we obtain

$$p^{(1)}_n(\lambda_1, \ldots, \lambda_n) = \exp\Big(-\frac{1}{4}\sum_{j=1}^n \lambda_j^2 + c^{(1)}_n \Big) \prod_{j<k} |\lambda_j - \lambda_k|,$$

where $c^{(1)}_n$ is some positive constant. $\square$

We may also study the eigenvectors of $\mathbf{X}$. Let us denote by $\mathbf{u}_j := (u_{j1}, ..., u_{jn})$ the eigenvectors of $\mathbf{W}$ corresponding to the eigenvalue $\lambda_j(\mathbf{W})$.

Lemma 2.3. *The collections* $(\mathbf{u}_1, ..., \mathbf{u}_n)$ *is independent of the eigenvalues* $(\lambda_1(\mathbf{W}), ..., \lambda_n(\mathbf{W}))$. *Each of the eigenvectors is distributed uniformly on*

$$S^{n-1}_+ \stackrel{\mathsf{def}}{=} \{x = (x_1, ..., x_n) : x_j \in \mathbb{R}, \|x\| = 1, x_1 > 0\}.$$

Further, $(\mathbf{u}_1, ..., \mathbf{u}_n)$ *is distributed like a sample of Haar measure on* $\mathbb{O}(n)$*, with each column multiplied by a norm one scalar so that the columns all belong to* S^{n-1}_+.

Proof. $\mathbf{W} = \mathbf{U}\Lambda\mathbf{U}^\mathsf{T}$. Let $\mathbf{T} \in \mathbb{O}(n)$. Then $\mathbf{T}\mathbf{W}\mathbf{T}^\mathsf{T}$ possesses the same eigenvalues as $\mathbf{W}$ and is distributed like $\mathbf{W}$. Choosing $\mathbf{T}$ uniformly according to Haar measure and independent of $\mathbf{U}$ makes $\mathbf{T}\mathbf{U}$ Haar distributed and independent of $\mathbf{U}$ and $\lambda_1(\mathbf{W}), ..., \lambda_n(\mathbf{W})$. $\square$

Definition 2.4 (GUE). We say that a random matrix $\mathbf{X} \in \mathbb{C}^{n\times n}$ belongs to the *Gaussian unitary ensemble (GUE)* if:

- $X_{jk}, 1 \le j < k \le n$, are independent complex r.v.;
- $\operatorname{Re} X_{jk}$ and $\operatorname{Im} X_{jk}$ are independent $\mathcal{N}\big(0, \frac{1}{2}\big)$, $1 \le j < k \le n$;
- X_{jj} are i.i.d. $\mathcal{N}(0, 1)$ real random variables independent of off-diagonal entries.

EXERCISE 2.5. Let H_{jk} denotes the entry of $\mathbf{H} \in \mathbb{C}^{n\times n}, \mathbf{H}^* = \mathbf{H}$ and $\mathbb{P}_n^{(2)}$ be the law of $\mathbf{X}$ from GUE. Show that

$$d\,\mathbb{P}_n^{(2)}(\mathbf{H}) = \frac{1}{2^{n/2}\pi^{n^2/2}} \exp(-\operatorname{Tr}\mathbf{H}^2/2)d\mathbf{H}, \tag{2.5}$$

where $d\mathbf{H} = \prod_{j<k} dH_{jk} \prod_{j=1}^n dH_{jj}$. It is easy to see that $d\,\mathbb{P}_n^{(2)}(\mathbf{H}) = d\,\mathbb{P}_n^{(2)}(\mathbf{U}\mathbf{H}\mathbf{U}^\mathsf{T})$ for all $\mathbf{U} \in \mathbb{U}_n$ (invariance with respect to unitary group transformations).

Similarly, one may show that for $\mathbf{X}$ from GUE the joint density function of $\lambda_1(\mathbf{X}) \le \lambda_2(\mathbf{X}) \le \ldots \le \lambda_n(\mathbf{X})$ is given by

$$p_n^{(2)}(x_1, ..., x_n) = n!C_n^{(2)}\mathbb{I}[x_1 \le ... \le x_n]|\Delta(x)|^2 \prod_{i=1}^n e^{-x_i^2/2},$$

where $C_n^{(2)}$ is some constant.

2.2. Wishart–Laguerre Ensemble.

Let $X, X_1, ..., X_n$ be i.i.d. random vectors in $\mathbb{R}^p$. Denote by $\mathbf{\Sigma}$ its $p \times p$ symmetric covariance matrix defined as

$$\mathbf{\Sigma} \overset{\text{def}}{=} \mathbb{E}(XX^\mathsf{T}).$$

We also consider the sample covariance matrix $\widehat{\mathbf{\Sigma}}$ of the observations $X_1, \ldots, X_n$ defined as the sum of $X_j X_j^\mathsf{T}$:

$$\widehat{\mathbf{\Sigma}} \overset{\text{def}}{=} \frac{1}{n}\sum_{j=1}^n X_j X_j^\mathsf{T} = \frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}, \tag{2.6}$$

where $\mathbf{X} \overset{\text{def}}{=} [X_1, \ldots, X_n] \in \mathbb{R}^{p\times n}$. In statistical applications, the true covariance matrix $\mathbf{\Sigma}$ is typically unknown and one often uses the sample covariance matrix $\widehat{\mathbf{\Sigma}}$ as its estimator.

DEFINITION 2.6 (LOE). We say that $\mathbf{S}$, where $\mathbf{S} = \mathbf{X}\mathbf{X}^\mathsf{T}$, $\mathbf{X} = [X_1, ..., X_n], X_j \in \mathbb{R}^p, j = 1, \ldots, n$, belongs to *Wishart–Laguerre Ensemble (LOE)* if X_j are i.i.d. $\mathcal{N}(0, \mathbf{I})$ random vectors.

