Statistica Sinica **24** (2014), 919-936
doi:http://dx.doi.org/10.5705/ss.2012.130

A LOCAL MOMENT ESTIMATOR OF THE SPECTRUM OF A LARGE DIMENSIONAL COVARIANCE MATRIX

Weiming Li[1,2] and Jianfeng Yao[3]

[1]*Beihang University,* [2]*Beijing University of Posts and Telecommunications and* [3]*The University of Hong Kong*

Abstract: This paper considers the problem of estimating the population spectral distribution from a sample covariance matrix when its dimension is large. We generalize the contour-integral based method in Mestre (2008) and present a local moment estimation procedure. Compared with the original, the new procedure can be applied successfully to models where the asymptotic clusters of sample eigenvalues generated by different population eigenvalues are not all separate. The proposed estimates are proved to be consistent. Numerical results illustrate the implementation of the estimation procedure and demonstrate its efficiency in various cases.

Key words and phrases: Empirical spectral distribution, large covariance matrix, moment estimation, population spectral distribution, Stieltjes transform.

1. Introduction

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be a sequence of i.i.d. zero-mean random vectors in $\mathbb{R}^p$ or $\mathbb{C}^p$, with a common population covariance matrix Σ_p. When the population size p is not negligible with respect to the sample size n, modern random matrix theory indicates that the sample covariance matrix $S_n = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^*/n$ does not approach Σ_p. Therefore, classical statistical procedures based on an approximation of Σ_p by S_n become inconsistent.

The *spectral distribution* (SD) F^A of an $m \times m$ Hermitian matrix (or real symmetric) A is the measure generated by its eigenvalues $\{\lambda_i^A\}$,

$$F^A = \frac{1}{m} \sum_{i=1}^m \delta_{\lambda_i^A} \ ,$$

where δ_b denotes the Dirac point measure at b. Denote by $(\sigma_i)_{1 \le i \le p}$ the p eigenvalues of Σ_p. We are particularly interested in the SD

$$H_p := F^{\Sigma_p} = \frac{1}{p} \sum_{i=1}^p \delta_{\sigma_i}.$$

In large-dimensional frameworks, both dimensions p and n grow to infinity. It is then natural to assume that H_p converges weakly to a limit H. Both the SD H_p

and its limit H are referred as the *population spectral distribution* (PSD) of the observation model.

The main observation is that for large-dimensional data, the empirical SD (ESD) $F_n := F^{S_n}$ of S_n is far from the PSD H_p. Indeed, under reasonable assumptions, when both dimensions p and n grow proportionally, almost surely, the ESD F_n weakly converges to a deterministic distribution F, which in general has no explicit form but is linked to the PSD H via the so-called Marčenko-Pastur equation, see Marčenko and Pastur (1967), Silverstein (1995), Silverstein and Bai (1995), and Section 2.1.

A natural question here is the recovery of the PSD H_p (or its limit H) from the ESD F_n. This question is of central importance in such statistical methodologies as principal component analysis Johnstone (2001) and factor analysis that rely on efficient estimations of some population covariance matrices.

El Karoui (2008) proposed a nonparametric approach by solving the Marčenko-Pastur equation on the upper complex plane, and then obtained consistent estimates of H. Rao et al. (2008) investigated the asymptotic distributions of the moments of the ESD F_n and introduced a Gaussian likelihood to get consistent estimates of H. In the work of Mestre (2008), each mass of a discrete PSD H was represented by a contour integral under a certain eigenvalue splitting condition and consistent estimators of H were then obtained. Recently, Bai, Chen, and Yao (2010) modified the approach in Rao et al. (2008) and turned it to a fully moments-based procedure. Beyond consistency, the authors proved a central limit theorem for the estimator. Li et al. (2013) synthesized the optimization approach in El Karoui (2008) and the parametric setup in Bai, Chen, and Yao (2010); they changed the optimization problem from the complex plane to the real line by considering the extension of the Stieltjes transform on the real line.

The contour-integral based method in Mestre (2008) is well known for its high efficiency and easy computation. However, it is limited to a small class of discrete PSDs where the imposed eigenvalue splitting condition states that distinct population eigenvalues should generate non-overlapping clusters of sample eigenvalues. This method has been recently employed for subspace estimation in a so-called "information plus noise" model in Hachem et al. (2012).

Our purpose in this paper is to extend Mestre's method to situations where the splitting condition may not be satisfied. For a discrete PSD H with finite support on $\mathbb{R}^+$, one separate interval of the support S_F of the limiting SD (LSD) F corresponds to only one atom of H if the dimension ratio c is close to zero (the splitting condition holds). When c is increased gradually, adjacent intervals of S_F grow closer, and some may merge into a larger interval (the splitting condition fails). Such merged intervals correspond to more than one atom of H, and establishing their relationship gives rise to our local estimation method.

Our strategy is to first divide the PSD H into a number of sub-probability measures: $H_1, \ldots, H_m$, such that each H_i corresponds to a separate interval of S_F. We then develop a method to approximate the moments of H_i; an estimate of H_i can be obtained by solving a system of moment equations. Collecting these estimates produces an estimator of H. It is shown that, when m is equal to the number of atoms of H, this estimator reduces to the one in Mestre (2008); if $m = 1$, the estimator is equivalent to the one in Bai, Chen, and Yao (2010).

The rest of the paper is organized as follows. In the next section, we review some useful results from Random Matrix Theory and introduce the division of a PSD H according to the separation of the corresponding LSD F. A fast algorithm to solve the associated moment equations is also given. In Section 3, we present the theoretical supports for and the detailed procedure of our estimation. In Section 4, simulation experiments are carried out to compare our estimator with the estimator in Mestre (2008) and the moment estimator in Bai, Chen, and Yao (2010). Some conclusions and remarks are presented in Section 5.

2. Limiting Spectral Distribution and Division of a PSD H

2.1. The Marčenko-Pastur equation

The Stieltjes transform of G, a measure supported on the real line, is

$$s_G(z) = \int \frac{1}{x - z} dG(x), \quad z \in \mathbb{C}^+,$$

where $\mathbb{C}^+$ is the set of complex numbers with positive imaginary part. Let S_G be the support set of G and S_G^c its complementary set. We need to extend the Stieltjes transform to $\mathbb{C} \setminus S_G$ by

$$s(z) = \begin{cases} s^*(z^*) & (z \in \mathbb{C}^- = \{z \in \mathbb{C} : \Im(z) < 0\}), \\ \lim\limits_{\varepsilon \to 0^+} s(x + \varepsilon \mathrm{i}) & (z = x \in \mathbb{R} \setminus S_G), \end{cases}$$

where a^* denotes the complex conjugate of a. The existence of the limit in the second term follows from the Dominated Convergence Theorem.