Assume that $\mathbf{S}$ belongs to LOE, i.e. $\mathbf{\Sigma} = \mathbf{I}$. Let $\mathbf{H} \in \mathbb{R}^{p\times n}$. It is straightforward to check that

$$d\,\mathbb{P}_{n,p}^{\mathbf{X}}(\mathbf{H}) = (2\pi)^{-\frac{np}{2}} \exp\left(-\frac{\operatorname{Tr}(\mathbf{H}\mathbf{H}^\mathsf{T})}{2}\right) d\mathbf{H},$$

where $d\mathbf{H} = \prod_{j=1,k=1}^{p,n} dH_{jk}$. Moreover, let $\mathbf{M} \in \mathbb{R}^{p\times p}$ then

$$d\,\mathbb{P}_{n,p}^{\mathbf{S}}(\mathbf{M}) = c_{n,p} \det(\mathbf{M})^{n-p-1} \exp\left(-\frac{\operatorname{Tr}(\mathbf{M})}{2}\right) d\mathbf{M}\,\mathbb{I}[\mathbf{M} \ge 0],$$

where $d\mathbf{M} = \prod_{1\le j\le k\le p} dM_{jk}$ (see J. Wishart, The generalized product moment distribution in samples from a normal multivariate population, Biometrika, 1928).

Let $\lambda_1(\mathbf{S}) \leq \lambda_2(\mathbf{S}) \leq \ldots \leq \lambda_p(\mathbf{S})$ be the eigenvalues of $\mathbf{S}$ (squared singular values of $\mathbf{X}$). One may write down the joint density of $\lambda_1, ..., \lambda_p$:

$$p(x_1, ..., x_p) = p! C_{n,p} \mathbb{I}[0 \leq x_1 \leq ... \leq x_p] \prod_{i<j} |x_j - x_i| \prod_{i=1}^{p} x_j^{\frac{n-p-1}{2}} e^{-x_i/2}.$$

Lecture 12

3.1. Wigner's semicirle law

We consider a random symmetric matrix $\mathbf{X} = [X_{jk}]_{j,k=1}^{n}$ where the upper triangular entries are independent random variables with mean zero and unit variance. We will be mostly interested in limiting laws for the eigenvalues and eigenvectors of large $n \times n$ symmetric random matrices in the asymptotic limit as n goes to infinity.

For the symmetric matrix $\mathbf{W} := \frac{1}{\sqrt{n}}\mathbf{X}$ we denote its n eigenvalues in the increasing order as

$$\lambda_1(\mathbf{W}) \le \ldots \le \lambda_n(\mathbf{W})$$

and introduce the eigenvalue counting function

$$N_I(\mathbf{W}) := |\{1 \le k \le n : \lambda_k(\mathbf{W}) \in I\}|$$

for any interval $I \subset \mathbb{R}$, where $|A|$ denotes the number of elements in the set A. Note that we shall sometimes omit $\mathbf{W}$ from the notation of $\lambda_j(\mathbf{W})$.

Applying $\mathbf{W} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\mathsf{T}$ ($\mathbf{U}\mathbf{U}^\mathsf{T} = \mathbf{I}, \boldsymbol{\Lambda} := \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$) and $\operatorname{Tr}\mathbf{W}^2 = \operatorname{Tr}\boldsymbol{\Lambda}^2$ we obtain

$$\mathbb{E}\left[\frac{1}{n}\sum_{j=1}^{n}\lambda_j^2(\mathbf{W})\right] = \frac{1}{n}\,\mathbb{E}\operatorname{Tr}\mathbf{W}^2 = \frac{1}{n^2}\sum_{j,k=1}^{n}\mathbb{E}\,X_{jk}^2 = 1.$$

This means that λ_j are typically of order $\mathcal{O}(1)$. In Chapter we show that typically $\|\mathbf{X}\|$ has the order $\sqrt{n}$.

The pioneering result of E. Wigner [Wig55] states that for any interval $I \subset \mathbb{R}$ of fixed length and independent of n

$$\lim_{n\to\infty}\frac{1}{n}\,\mathbb{E}\,N_I(\mathbf{W}) = \int_I g_{sc}(\lambda)\,d\lambda, \tag{3.1}$$

where $g_{sc}(\lambda) := \frac{1}{2\pi}\sqrt{(4-\lambda^2)_+}$ and $(x)_+ := \max(x, 0)$. Wigner considered the special case when all X_{jk} take only two values ± 1 with equal probabilities. Later on the result (3.1) was called *Wigner's semicircle law*. In what follows we call *Wigner's semicircle law or semicircle law* not only the result of type (3.1), but the limiting probability distribution as well.

To prove (3.1) Wigner used the moment method which may be described as follows. Since g_{sc} is compactly supported it is uniquely determined by the sequence of its moments given by

$$\beta_k = \begin{cases} \frac{1}{m+1}\binom{2m}{m}, & k = 2m, \\ 0, & k = 2m+1. \end{cases}, \qquad k \geq 1.$$

We remark here that $\beta_{2m}, m \geq 1$, are Catalan numbers. To establish the convergence (3.1) one needs to show that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} \lambda^k \, dF_n(\lambda) = \int_{-2}^{2} \lambda^k g_{sc}(\lambda) \, d\lambda,$$

where $F_n(\lambda) := \frac{1}{n} N_{(-\infty,\lambda]}(\mathbf{W})$ is the empirical spectral distribution function. Further details may be found in [**BS10**].