Denote by $\lambda_1 \leq \cdots \leq \lambda_p$ the eigenvalues of the sample covariance matrix S_n. Then the ESD F_n of S_n is

$$F_n = \frac{1}{p} \sum_{i=1}^{p} \delta_{\lambda_i},$$

whose Stieltjes transform is

$$s_n(z) = \int \frac{1}{x - z} dF_n(x) = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{\lambda_i - z}.$$

We use a convergence result of Silverstein (1995).

Lemma 1. *Suppose the entries of $X_n(p \times n)$ are complex random variables that are independent for each n and identically distributed for all n, and that $\mathrm{E}(x_{11}) = 0$ and $\mathrm{E}(|x_{11}|^2) = 1$. Suppose T_n is a $p \times p$ random Hermitian nonnegative definite matrix, independent of X_n, and that the empirical distribution F^{T_n} converges almost surely to a probability measure H on $[0, \infty)$ as $n \to \infty$. If $B_n = T_n^{1/2} X_n X_n^* T_n^{1/2}/n$, and $p = p(n)$ satisfies $p/n \to c > 0$ as $n \to \infty$, then, almost surely, the empirical spectral distribution $F^{\mathrm{B_n}}$ converges in distribution, as $n \to \infty$, to a (non-random) probability measure F, whose Stieltjes transform $s = s(z)$ is a solution to the equation*

$$s = \int \frac{1}{t(1 - c - czs) - z} dH(t). \tag{2.1}$$

The solution is unique in the set $\{s \in \mathbb{C} : -(1-c)/z + cs \in \mathbb{C}^+\}$.

It is more convenient to use a companion distribution $\underline{F}_n = (1 - p/n)\delta_0 + (p/n)F_n$ with Stiletjes transform

$$\underline{s}_n(z) = -\frac{1 - p/n}{z} + \frac{p}{n} s_n(z) = -\frac{1 - p/n}{z} + \frac{1}{n} \sum_{i=1}^{p} \frac{1}{\lambda_i - z}.$$

The corresponding limit is $\underline{s}(z) = -(1-c)/z + cs(z)$ and it satisfies a variant of (2.1),

$$z = -\frac{1}{\underline{s}} + c \int \frac{t}{1 + t\underline{s}} dH(t). \tag{2.2}$$

Both (2.1) and (2.2) are referred as the Marčenko-Pastur equation.

Since the convergence in distribution of probability measures implies the pointwise convergence of the associated Stieltjes transforms, by Lemma 1, $\underline{s}_n(z)$ converges to $\underline{s}(z)$ almost surely, for any $z \in \mathbb{C} \setminus \mathbb{R}$. In Silverstein and Choi (1995), the convergence is extended to $S_F \setminus \{0\}$, and thus we conclude that for sufficiently large n, $\underline{s}_n(z)$ converges to $\underline{s}(z)$ almost surely for every $z \in \mathbb{C} \setminus (S_F \cup \{0\})$.

2.2. Division of a PSD H

Our method relies on a division of a PSD H according to the separation of the corresponding LSD F. Suppose the support S_F of F consists of m ($m \geq 1$) disjoint compact intervals, $S_1 = [x_1^-, x_1^+], \ldots, S_m = [x_m^-, x_m^+]$, sorted in an increasing order. Choose δ_i^-, δ_i^+ $(i = 1, \ldots, m)$ satisfying

$$\delta_1^- < x_1^- < x_1^+ < \delta_1^+ < \delta_2^- < \cdots < \delta_{m-1}^+ < \delta_m^- < x_m^- < x_m^+ < \delta_m^+. \tag{2.3}$$

Notice that when $z = x$ is restricted to S_F^c, $u(x) = -1/\underline{s}(x)$ is monotonically increasing and takes values in S_H^c Silverstein and Choi (1995). Then

$$u(\delta_1^-) < u(\delta_1^+) < u(\delta_2^-) < \cdots < u(\delta_{m-1}^+) < u(\delta_m^-) < u(\delta_m^+),$$

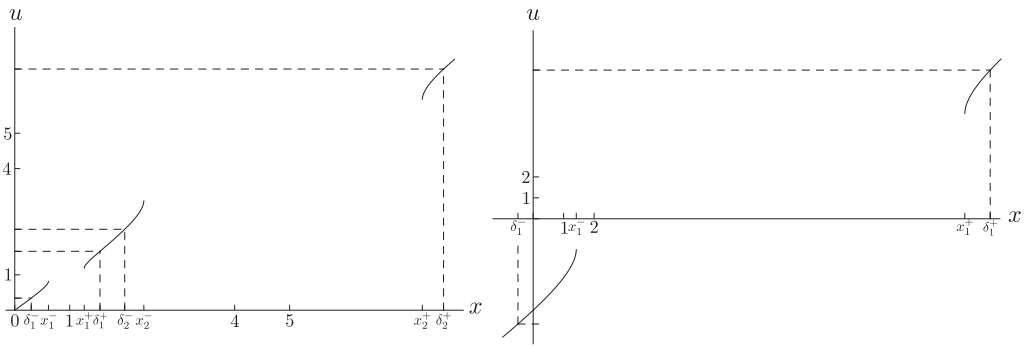

Figure 1. The curves of $u(x)$ on $S_F^c \cap \mathbb{R}^+$, with $H_1 = 0.3\delta_1 + 0.4\delta_4 + 0.3\delta_5$ and $c_1 = 0.1$ (left), and $H_2 = 0.5\delta_1 + 0.5\delta_2$ and $c_2 = 4$ (right).

$$S_H \subset \bigcup_{i=1}^{m} \left[u(\delta_i^-), u(\delta_i^+)\right].$$

Consequently, we can match each compact interval of S_F with a disjoint part of S_H by

$$S_i \to S_H \cap [u(\delta_i^-), u(\delta_i^+)], \quad i = 1, \ldots, m, \tag{2.4}$$

and hence, the PSD H admits a division as

$$H_i(A) = \int_{[u(\delta_i^-), u(\delta_i^+)] \cap A} dH, \quad A \in \mathcal{B}, \quad i = 1, \ldots, m,$$

where $\mathcal{B}$ is the class of Borel sets of $\mathbb{R}$. Obviously, $\sum_{i=1}^m H_i = H$.

The map in (2.4) can be easily found from the graph of $u(x)$ on S_F^c. Two typical representations of the graph are shown in Figure 1. The figures show that when $c < 1$, each compact interval of S_F corresponds to masses of H that fall within this interval; this is not true when $c > 1$, as shown in the right panel of Figure 1 where the mass 1 falls outside the interval $[x_1^-, x_1^+]$.