THEOREM 3.1 (Wigner's semicircle law). *Let $X_{jk}, 1 \leq j \leq k \leq N$, be independent standard Gaussian r.v. Then for all intervals $I \subset \mathbb{R}$ of fixed length*

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}\, N_I(\mathbf{W}) = \int_I g_{sc}(\lambda) \, d\lambda.$$

3.2. Stieltjes transform

Let μ be an arbitrary probability measure on $\mathbb{R}$. An appropriate analytical tool to investigate the spectrum of random matrices is the Stieltjes transform.

DEFINITION 3.2. The Stieltjes transform of μ is given by

$$s_\mu(z) \overset{\text{def}}{=} \int_{-\infty}^{\infty} \frac{\mu(d\lambda)}{\lambda - z}, \quad z \in \mathbb{C}^+.$$

LEMMA 3.3. *Let $z = u + iv \in \mathbb{C}^+$. The following statements hold:*

(1) $s_\mu(z)$ is analytic function in $\mathbb{C}/\mathbb{R}$.
(2) $\operatorname{Im} s_\mu(z) > 0$ for all $z \in \mathbb{C}^+$.
(3) $|s_\mu(z)| \leq \frac{1}{v}$.
(4) $\operatorname{Im} s_\mu(z)$ is a convolution $\mu \star \nu_v(\cdot)$, where

$$\nu_v(I) \overset{\text{def}}{=} \frac{1}{\pi} \int_I \frac{v}{x^2 + v^2} \, dx$$

is a Cauchy distribution.

(5) $\mu(I) = \frac{1}{\pi} \lim_{\varepsilon\to 0} \int_I \operatorname{Im} s_\nu(u + i\varepsilon) \, du$.
(6) If μ_n weakly converges to μ then $s_{\mu_n}(z) \to s_\mu(z)$ for every $z \in \mathbb{C}^+$. Let μ_n be a sequence of random probability measures and μ be a probability measure. If $s_{\mu_n}(z) \to s_\mu(z)$ a.s. (in probability or in average sense) for all $z \in \mathbb{C}^+$, then μ_n weakly converges to μ a.s (in probability or in average sense respectively).

Statements (1) and (2) together imply that $s_\mu(z)$ is Nevanlinna function (complex function which is an analytic function on the open upper half-plane $\mathbb{C}^+$ and has non-negative imaginary part. A Nevanlinna function maps the upper half-plane to itself). Statement (5) implies that one may recover $\mu(\cdot)$ from $s_\mu(z)$.

The Stieltjes transform of Wigner's semicircle law is given by

$$s_{sc}(z) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \frac{g_{sc}(\lambda)\, d\lambda}{\lambda - z} = -\frac{z}{2} + \sqrt{\frac{z^2}{4} - 1}. \tag{3.2}$$

Moreover, $s_{sc}(z)$ satisfies the following quadratic equation

$$1 + z s_{sc}(z) + s_{sc}^2(z) = 0 \text{ or } s_{sc}(z) = -\frac{1}{z + s_{sc}(z)}. \tag{3.3}$$

EXERCISE 3.4. Prove (3.2) and (3.3).

Define the Stieltjes transform of F_n by

$$m_n(z) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \frac{dF_n(\lambda)}{\lambda - z} = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{\lambda_j - z} = \frac{1}{n} \operatorname{Tr}(\mathbf{W} - z\mathbf{I})^{-1},$$

where $z = u + iv, v \geq 0$ (i.e. $z \in \mathbb{C}^+$). Denote the resolvent of $\mathbf{W}$ by

$$\mathbf{R}(z) \stackrel{\text{def}}{=} (\mathbf{W} - z\mathbf{I})^{-1}. \tag{3.4}$$

EXERCISE 3.5. Show that

$$\|\mathbf{R}(z)\| \leq \frac{1}{v} \quad \text{for all } z = u + iv, v > 0.$$

In these notations

$$m_n(z) = \frac{1}{n} \operatorname{Tr} \mathbf{R}(z) = \frac{1}{n} \frac{\partial}{\partial z} \log \det \mathbf{R}(z).$$

Taking the imaginary part of $m_n(z)$ (compare with the statements (3) and (4)) we get

$$\operatorname{Im} m_n(u + iv) = \int_{-\infty}^{\infty} \frac{v}{(\lambda - u)^2 + v^2}\, dF_n(\lambda) = \frac{1}{v} \int_{-\infty}^{\infty} K\left(\frac{\lambda - u}{v}\right) dF_n(\lambda) \tag{3.5}$$

which is the kernel density estimator with Poisson's kernel

$$K(x) := \frac{1}{x^2 + 1}$$

and bandwidth v. For a meaningful estimator of the spectral density we cannot allow v to be smaller than the typical $\frac{1}{n}$ -distance between eigenvalues. Hence, we shall be mostly interested in the situations when

$$v \geq \frac{c}{n}, c > 0.$$

Figure 3.2 illustrates (3.5) for different values of v.