2.3. Moments of a discrete measure

Let be a discrete measure $G = \sum_{i=1}^k m_i \delta_{b_i}$ where $b_1 < \cdots < b_k$ are k masses with respective positive weights $\{m_i\}$. Here G can be a sub-probability measure. The lth moment of G is

$$\gamma_l = \sum_{i=1}^{k} m_i b_i^l, \quad l = 0, 1, \ldots,$$

and the Nth Hankel matrix related to G is

$$\Gamma(G, N) = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{N-1} \\ \gamma_1 & \gamma_2 & \cdots & \gamma_N \\ \vdots & \vdots & & \vdots \\ \gamma_{N-1} & \gamma_N & \cdots & \gamma_{2N-2} \end{pmatrix}.$$

Proposition 1. *The Hankel matrix* $\Gamma(G,k)$ *is positive definite, with determinant*

$$\det(\Gamma(G,k)) = \prod_{i=1}^{k} m_i \prod_{1\le i<j\le k} (b_i - b_j)^2. \tag{2.5}$$

Furthermore,

$$\det(\Gamma(G,N)) = 0, \quad N > k. \tag{2.6}$$

Proof. Write $M = \mathrm{diag}(m_1,\dots,m_k)$ and

$$B = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ b_1 & b_2 & \cdots & b_k \\ \vdots & \vdots & & \vdots \\ b_1^{k-1} & b_2^{k-1} & \cdots & b_k^{k-1} \end{pmatrix}. \tag{2.7}$$

Here B is a Vandermonde matrix whose determinant is $\prod_{1\le i<j\le k}(b_j - b_i)$. From this and the fact that $\Gamma(G,k) = BMB^T$, we get (2.5). Then (2.6) and the positive definiteness of $\Gamma(G,k)$ can be verified by a direct calculation.

Our aim is to find an efficient inversion formula for these moment equations; it will be the basis of our inference procedure. Now if

$$P(x) = \prod_{i=1}^{k}(x - b_i) = \sum_{i=0}^{k} c_i x^i, \quad c_k = 1,$$

then the coefficients of $P(x)$ and the moments of G have the following relationship.

Proposition 2. *If* $\mathbf{c} = (c_0,\dots,c_{k-1})'$ *and* $\boldsymbol{\gamma} = (\gamma_k,\dots,\gamma_{2k-1})'$, *then* $\Gamma(G,k)\cdot \mathbf{c} + \boldsymbol{\gamma} = \mathbf{0}$.

Proof. From $P(b_j) = \sum_{i=0}^{k} c_i b_j^i = 0$ for $j = 1,\dots,k$, we have $B'\mathbf{c} = -(b_1^k,\dots,b_k^k)'$, where the matrix B is defined at (2.7). From this and the decomposition of $\Gamma(G,k)$, we get

$$\begin{aligned} \Gamma(G,k)\cdot\mathbf{c} + \boldsymbol{\gamma} &= BMB'\mathbf{c} + \boldsymbol{\gamma} \\ &= \boldsymbol{\gamma} - BM(b_1^k,\dots,b_k^k)' \\ &= \mathbf{0}. \end{aligned}$$

Propositions 1 and 2 tell us that the masses of G are the zeros of $P(x)$ with coefficients $\mathbf{c} = -(\Gamma(G,k))^{-1}\cdot\boldsymbol{\gamma}$ and $c_k = 1$. The weights of G can be obtained by solving, with b_i's known,

$$\sum_{i=1}^{k} m_i b_i^l = \gamma_l, \quad l = 0,\dots,k-1.$$

3. Estimation

3.1 Model and estimation strategy

We consider a class of discrete PSDs with finite support on $\mathbb{R}^+$,

$$H(\boldsymbol{\theta}) = w_1\delta_{a_1} + \cdots + w_k\delta_{a_k}, \quad \boldsymbol{\theta} \in \Theta,$$

where

$$\Theta = \left\{\boldsymbol{\theta} = (a_1, w_1, \ldots, a_k, w_k) : 0 < a_1 < \cdots < a_k < \infty;\ w_i > 0,\ \sum_{i=1}^{k} w_i = 1\right\}.$$

Here, the order k of H is assumed known (when k is also to be estimated, a consistent estimator of k is given in Chen, Delyon, and Yao (2011)).

If the support S_F of the LSD F associated to H and c has m ($1 \le m \le k$) disjoint compact intervals, H can be divided into m parts, $H_1, \ldots, H_m$, with H_i consisting of k_i masses of H, $k_i \ge 1$ and $\sum_{i=1}^{m} k_i = k$. When the k_i's are all 1, one has the split case in Mestre (2008). We take the k_i's as unknown, possibly larger than 1, and not necessarily equal.

Our estimation strategy is the following.

(1) Determine the division of H according to the separation of clusters of sample eigenvalues.
(2) For each part H_i, obtain strongly consistent estimators of its moments.
(3) Obtain a strongly consistent estimator $\widehat{\mathbf{k}}_n$ of the partition $(k_1, \ldots, k_m)$ of numbers of masses in the m parts $H_1, \ldots, H_m$.
(4) By combination of these estimators, and using the method of moments, obtain consistent estimators of all the weights and masses (w_i, a_i).

Note that in the first step, an accurate division of H may not be always achieved, especially when sample sizes are relatively small. A solution to this problem will be given later.

3.2. Estimation of the moments of H_i

One can express the moments of H_i as contour integrals related to the companion Stieltjes transform $\underline{s}(z)$.

Theorem 1. *If the conditions of Lemma* 1 *hold, then the* l*th moment of* H_i *is*

$$\gamma_{i,l} = (-1)^l \frac{1}{c}\frac{1}{2\pi \mathrm{i}} \oint_{C_i} \frac{z\underline{s}'(z)}{\underline{s}^l(z)} dz, \quad l = 1, 2, \ldots, \tag{3.1}$$

where C_i *is a positively oriented contour described by the boundary of the rectangle*

$$\{z \in \mathbb{C} : \delta_i^- \le \Re(z) \le \delta_i^+, |\Im(z)| \le 1\},$$

where δ_i^-, δ_i^+ $(i = 1, \ldots, m)$ *are defined by* (2.3) *and* $\delta_1^- < 0$ *if* $c \ge 1$.

Proof. Let the image of C_i under $u(z) = 1/\underline{s}(z)$ be

$$u(C_i) = \{u(z) : z \in C_i\}.$$

Notice that $\underline{s}(z)$ is holomorphic on C_i. Then $u(C_i)$ is a simple closed curve taking values on $\mathbb{C} \setminus (S_H \cup \{0\})$. (The function $u(z) = -1/\underline{s}(z)$ is analytic on C_i and is a one-to-one map from C_i to its image $u(C_i)$. Thus, the two curves C_i and $u(C_i)$ are homeomorphic. Since C_i is simple and closed (homeomorphic to a unit circle in $\mathbb{C}$), its image is also simple and closed.) Moreover, since $\Im(u(z)) \neq 0$ for all z with $\Im(z) \neq 0$, we have $u(C_i) \cap \mathbb{R} = \{u(\delta_i^-), u(\delta_i^+)\}$ and $u(C_i)$ encloses $[u(\delta_i^-), u(\delta_i^+)]$. Therefore, $u(C_i)$ encloses only S_{H_i} and no other masses of H.