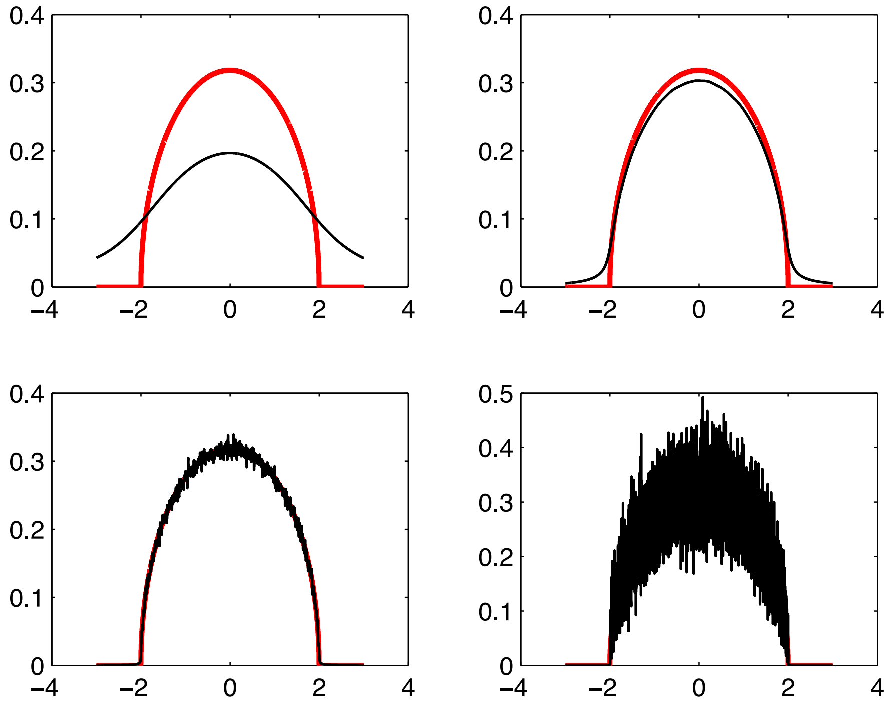

FIGURE 1. Let $n = 5000$. In the top row $v = 1$ (on the left) and $v = 0.1$ (on the right) In the bottom row $v = 0.005$ (on the left) and $v = 0.0007$ (on the right)

3.3. Proof of Wigner's semicircle law

PROOF. We show that

$$\lim_{n\to\infty} \mathbb{E}\, m_n(z) = s(z) \qquad \text{for all } z = u + iv, v > 0.$$

By definition of $\mathbf{R}(z)$

$$\mathbf{R}(z)(\mathbf{W} - z\mathbf{I}) = \mathbf{I}.$$

Taking trace from the both sides we get

$$\frac{1}{n} \operatorname{Tr} \mathbf{W}\mathbf{R}(z) - z\frac{1}{n} \operatorname{Tr} \mathbf{R}(z) = 1.$$

Hence,

$$1 + z\, \mathbb{E}\, m_n(z) = \frac{1}{n}\, \mathbb{E} \operatorname{Tr} \mathbf{W}\mathbf{R}(z).$$

Comparing this equation with (3.3) it is obvious that we need to show the the right hand side is equal to $(\mathbb{E}\, m_n(z))^2 + r_n(v)$, where $r_n(v) \to 0$ as $n \to \infty$ for all $v \geq v_0$.

Here v_0 is some positive number defined below. Applying Lemma 1.1 we obtain

$$\frac{1}{n}\,\mathbb{E}\,\mathrm{Tr}\,\mathbf{W}\mathbf{R}(z) = \frac{1}{n^{\frac{3}{2}}}\sum_{j,k=1}\mathbb{E}\,X_{jk}\mathbf{R}_{kj} = \frac{1}{n^{\frac{3}{2}}}\sum_{j,k=1}\mathbb{E}\,\frac{\partial \mathbf{R}_{kj}}{\partial X_{jk}}$$
$$= -\frac{1}{n^2}\sum_{j,k=1}^{n}[\mathbb{E}\,\mathbf{R}_{kk}\mathbf{R}_{jj} + \mathbb{E}[\mathbf{R}_{kj}]^2]$$

Hence,

$$\frac{1}{n}\,\mathbb{E}\,\mathrm{Tr}\,\mathbf{W}\mathbf{R}(z) = -\,\mathbb{E}[m_n(z)]^2 - \frac{1}{n^2}\,\mathbb{E}\,\|\mathbf{R}(z)\|_2^2.$$

From $\|\mathbf{R}(z)\| \le v^{-1}$ it follows that

$$\frac{1}{n^2}\,\mathbb{E}\,\|\mathbf{R}(z)\|_2^2 \le \frac{1}{n}\,\mathbb{E}\,\|\mathbf{R}(z)\|^2 \le \frac{1}{nv^2}$$

It is easy to see that

$$\mathbb{E}[m_n(z)]^2 = [\mathbb{E}\,m_n(z)]^2 + \mathrm{Var}[m_n(z)].$$

Later on (see Lemma 3.6) we show that

$$\mathrm{Var}[m_n(z)] \le \frac{1}{nv^2}.$$

Finally

$$-(\mathbb{E}\,m_n(z))^2 = 1 + z\,\mathbb{E}\,m_n(z) + r_n(v),$$

or

$$1 + z\,\mathbb{E}\,m_n(z) + (\mathbb{E}\,m_n(z))^2 = r_n(v).$$

We take $n \to \infty$ and compare with (3.3). □

LEMMA 3.6. *Let $X = (X_k, \mathfrak{F}_k)$ be a complex martingale-difference. Let $\mathbb{E}_k(\cdot) \stackrel{\mathsf{def}}{=} \mathbb{E}(\cdot\,|\mathfrak{F}_k)$. Then for all $p \ge 2$*

$$\mathbb{E}\left|\sum_{k=1}^{n} X_k\right|^p \le C_p\left(\mathbb{E}\left(\sum_{k=1}^{n}\mathbb{E}_{k-1}|X_k|^2\right)^{\frac{p}{2}} + \sum_{k=1}^{n}\mathbb{E}\,|X_k|^p\right) \tag{3.6}$$

EXERCISE 3.7. Prove that in Wigner's semicircle law one may replace convergence of expectations by a.s. (in probability) convergence. Hint: Use the following Girko's trick. First prove:

$$\mathbb{E}\left|\frac{1}{n}\,\mathrm{Tr}\,\mathbf{R}(z) - \mathbb{E}\,\mathrm{Tr}\,\mathbf{R}(z)\right|^4 \le \frac{C}{n^2v^4} \tag{3.7}$$

and Borel-Cantelli lemma. Indeed, let us introduce the following σ-algebras:

$$\mathfrak{F}^{(k)} \stackrel{\mathsf{def}}{=} \sigma\{X_{jl}, j, l > k\}, \quad \mathfrak{F}^{(0)} \stackrel{\mathsf{def}}{=} \mathfrak{F}, \quad \mathfrak{F}^{(n)} \stackrel{\mathsf{def}}{=} \{\Omega, \emptyset\}.$$

Denote $\mathbb{E}_k(\cdot) \stackrel{\mathsf{def}}{=} \mathbb{E}(\cdot \big| \mathfrak{F}^{(k)})$. Then

$$\begin{aligned}\frac{1}{n}\operatorname{Tr}\mathbf{R}(z) - \frac{1}{n}\mathbb{E}\operatorname{Tr}\mathbf{R}(z) &= \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}_{k-1}\operatorname{Tr}\mathbf{R}(z) - \mathbb{E}_k\operatorname{Tr}\mathbf{R}(z)\\ &= \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}_{k-1}[\operatorname{Tr}\mathbf{R}(z) - \operatorname{Tr}\mathbf{R}^{(k)}(z)] - \mathbb{E}_k[\operatorname{Tr}\mathbf{R}(z) - \operatorname{Tr}\mathbf{R}^{(k)}(z)]\\ &= \sum_{k=1}^{n}\eta_k,\end{aligned}$$

where $\mathbf{R}^{(k)}$ is $\mathbf{R}$ with k-th row and column deleted from $\mathbf{X}$, and $\eta = (\eta_k, \mathfrak{F}^{(k)})_{k\geq 0}$ is a martingale-difference. Show that

$$|\operatorname{Tr}\mathbf{R}(z) - \operatorname{Tr}\mathbf{R}^{(k)}(z)| \leq \frac{1}{v}. \tag{3.8}$$

Apply Lemma 3.6 to prove (3.7).

3.4. Covariance matrices

Let $X, X_1, ..., X_n$ be i.i.d. random vectors in $\mathbb{R}^p$ with $\mathbb{E}\,X = 0, \operatorname{Cov} X = \mathbf{I}$. Recall the definition (2.6) of sample covariance matrix $\mathbf{\Sigma}$

$$\widehat{\mathbf{\Sigma}} \stackrel{\mathsf{def}}{=} \frac{1}{n}\sum_{j=1}^{n} X_j X_j^{\mathsf{T}} = \frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{T}},$$

where $\mathbf{X} \stackrel{\mathsf{def}}{=} [X_1, \ldots, X_n] \in \mathbb{R}^{p\times n}$. In statistical applications, the true covariance matrix $\mathbf{\Sigma}$ is typically unknown and one often uses the sample covariance matrix $\widehat{\mathbf{\Sigma}}$ as its estimator.

Let $\lambda_1(\widehat{\mathbf{\Sigma}}) \leq \lambda_2(\widehat{\mathbf{\Sigma}}) \leq \ldots \leq \lambda_p(\widehat{\mathbf{\Sigma}})$ be the eigenvalues of $\widehat{\mathbf{\Sigma}}$ (squared singular values of $\mathbf{W} \stackrel{\mathsf{def}}{=} \frac{1}{\sqrt{n}}\mathbf{X}$). Define

$$N_I(\widehat{\mathbf{\Sigma}}) \stackrel{\mathsf{def}}{=} |j : \lambda_j(\widehat{\mathbf{\Sigma}}) \in I|.$$

Theorem 3.8. *Let $X, X_1, \ldots, X_n$ be i.i.d. random vectors taking values in $\mathbb{R}^p$ with $\mathbb{E}\,X = 0, \operatorname{Cov} X = \mathbf{I}$. Assume that $p = p(n)$ and $\lim_{n\to\infty}\frac{p}{n} = c \in [0,1]$. Then a.s.*

$$\lim_{n\to\infty}\frac{1}{p}N_I(\widehat{\mathbf{\Sigma}}) = \int_I p_c(\lambda)\,d\lambda,$$

where $p_c(x)$ is the density of the Marchenko–Pastur law

$$p_c(\lambda) \stackrel{\mathsf{def}}{=} \frac{1}{2\pi\lambda c}\sqrt{(\lambda-a)(b-\lambda)}\mathbb{I}[\lambda\in[a,b]]$$

where $a = (1-\sqrt{c})^2, b = (1+\sqrt{c})^2$.

On Figure 2 we plotted p_ρ for different c.

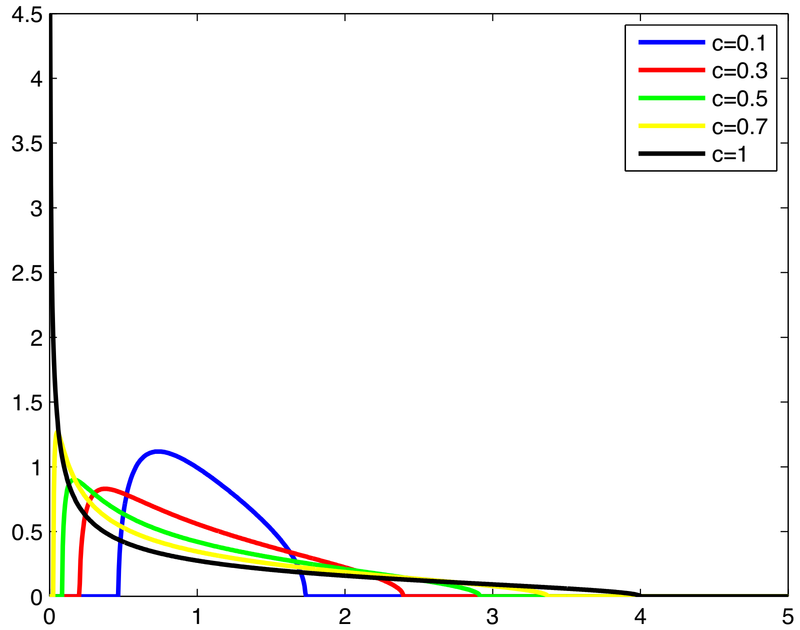

FIGURE 2. MP distribution for different c.