Applying this change of variable to the right hand side of (3.1), we have

$$\begin{aligned}
(-1)^l \frac{1}{c} \frac{1}{2\pi \mathrm{i}} \oint_{C_i} \frac{z\underline{s}'(z)}{\underline{s}^l(z)} dz &= \frac{1}{c} \frac{1}{2\pi \mathrm{i}} \oint_{u(C_i)} z(u) u^{l-2} du \\
&= \frac{1}{c} \frac{1}{2\pi \mathrm{i}} \oint_{u(C_i)} u^{l-1} + c \int \frac{t u^{l-1}}{u - t} dH(t) du \\
&= \frac{1}{2\pi \mathrm{i}} \int \oint_{u(C_i)} \frac{t u^{l-1}}{u - t} du dH(t) \\
&= \gamma_{i,l},
\end{aligned}$$

where the second equality is from the Marčenko-Pastur equation, and the last follows from the residue theorem.

By substituting the empirical Stieltjes transform $\underline{s}_n(z)$ for $\underline{s}(z)$ in (3.1), we get a natural estimator of $\gamma_{i,l}$:

$$\widehat{\gamma}_{i,l} = (-1)^l \frac{n}{p} \frac{1}{2\pi \mathrm{i}} \oint_{C_i} \frac{z\underline{s}_n'(z)}{\underline{s}_n^l(z)} dz, \quad l = 1, 2, \ldots. \tag{3.2}$$

Theorem 2. *If the conditions of Lemma* 1 *hold, for each* l $(l \geq 1)$, $\widehat{\gamma}_{i,l}$ *converges almost surely to* $\gamma_{i,l}$.

Proof. From the "no eigenvalues" conclusion in Bai and Silverstein (1998) we have, for sufficiently large n, $z\underline{s}_n'(z)/\underline{s}_n^l(z)$ as well as $z\underline{s}'(z)/\underline{s}^l(z)$ are continuous on C_i, and thus bounded on the contour. By the convergence of $\underline{s}_n(z)$ and the Dominated Convergence Theorem, almost surely,

$$\begin{aligned}
|\gamma_{i,l} - \widehat{\gamma}_{i,l}| &= \left| \oint_{C_i} \frac{z\underline{s}'(z)}{\underline{s}^l(z)} - \frac{z\underline{s}_n'(z)}{\underline{s}_n^l(z)} dz \right| \\
&\leq \oint_{C_i} \left| \frac{z\underline{s}'(z)}{\underline{s}^l(z)} - \frac{z\underline{s}_n'(z)}{\underline{s}_n^l(z)} \right| |dz| \\
&\to 0, \quad n \to \infty.
\end{aligned}$$

A technical issue here is the contour integration in (3.2). It can be calculated by the residue theorem; an algorithm is described in the Appendix.

3.3. Estimation of the partition $(k_1, \ldots, k_m)$

With $\mathbf{k} = (k_1, \ldots, k_m)'$ the vector of orders of H_i's, the collection of all possible values of $\mathbf{k}$ is

$$\mathbb{K} = \{\mathbf{k} : k_i \geq 1,\ \sum_{i=1}^{m} k_i = k\}.$$

Let $\mathbf{k}_0 = (k_{0,1}, \ldots, k_{0,m})'$ be the true value of $\mathbf{k}$. From Proposition 1, the smallest eigenvalue $\lambda_{\min}(\Gamma(H_i, k_i))$ of the Hankel matrix $\Gamma(H_i, k_i)$ is positive if $k_i \leq k_{0,i}$, and otherwise 0. Then

$$g(\mathbf{k}) = \min\left\{\lambda_{\min}(\Gamma(H_i, k_i)), \quad i = 1, \ldots, m\right\}, \quad \mathbf{k} \in \mathbb{K},$$

satisfies

$$g(\mathbf{k}_0) > 0 \quad \text{and} \quad g(\mathbf{k}) = 0 \quad (\mathbf{k} \neq \mathbf{k}_0).$$

An estimator of $\mathbf{k}_0$ can be obtained as

$$\begin{aligned}\widehat{\mathbf{k}}_n &= \arg\max_{\mathbf{k}\in\mathbb{K}} \widehat{g}(\mathbf{k}) \\ &= \arg\max_{\mathbf{k}\in\mathbb{K}} \min\left\{\lambda_{\min}(\widehat{\Gamma}(H_i, k_i)), \quad i = 1, \ldots, m\right\},\end{aligned}$$

where $\widehat{\Gamma}(H_i, k_i) = (\widehat{\gamma}_{i,r+s-2})_{1\leq r,\ s\leq k_i}$ with its entries defined by (3.2).

When evaluating the estimator $\widehat{\mathbf{k}}_n$, it is not necessary to compare $\widehat{g}(\mathbf{k})$'s at all $\mathbf{k}$-points. For the ith element k_i of $\mathbf{k}$, in theory, its value may range from 1 to $k - m + 1$ and its true value $k_{0,i}$ makes $\Gamma(H_i, k_i)$ positive definite. Thus if $\Gamma(H_i, k_i)$ is non-positive definite, $k_i \neq k_{0,i}$ (actually $k_i > k_{0,i}$). It is enough to consider k_i that belongs to a set $\{1, \ldots, d_i\}$, where $d_i \leq k - m + 1$ stands for the largest integer such that $\widehat{\Gamma}(H_i, d_i)$ is positive definite. This effectively reduces the computational burden when the cardinality of $\mathbb{K}$ is large.

Theorem 3. *Under the assumptions of Lemma* 1*, almost surely,* $\widehat{\mathbf{k}}_n \to \mathbf{k}_0$ *as* $n \to \infty$.

Proof. The conclusion follows from Theorem 2 and the fact that $\mathbf{k}_0$ is the unique maximizer of the function $g(\mathbf{k})$ on the finite set $\mathbb{K}$.

3.4. Estimation of $\boldsymbol{\theta}$

By Theorem 3 and the finiteness of $\mathbb{K}$ is finite, $\widehat{\mathbf{k}}_n = \mathbf{k}_0$ eventually, almost

surely. For consistency in estimation of $\boldsymbol{\theta}$, we may assume that the partition $\mathbf{k}$ is known. The estimator $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ is taken to be a solution of the equations

$$\int x^l dH_i(\boldsymbol{\theta}) = \widehat{\gamma}_{i,l}, \quad l = 0, \ldots, 2k_i - 1, \quad i = 1, \ldots, m, \tag{3.3}$$

where $\widehat{\gamma}_{i,0} = v_i/v$, $i = 1, \ldots, m$, (v is the total number of positive sample eigenvalues and v_i is the number of those forming the ith cluster). We call $\widehat{\boldsymbol{\theta}}_n$ the *local moment estimator* (LME) of $\boldsymbol{\theta}$. Accordingly, the LME of H is $\widehat{H} = H(\widehat{\boldsymbol{\theta}}_n)$. When $k_1 = \cdots = k_m = 1$, the LME reduces to the one in Mestre (2008).