Bibliography

[AGZ10] G. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010. MR 2760897 (2011m:60016)

[Arn67] L. Arnold, *On the asymptotic distribution of the eigenvalues of random matrices*, J. Math. Anal. Appl. **20** (1967), 262–268. MR 0217833 (36 #922)

[Ben03] V. Bentkus, *A new approach to approximations in probability theory and operator theory*, Liet. Mat. Rink. **43** (2003), no. 4, 444–470. MR 2058566 (2005a:60045)

[BMP15] M. Banna, F. Merlevede, and M. Peligrad, *On the limiting spectral distribution for a large class of symmetric random matrices with correlated entries*, Stoch. Proc. and their Appl. **125** (2015), 2700–2726.

[BS10] Z. Bai and J. Silverstein, *Spectral analysis of large dimensional random matrices*, second ed., Springer, New York, 2010. MR 2567175 (2011d:60014)

[CGS11] L. Chen, L. Goldstein, and Q.M. Shao, *Normal approximation by Stein's method*, Springer (2011).

[Gir85] V. Girko, *Spectral theory of random matrices*, Uspekhi Mat. Nauk **40** (1985), no. 1(241), 67–106, 256. MR 783605 (86m:60094)

[Gir90] ______, *Theory of random determinants*, Mathematics and its Applications (Soviet Series), vol. 45, Kluwer Academic Publishers Group, Dordrecht, 1990, Translated from the Russian. MR 1080966 (91k:60001)

[GNT15] F. Götze, A. Naumov, and A. Tikhomirov, *Limit theorems for two classes of random matrices with dependent entries*, Theory Probab. Appl. **59 (114)** (2015), no. 1, 23–39.

[GT06] F. Götze and A. Tikhomirov, *Limit theorems for spectra of random matrices with martingale structure*, Teor. Veroyatn. Primen. **51** (2006), no. 1, 171–192. MR 2324173 (2008d:60052)

[KL15] V. Koltchinskii and K. Lounici, *Concentration inequalities and moment bounds for sample covariance operators*, arXiv:1405.2468 (2015).

[Meh91] Madan Lal Mehta, *Random matrices*, second ed., Academic Press Inc., Boston, MA, 1991. MR 1083764 (92f:82002)

[Nau14] A. Naumov, *Limit theorems for two classes of random matrices with Gaussian elements*, Journal of Mathematical Sciences **204** (2014), no. 1, 140–147.

[Pas73] L. Pastur, *Spectra of random selfadjoint operators*, Uspehi Mat. Nauk **28** (1973), no. 1(169), 3–64. MR 0406251 (53 #10042)

[Shl96] D. Shlyakhtenko, *Random gaussian band matrices and freeness with amalgamation*, International Mathematics Research Notices (1996), no. 20, 1014–1025.

[Tro12] J. Tropp, *User-friendly tail bounds for sums of random matrices*, Found. Comput. Math. **12** (2012), no. 4, 389–434. MR 2946459

[Ver12] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*, Compressed sensing, Cambridge Univ. Press, Cambridge, 2012, pp. 210–268. MR 2963170

[vH15] Ramon van Handel, *On the spectral norm of gaussian random matrices*, 2015, arXiv:1502.05003.

[Wig55] E. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. of Math. (2) **62** (1955), 548–564. MR 0077805 (17,1097c)

Appendix A
Some Results in Linear Algebra

In this chapter, the reader is assumed to have a college-level knowledge of linear algebra. Therefore, we only introduce those results that will be used in this book.

A.1 Inverse Matrices and Resolvent

A.1.1 Inverse Matrix Formula

Let $\mathbf{A} = (a_{ij})$ be an $n \times n$ matrix. Denote the cofactor of a_{ij} by A_{ij}. The Laplace expansion of the determinant states that, for any j,

$$\det(\mathbf{A}) = \sum_{i=1}^{n} a_{ij} A_{ij}. \tag{A.1.1}$$

Let $\mathbf{A}^a = (A_{ij})'$ denote the adjacent matrix of $\mathbf{A}$. Then, applying the formula above, one immediately gets

$$\mathbf{A}\mathbf{A}^a = \det(\mathbf{A})\mathbf{I}_n.$$

This proves the following theorems.

Theorem A.1. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix with a nonzero determinant. Then, it is invertible and*

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})}\mathbf{A}^a. \tag{A.1.2}$$

Theorem A.2. *We have*

$$\mathrm{tr}(\mathbf{A}^{-1}) = \sum_{k=1}^{n} A_{kk}/\det(\mathbf{A}). \tag{A.1.3}$$

A.1.2 Holing a Matrix

The following is known as Hua's holing method:

$$\begin{pmatrix} \mathbf{I} & \mathbf{O} \\ -\mathbf{C}\mathbf{A}^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{O} & \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \end{pmatrix}. \tag{A.1.4}$$

In application, this formula can be considered as making a row Gaussian elimination on the matrix $\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}$ to eliminate the (2,1)-th block. A similar column transformation also holds. An important application of this formula is the following theorem.

Theorem A.3. *If $\mathbf{A}$ is a squared nonsingular matrix, then*

$$\det \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \det(\mathbf{A})\det(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}). \tag{A.1.5}$$

This theorem follows by taking determinants on both sides of (A.1.4).

Note that the transformation (A.1.4) does not change the rank of the matrix. Therefore, it is frequently used to prove rank inequalities.

A.1.3 Trace of an Inverse Matrix

For $n \times n$ $\mathbf{A}$, define $\mathbf{A}_k$, called a major submatrix of order $n-1$, to be the matrix resulting from deleting the k-th row and column from $\mathbf{A}$. Applying (A.1.2) and (A.1.5), we obtain the following useful theorem.

Theorem A.4. *If both $\mathbf{A}$ and $\mathbf{A}_k$, $k = 1, 2, \cdots, n$, are nonsingular, and if we write $\mathbf{A}^{-1} = \left[a^{k\ell}\right]$, then*

$$a^{kk} = \frac{1}{a_{kk} - \boldsymbol{\alpha}_k' \mathbf{A}_k^{-1} \boldsymbol{\beta}_k}, \tag{A.1.6}$$

and hence

$$\mathrm{tr}(\mathbf{A}^{-1}) = \sum_{k=1}^{n} \frac{1}{a_{kk} - \boldsymbol{\alpha}_k' \mathbf{A}_k^{-1} \boldsymbol{\beta}_k}, \tag{A.1.7}$$

where a_{kk} is the k-th diagonal entry of $\mathbf{A}$, $\mathbf{A}_k$ is defined above, $\boldsymbol{\alpha}_k'$ is the vector obtained from the k-th row of $\mathbf{A}$ by deleting the k-th entry, and $\boldsymbol{\beta}_k$ is the vector from the k-th column by deleting the k-th entry.

If $\mathbf{A}$ is an $n \times n$ symmetric nonsingular matrix and all its major submatrices of order $(n-1)$ are nonsingular, then from (A.1.7) it follows immediately that

$$\operatorname{tr}(\mathbf{A}^{-1}) = \sum_{k=1}^{n} \frac{1}{a_{kk} - \boldsymbol{\alpha}_k' \mathbf{A}_k^{-1} \boldsymbol{\alpha}_k}. \tag{A.1.8}$$

If $\mathbf{A}$ is an $n \times n$ Hermitian nonsingular matrix and all its major submatrices of order $(n-1)$ are nonsingular, similarly we have

$$\operatorname{tr}(\mathbf{A}^{-1}) = \sum_{k=1}^{n} \frac{1}{a_{kk} - \boldsymbol{\alpha}_k^* \mathbf{A}_k^{-1} \boldsymbol{\alpha}_k},$$

where * denotes the complex conjugate transpose of matrices or vectors.

In this book, we shall frequently consider the resolvent of a Hermitian matrix $\mathbf{X} = (x_{jk})$ (i.e., $\mathbf{A} = (\mathbf{X} - z\mathbf{I})^{-1}$), where z is a complex number with positive imaginary part. In this case, we have

$$\operatorname{tr}((\mathbf{X} - z\mathbf{I})^{-1}) = \sum_{k=1}^{n} \frac{1}{x_{kk} - z - \mathbf{x}_k^* \mathbf{H}_k^{-1} \mathbf{x}_k}, \tag{A.1.9}$$

where $\mathbf{H}_k$ is the matrix obtained from $\mathbf{X} - z\mathbf{I}$ by deleting the k-th row and the k-th column and $\mathbf{x}_k$ is the k-th column of $\mathbf{X}$ with the k-th element removed.

A.1.4 Difference of Traces of a Matrix A *and Its Major Submatrices*

Suppose that the matrix $\boldsymbol{\Sigma}$ is positive definite and has the partition as given by $\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$. Then, the inverse of $\boldsymbol{\Sigma}$ has the form

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1} + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22.1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22.1}^{-1} \\ -\boldsymbol{\Sigma}_{22.1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \boldsymbol{\Sigma}_{22.1}^{-1} \end{pmatrix},$$

where $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{12}$. In fact, the formula above can be derived from the identity (by applying (A.1.4))

$$\begin{pmatrix} \mathbf{I} & \mathbf{O} \\ -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{I} & -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \\ \mathbf{O} & \mathbf{I} \end{pmatrix}$$
$$= \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \end{pmatrix}$$

and the fact that

$$\begin{pmatrix} \mathbf{I} & 0 \\ -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{I} & \mathbf{O} \\ \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I} \end{pmatrix}.$$

Making use of this identity, we obtain the following theorem.

Theorem A.5. *If the matrix* $\mathbf{A}$ *and* $\mathbf{A}_k$*, the* k*-th major submatrix of* $\mathbf{A}$ *of order* $(n-1)$*, are both nonsingular and symmetric, then*

$$\operatorname{tr}(\mathbf{A}^{-1}) - \operatorname{tr}(\mathbf{A}_k^{-1}) = \frac{1 + \boldsymbol{\alpha}_k' \mathbf{A}_k^{-2} \boldsymbol{\alpha}_k}{a_{kk} - \boldsymbol{\alpha}_k' \mathbf{A}_k^{-1} \boldsymbol{\alpha}_k}. \tag{A.1.10}$$

If $\mathbf{A}$ *is Hermitian, then* α_k' *is replaced by* α_k^* *in the equality above.*

A.1.5 Inverse Matrix of Complex Matrices

Theorem A.6. *If Hermitian matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *are commutative and such that* $\mathbf{A}^2 + \mathbf{B}^2$ *is nonsingular, then the complex matrix* $\mathbf{A} + i\mathbf{B}$ *is nonsingular and*

$$(\mathbf{A} + i\mathbf{B})^{-1} = (\mathbf{A} - i\mathbf{B})(\mathbf{A}^2 + \mathbf{B}^2)^{-1}. \tag{A.1.11}$$

This can be directly verified.