The solution of (3.3) exists and is unique if the matrices $\widehat{\Gamma}(H_i, k_i)$'s are all positive definite. Moreover, a fast algorithm for the solution exists.

Theorem 4. *Let the conditions of Lemma* 1 *hold, and suppose that the true value* $\boldsymbol{\theta}_0$ *is an inner point of* Θ. *Then* $\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ *almost surely as* $n \to \infty$.

Proof. Write $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{\theta}}_{1n}, \ldots, \widehat{\boldsymbol{\theta}}_{mn})$, where $\widehat{\boldsymbol{\theta}}_{in}$ is the LME of the parameter vector $\boldsymbol{\theta}_{i0}$ of H_i ($i = 1, \ldots, m$). It is sufficient to prove that $\widehat{\boldsymbol{\theta}}_{in} \to \boldsymbol{\theta}_{i0}$ almost surely for each i ($i = 1, \ldots, m$). Let h_i be the function $R^{2k_i} \to R^{2k_i}$: $\boldsymbol{\theta}_i \mapsto \boldsymbol{\gamma}_i = (\gamma_{i,0}, \ldots, \gamma_{i,2k_i-1})$. The multivariate function h_i is invertible by Propositions 1 and 2.

Let $\widehat{\boldsymbol{\gamma}}_{in} = (\widehat{\gamma}_{i,0}, \ldots, \widehat{\gamma}_{i,2k_i-1})$ and $\boldsymbol{\gamma}_{i0} = h_i(\boldsymbol{\theta}_{i0})$. By the convergence of $\widehat{\boldsymbol{\gamma}}_{in}$ (Theorem 2) and the Implicit Function Theorem, there exists a neighborhood U_i of $\boldsymbol{\theta}_{i0}$ and a neighborhood V_i of $\boldsymbol{\gamma}_{i0}$, such that h_i is a differomorphism from U_i onto V_i. Moreover, $\widehat{\boldsymbol{\theta}}_{in} = h_i^{-1}(\widehat{\boldsymbol{\gamma}}_{in}) \in U_i$ exists almost surely for all large n. Therefore, $\widehat{\boldsymbol{\theta}}_{in}$ converges to $\boldsymbol{\theta}_{i0} = h_i^{-1}(\boldsymbol{\gamma}_{i0})$ almost surely, as $n \to \infty$.

3.5. A generalization of the local moment estimator

The proposed estimation procedure needs good judgment on the separation of clusters of sample eigenvalues when the sample size is small. To handle this problem, we introduce a generalized version of the estimation procedure, the resulting estimator is termed the generalized LME (GLME).

Suppose the support S_F has m (≥ 1) disjoint compact intervals and, accordingly, H gains a division of m parts: $H_1, \ldots, H_m$. Without loss of generality, we suppose that the first two clusters of sample eigenvalues have no clear separation. Our strategy is to simply merge these clusters and treat H_1 and H_2 as a whole. Then, the GMLE can be obtained by conducting a similar procedure of estimation as before.

If all clusters are merged into one, then the GLME becomes a "full moment" estimator equivalent to the moment estimator in Bai, Chen, and Yao (2010). However, the merging procedure may result in a reduction of estimation efficiency; this is illustrated numerically in the next section.

It can be easily shown that Theorems 1−3 still hold true after merging so we obtain the strong convergence of the GLME by a similar proof. The details are omitted.

4. Simulation

We report on simulations carried out to compare the performance of the proposed estimator with the estimator in Mestre (2008) (referred as ME), and the one in Bai, Chen, and Yao (2010) (referred as BCY). Samples were drawn from mean-zero normal distribution with $(p, n) = (320,\ 1{,}000)$ for the estimation of H, and $(p, n) = (320,\ 1{,}000)$, $(160,\ 500)$, $(64,\ 200)$, $(32,\ 100)$, and $(16,\ 50)$ for the estimation of the partition $\mathbf{k}$ of H. The independent replications numbered 1,000. More p/n combinations were considered for the partition estimator $\widehat{\mathbf{k}}_n$ since this step has a primary importance on the overall performance of the procedure.

To measure the distance between H and its estimate $\widehat{H}$, we took the Wasserstein distance $d = \int |Q_H(t) - Q_{\widehat{H}}(t)|dt$, $Q_\mu(t)$ the quantile function of a probability measure μ. Execution times are provided for one realization of $\widehat{H}$, in seconds. All programs were realized in Mathematica 8 software, and run on a PC equipped with 3.5GHk CPU and 8GB physical RAM.

We consider a case in Mestre (2008) with $H = 0.5\delta_1 + 0.25\delta_7 + 0.125\delta_{15} + 0.125\delta_{25}$ and $c = 0.32$. Here H has four atoms at 1, 7, 15, and 25, while the sample eigenvalues form three clusters, and spread over $S_F = [0.2615, 1.6935] \cup [3.2610, 10.1562] \cup [10.2899, 38.0931]$ in the limit, see Figure 2. In Mestre (2008), it was shown that the ME performed very well by assuming all weight parameters (multiplicities) being known even if the splitting condition did not hold.

In the LME method, the PSD H can only be divided into $H_1 = 0.5\delta_1$, $H_2 = 0.25\delta_7$, and $H_3 = 0.125\delta_{15} + 0.125\delta_{25}$, with true partition $\mathbf{k}_0 = (1, 1, 2)$. Table 1 presents the frequency of estimates of the partition $\mathbf{k}$. The results show that the true model can be identified with an accuracy of 100% when the sample size n is larger than 200.

Table 2 presents statistics for the three estimators of H. The first six rows are results assuming all the weights $\{w_i\}$ are known, while in the last four rows it is assumed only $\{w_1, w_2\}$ are known and w_3 is to be estimated (w_4 is determined by $\sum w_i = 1$). Overall, the LME is as good as the ME when all weights are known, and is much better than the BCY in all cases. When w_3 is unknown, one finds larger distance values of d for both LME and BCY. This is also reflected in larger variances of the estimates of a_3 and a_4. As to execution times, the BCY is the fastest followed by the ME, and then by the LME. However, the elapsed time of the BCY estimation increases rapidly as the number of unknown parameters increases.