Let $z = u + iv$, $v > 0$, and let $\mathbf{A}$ be an $n \times n$ Hermitian matrix. Then

$$\left|\operatorname{tr}(\mathbf{A} - z\mathbf{I}_n)^{-1} - \operatorname{tr}(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-1}\right| \le v^{-1}. \tag{A.1.12}$$

Proof. By (A.1.10), we have

$$\operatorname{tr}(\mathbf{A} - z\mathbf{I}_n)^{-1} - \operatorname{tr}(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-1} = \frac{1 + \boldsymbol{\alpha}_k^*(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-2}\boldsymbol{\alpha}_k}{a_{kk} - z - \boldsymbol{\alpha}_k^*(\mathbf{A} - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k}.$$

If we denote $\mathbf{A}_k = \mathbf{E}^* \operatorname{diag}[\lambda_1 \cdots \lambda_{n-1}]\mathbf{E}$ and $\boldsymbol{\alpha}_k^* \mathbf{E}^* = (y_1, \cdots, y_{n-1})$, where $\mathbf{E}$ is an $(n-1) \times (n-1)$ unitary matrix, then we have

$$\begin{aligned}
|1 + \boldsymbol{\alpha}_k^*(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-2}\boldsymbol{\alpha}_k| &= |1 + \sum_{\ell=1}^{n-1} |y_\ell^2|(\lambda_\ell - z)^{-2}| \\
&\le 1 + \sum_{\ell=1}^{n-1} |y_\ell^2|((\lambda_\ell - u)^2 + v^2)^{-1} \\
&= 1 + \boldsymbol{\alpha}_k^*((\mathbf{A}_k - u\mathbf{I}_{n-1})^2 + v^2\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k.
\end{aligned}$$

On the other hand, by (A.1.11) we have

$$\begin{aligned}
&\Im(a_{kk} - z - \boldsymbol{\alpha}_k^*(\mathbf{A} - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k) \\
&= v(1 + \boldsymbol{\alpha}_k^*((\mathbf{A}_k - u\mathbf{I}_{n-1})^2 + v^2\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k).
\end{aligned} \tag{A.1.13}$$

From these estimates, (A.1.12) follows.

Making use of this identity, we obtain the following theorem.

Theorem A.5. *If the matrix* $\mathbf{A}$ *and* $\mathbf{A}_k$*, the k-th major submatrix of* $\mathbf{A}$ *of order* $(n-1)$*, are both nonsingular and symmetric, then*

$$\operatorname{tr}(\mathbf{A}^{-1}) - \operatorname{tr}(\mathbf{A}_k^{-1}) = \frac{1 + \boldsymbol{\alpha}_k' \mathbf{A}_k^{-2} \boldsymbol{\alpha}_k}{a_{kk} - \boldsymbol{\alpha}_k' \mathbf{A}_k^{-1} \boldsymbol{\alpha}_k}. \tag{A.1.10}$$

If $\mathbf{A}$ *is Hermitian, then* α_k' *is replaced by* α_k^* *in the equality above.*

A.1.5 Inverse Matrix of Complex Matrices

Theorem A.6. *If Hermitian matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *are commutative and such that* $\mathbf{A}^2 + \mathbf{B}^2$ *is nonsingular, then the complex matrix* $\mathbf{A} + i\mathbf{B}$ *is nonsingular and*

$$(\mathbf{A} + i\mathbf{B})^{-1} = (\mathbf{A} - i\mathbf{B})(\mathbf{A}^2 + \mathbf{B}^2)^{-1}. \tag{A.1.11}$$

This can be directly verified.

Let $z = u + iv$, $v > 0$, and let $\mathbf{A}$ be an $n \times n$ Hermitian matrix. Then

$$\left|\operatorname{tr}(\mathbf{A} - z\mathbf{I}_n)^{-1} - \operatorname{tr}(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-1}\right| \le v^{-1}. \tag{A.1.12}$$

Proof. By (A.1.10), we have

$$\operatorname{tr}(\mathbf{A} - z\mathbf{I}_n)^{-1} - \operatorname{tr}(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-1} = \frac{1 + \boldsymbol{\alpha}_k^*(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-2}\boldsymbol{\alpha}_k}{a_{kk} - z - \boldsymbol{\alpha}_k^*(\mathbf{A} - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k}.$$

If we denote $\mathbf{A}_k = \mathbf{E}^* \operatorname{diag}[\lambda_1 \cdots \lambda_{n-1}]\mathbf{E}$ and $\boldsymbol{\alpha}_k^* \mathbf{E}^* = (y_1, \cdots, y_{n-1})$, where $\mathbf{E}$ is an $(n-1) \times (n-1)$ unitary matrix, then we have

$$\begin{aligned} |1 + \boldsymbol{\alpha}_k^*(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-2}\boldsymbol{\alpha}_k| &= |1 + \sum_{\ell=1}^{n-1} |y_\ell^2|(\lambda_\ell - z)^{-2}| \\ &\le 1 + \sum_{\ell=1}^{n-1} |y_\ell^2|((\lambda_\ell - u)^2 + v^2)^{-1} \\ &= 1 + \boldsymbol{\alpha}_k^*((\mathbf{A}_k - u\mathbf{I}_{n-1})^2 + v^2\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k. \end{aligned}$$

On the other hand, by (A.1.11) we have

$$\begin{aligned} &\Im(a_{kk} - z - \boldsymbol{\alpha}_k^*(\mathbf{A} - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k) \\ &= v(1 + \boldsymbol{\alpha}_k^*((\mathbf{A}_k - u\mathbf{I}_{n-1})^2 + v^2\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k). \end{aligned} \tag{A.1.13}$$

From these estimates, (A.1.12) follows.