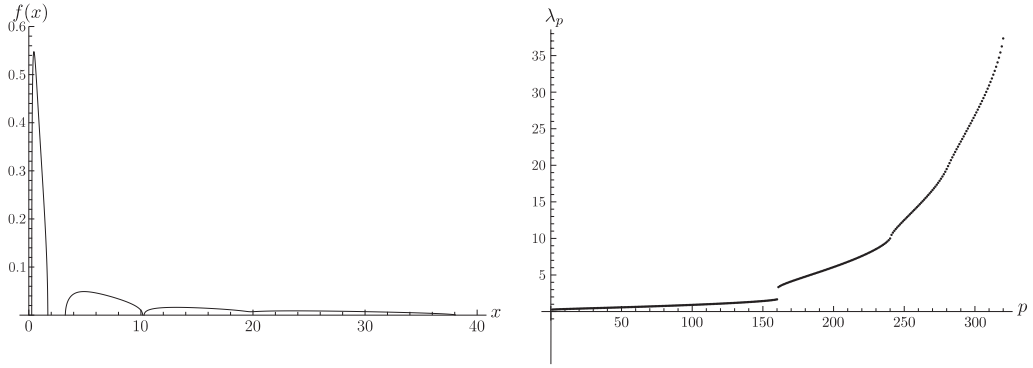

Figure 2. The density curve of F (left) and the average of the ith ($i = 1, \ldots, 320$) sample eigenvalues (right) from 1,000 replications for $H = 0.5\delta_1 + 0.25\delta_7 + 0.125\delta_{15} + 0.125\delta_{25}$ and $c = 0.32$.

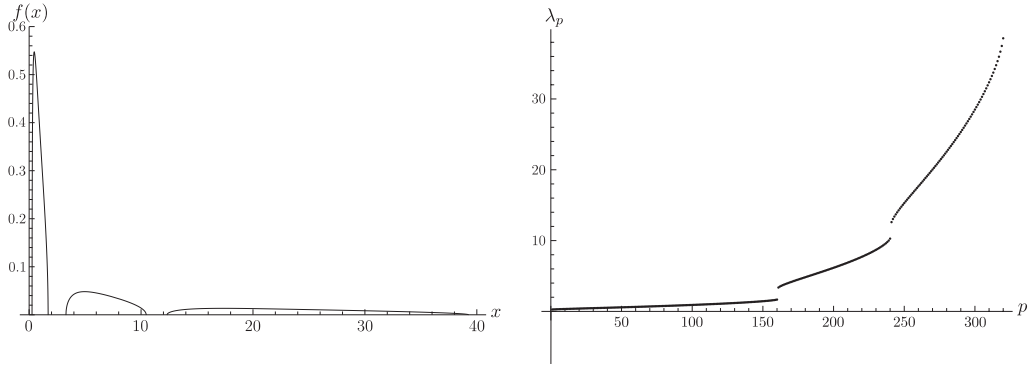

Figure 3. The density curve of F (left) and the average of the ith ($i = 1, \ldots, 320$) sample eigenvalues (right) from 1,000 replications for $H = 0.5\delta_1 + 0.25\delta_7 + 0.125\delta_{20} + 0.125\delta_{25}$ and $c = 0.32$.

Table 1. Frequency of estimates for the partition of H: $H_1 = 0.5\delta_1$, $H_2 = 0.25\delta_7$, $H_3 = 0.125\delta_{15} + 0.125\delta_{25}$ with $p/n = 0.32$.

Dimensions	$\mathbf{k}=(1,1,2)'$	$\mathbf{k}=(1,2,1)'$	$\mathbf{k}=(2,1,1)'$
$(p,n) = (320, 1000)$	1000	0	0
$(p,n) = (160, 500)$	1000	0	0
$(p,n) = (64, 200)$	999	0	1
$(p,n) = (32, 100)$	896	45	59
$(p,n) = (16, 50)$	623	169	208

In general, when the splitting condition is not satisfied the performance of the ME decreases, and the estimates may suffer from large biases. We examine the performances of the LME and the BCY in such situations. We consider a

Table 2. Estimates for $H = 0.5\delta_1 + 0.25\delta_7 + 0.125\delta_{15} + 0.125\delta_{25}$ with $p = 320$ and $n = 1{,}000$.

		a_1	a_2	a_3	w_3	a_4	d	Time
ME	Mean	1.0000	7.0031	14.9987	-	25.0001	0.0425	0.533s
	St. D.	0.0041	0.0407	0.1368	-	0.1964	0.0199	
LME	Mean	1.0000	7.0060	14.9533	-	25.0381	0.0447	0.578s
	St. D.	0.0040	0.0401	0.1371	-	0.2033	0.0205	
BCY	Mean	0.9924	7.0387	14.8968	-	25.0658	0.0887	0.147s
	St. D.	0.0189	0.1204	0.3027	-	0.2312	0.0554	
LME*	Mean	1.0000	7.0027	14.9935	0.1259	25.0772	0.1136	0.890s
	St. D.	0.0040	0.0401	0.2398	0.0059	0.3520	0.0662	
BCY*	Mean	1.0012	6.9806	15.1350	0.1288	25.1728	0.2143	0.710s
	St. D.	0.0082	0.0753	0.5738	0.0113	0.4903	0.1368	

Table 3. Frequency of estimates for the partition of H: $H_1 = 0.5\delta_1$, $H_2 = 0.25\delta_7$, $H_3 = 0.125\delta_{20} + 0.125\delta_{25}$ with $p/n = 0.32$.

Dimensions	$\mathbf{k}=(1,1,2)'$	$\mathbf{k}=(1,2,1)'$	$\mathbf{k}=(2,1,1)'$
$(p,n) = (320, 1000)$	1000	0	0
$(p,n) = (160, 500)$	922	28	50
$(p,n) = (64, 200)$	595	183	222
$(p,n) = (32, 100)$	455	267	278
$(p,n) = (16, 50)$	376	260	364

similar model with $H = 0.5\delta_1 + 0.25\delta_7 + 0.125\delta_{20} + 0.125\delta_{25}$ and $c = 0.32$. The empirical and limiting distributions of sample eigenvalues are illustrated in Figure 3, where $S_F = [0.2617, 1.6951] \cup [3.2916, 10.4557] \cup [12.3253, 39.2608]$. Analogous statistics are summarized in Tables 3 and 4. The results in Table 3 show that the estimation of the partition $\mathbf{k}$ is more difficult in this case, but its accuracy is 100% at $n = 1{,}000$. Table 4 reveals that the estimators of a_3 and a_4 from the ME have a bias as large as 0.85 on average when all weight parameters are assumed known, while the LME and the BCY are unbiased in the same settings. It is again confirmed that the LME improves upon the BCY, especially when the weight parameters are partially unknown.

Finally, we took $H = 0.5\delta_1 + 0.25\delta_3 + 0.125\delta_{15} + 0.125\delta_{25}$ and $c = 0.32$ to examine the performance of the GLME. The empirical and limiting distributions of sample eigenvalues are illustrated in Figure 4, where $S_F = [0.2552, 1.6086] \cup [1.6609, 4.7592] \cup [9.1912, 37.6300]$. The first two clusters of sample eigenvalues are too close to be identified, and we have to merge them into one to get the GLME of H (thus no weight parameters are known at all). For comparison, we also present the LME by assuming that we know the true separation of S_F into three

Table 4. Estimates for $H = 0.5\delta_1 + 0.25\delta_7 + 0.125\delta_{20} + 0.125\delta_{25}$ with $p = 320$ and $n = 1{,}000$.

		a_1	a_2	a_3	w_3	a_4	d	Time
ME	Mean	1.0001	6.9996	**19.1483**	-	**25.8521**	**0.2224**	0.533s
	St. D.	0.0041	0.0395	0.1836	-	0.2068	0.0404	
LME	Mean	1.0000	7.0006	19.9157	-	25.0811	0.0620	0.575s
	St. D.	0.0040	0.0391	0.2404	-	0.2631	0.0341	
BCY	Mean	0.9965	7.0090	19.9028	-	25.0874	0.0875	0.142s
	St. D.	0.0126	0.0692	0.3456	-	0.3155	0.0516	
LME*	Mean	1.0000	7.0003	19.8739	0.1282	25.2896	0.2588	0.896s
	St. D.	0.0039	0.0390	0.7883	0.0342	0.8857	0.1464	
BCY*	Mean	0.9993	6.9983	19.8587	0.1331	25.4569	0.3286	0.865s
	St. D.	0.0054	0.0446	1.2884	0.0437	1.0888	0.1685	

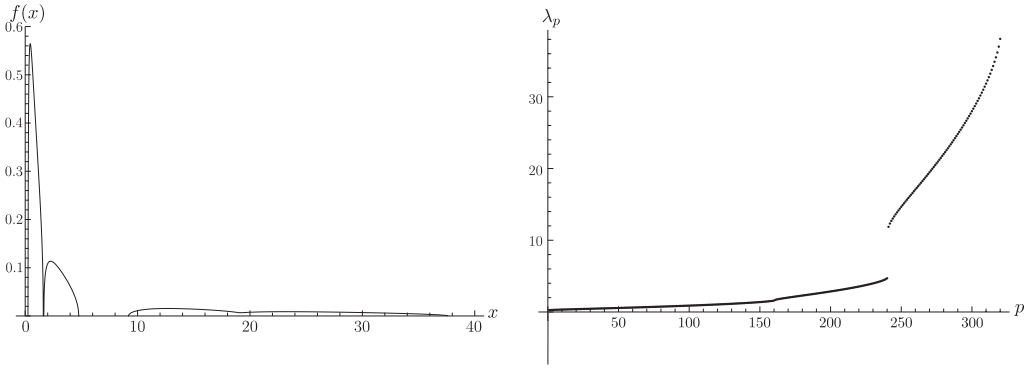

Figure 4. The density curve of F (left) and the average of the ith ($i = 1, \ldots, 320$) sample eigenvalues (right) from 1,000 replications for $H = 0.5\delta_1 + 0.25\delta_3 + 0.125\delta_{15} + 0.125\delta_{25}$ and $c = 0.32$.

intervals. Statistics in Table 5 show a perfect estimation of $\mathbf{k}$ with sample sizes $n = 500$, 1,000. Results in Table 6 demonstrate that the GMLE has a very good performance with only a slight reduction in estimation efficiency compared with the (impractical) LME. Note that the BCY becomes unstable for this model as, for example, the empirical moment equations defining the estimator often have no real solutions. A major reason is that the required estimates of the 6th and 7th moments of H have poor accuracy.

5. Conclusions and Remarks

This paper investigates the problem of estimating the population spectral distribution from the sample eigenvalues in a large-dimensional framework. A local moment estimation procedure is proposed by considering the division of a

Table 5. Frequency of estimates for the partition of H: $H_1 = 0.5\delta_1 + 0.25\delta_3$, $H_2 = 0.125\delta_{15} + 0.125\delta_{25}$ with $p/n = 0.32$.

Dimensions	$\mathbf{k}=(2,2)'$	$\mathbf{k}=(1,3)'$	$\mathbf{k}=(3,1)'$
$(p,n) = (320, 1000)$	1000	0	0
$(p,n) = (160, 500)$	1000	0	0
$(p,n) = (64, 200)$	984	0	16
$(p,n) = (32, 100)$	911	0	89
$(p,n) = (16, 50)$	865	0	135

Table 6. Estimates for $H = 0.5\delta_1 + 0.25\delta_3 + 0.125\delta_{15} + 0.125\delta_{25}$ with $p = 320$ and $n = 1{,}000$.

		a_1	w_1	a_2	w_2	a_3
GLME	Mean	1.0015	0.5015	3.0089	0.2485	15.0133
	St. D.	0.0080	0.0043	0.0270	0.0043	0.2243
LME	Mean	1.0003	-	2.9996	-	15.0061
	St. D.	0.0042	-	0.0165	-	0.2267
		w_3	a_4	w_4	d	Time
GLME	Mean	0.1265	25.1109	0.1235	0.1188	0.817s
	St. D.	0.0058	0.3361	0.0058	0.0639	
LME	Mean	0.1262	25.1058	0.1238	0.1074	0.820s
	St. D.	0.0058	0.3428	0.0058	0.0641	

discrete PSD H according to the separation of the LSD F. The new estimates are easy to compute and are proved to be consistent.

Our estimation procedure can be seen as an extension of the method in Mestre (2008). The extension mainly focus on two aspects: we drop the splitting condition; we do not need to know the weight parameters beforehand. These improvements enable our approach to be applied successfully to more complex PSDs.

The proposed method is more efficient than that in Bai, Chen, and Yao (2010). This could be attributed to two facts: our estimator uses lower moments of the PSD H; our estimator is more efficient from removing possible mixture effects brought by sample eigenvalues from different H_i's.

Acknowledgement

The authors would like to thank M. Yao Jianfeng (Telecom ParisTech, Paris) for insightful discussions on the relationship (3.2) between moments of a PSD and the contour integrals. They are grateful to the editors and referees whose careful

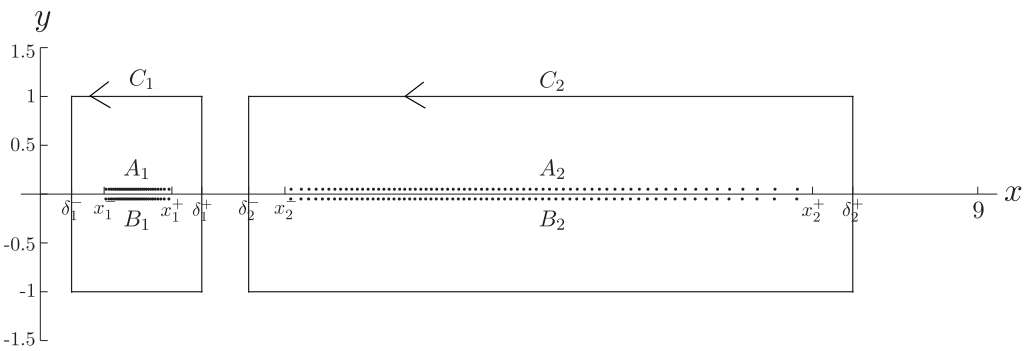

Figure A.1. Representation of A_i, B_i, and C_i $(i = 1, 2)$ where $H = 0.3\delta_1 + 0.4\delta_4 + 0.3\delta_5$, $(c, p, n) = (0.1, 100, 1000)$, and $S_F = [x_1^-, x_1^+] \cup [x_2^-, x_2^+] = [0.6127, 1.2632] \cup [2.3484, 7.4137]$.

and detailed comments have led to many improvements of the manuscript.

Appendix: Calculation of the contour integrals in Equation (3.2)

The possible poles in (3.2) are sample eigenvalues and zeros of $\underline{s}_n(u)$ on the real line. We need to determine which poles fall within the ith integration region C_i.

Let $v = \min\{p, n\}$ and $\lambda_1 < \cdots < \lambda_v$ be the nonzero sample eigenvalues. According to Bai and Silverstein (1999), these sample eigenvalues should form m separate clusters for all large p and n. Thus, with probability one, the ith cluster of sample eigenvalues, denoted by A_i, falls within C_i for all large p and n.

On the other hand, $\underline{s}_n(u) = 0$ is equivalent to $\sum_{i=1}^{v} \lambda_i/(\lambda_i - u) = n$ (except for $p/n = 1$, where the second equation would have an additional zero solution). For $\mu_1 < \cdots < \mu_v$ as the zeros of $\underline{s}_n(u)$ (define $\mu_1 = 0$ if $p/n = 1$), we have $\mu_1 < \lambda_1 < \mu_2 \cdots < \mu_v < \lambda_v$. Let $B_i = \{\mu_i : \mu_i \neq 0,\ \lambda_i \in A_i\}$ $(i = 1, \ldots, m)$. From the proof of Lemma 1 in Mestre (2008), with probability one, B_i falls within C_i for all large p and n. A representation of A_i's, B_i's, and C_i's is shown in Figure A.1 for a simple case. In order to differentiate between the A_i's and B_i's, the elements of A_i's are plotted on the line $y = 0.05$ and those of B_i's are plotted on the line $y = -0.05$.

The contour integral in (3.2) is formulated (approximately) as

$$\frac{1}{2\pi \mathrm{i}} \oint_{C_i} \frac{z\underline{s}_n'(z)}{\underline{s}_n^l(z)} dz = \sum_{\lambda \in A_i} \mathrm{Res}(f_{ln}, \lambda) + \sum_{\mu \in B_i} \mathrm{Res}(f_{ln}, \mu), \tag{A.1}$$

where $f_{ln}(z) = z\underline{s}_n'(z)/\underline{s}_n^l(z)$. The residues in (A.1) can be obtained with some calculation. Residues from A_i are simple:

$$\mathrm{Res}(f_{ln}, \lambda) = -\lambda I(l = 1).$$

Residues from B_i are listed for $l = 1, \ldots, 5$:

$$\operatorname{Res}(f_{ln}, \mu) = \begin{cases} \mu & (l = 1), \\ \dfrac{1}{\underline{s}_n'(\mu)} & (l = 2), \\ -\dfrac{\underline{s}_n''(\mu)}{2(\underline{s}_n'(\mu))^3} & (l = 3), \\ \dfrac{3(\underline{s}_n''(\mu))^2 - \underline{s}_n'(\mu)\underline{s}_n'''(\mu)}{6(\underline{s}_n'(\mu))^5} & (l = 4), \\ -\dfrac{15(\underline{s}_n''(\mu))^3 - 10\underline{s}_n'(\mu)\underline{s}_n''(\mu)\underline{s}_n'''(\mu) + (\underline{s}_n'(\mu))^2\underline{s}_n^{(4)}(\mu)}{24(\underline{s}_n'(\mu))^7} & (l = 5). \end{cases}$$

For larger order l, we can get an analytic expression for $\operatorname{Res}(f_{ln}, \mu)$ from the Mathematica code (here, the order of the moment is 3):

```
k = 3; * input the order of moment *
f = (z-mu)^k*z*D[sn[z],z]/(sn[z])^k;
D[f,{z,k-1}];
D[%*sn[z]^(2k-1),{z,2k-1}]/.z->mu;
D[sn[z],z]^(2k-1)(k-1)!(2k-1)!/.z->mu;
Simplify[%%/%,sn[mu]==0]
```

References

Bai, Z. D., Chen, J. Q. and Yao, J. F. (2010). On estimation of the population spectral distribution from a high-dimensional sample covariance matrix. *Aust. N. Z. J. Stat.* **52**, 423-437.

Bai, Z. D. and Silverstein, J. (1998). No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices. *Ann. Probab.* **26**, 316-345.

Bai, Z. D. and Silverstein, J. (1999). Exact separation of eigenvalues of large dimensional sample covariance matrices. *Ann. Probab.* **27**, 1536-1555.

Chen, J. Q., Delyon, B. and Yao, J. F. (2011). On a model selection problem from high-dimensional sample covariance matrices. *J. Multivariate Anal.* **102**, 1388-1398.

El Karoui, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Statist.* **36**, 2757-2790.

Hachem, W., Loubaton, P., Mestre, X., Najim, J. and Vallet, P. (2012). Large information plus noise random matrix models and consistent subspace estimation in large sensor networks. *Random Matrices: Theory Appl.* **1**, 1150006.

Johnstone I. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29**, 295-327.

Li, W. M., Chen, J. Q., Qin, Y. L., Yao, J. F. and Bai, Z. D. (2013). Estimation of the population spectral distribution from a large dimensional sample covariance matrix. *J. Statist. Plann. Inference* **143**, 1887-1897.

Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.* (*N.S.*) **72**, 507-536.

Mestre, X. (2008). Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Trans. Inform. Theory* **54**, 5113-5129.

Rao, N. R., Mingo, J. A., Speicher, R. and Edelman, A. (2008). Statistical eigen-inference from large Wishart matrices. *Ann. Statist.* **36**, 2850-2885.

Silverstein, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55**, 331-339.

Silverstein, J. W. and Bai, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.* **54**, 175-192.

Silverstein, J. W. and Choi, S. I. (1995). Analysis of the limiting spectral distribution of large-dimensional random matrices. *J. Multivariate Anal.* **54**, 295-309.

School of Mathematics and Systems Science, Beihang University, Beijing, 100191, China.

School of Science, Beijing University of Posts and Telecommunications, Beijing, 100876, China.

E-mail: liwm601@gmail.com

Department of Statistics and Actuarial Sciences, The University of Hong Kong, Hongkong, China.

E-mail: jeffyao@hku.hk

(Received April 2012; accepted April 2013)