On the Empirical Distribution of Eigenvalues of Large Dimensional Information-Plus-Noise Type Matrices

R. Brent Dozier[1] and Jack W. Silverstein[2]

May 2004

Abstract

Let X_n be $n \times N$ containing i.i.d. complex entries and unit variance (sum of variances of real and imaginary parts equals 1), $\sigma > 0$ constant, and R_n an $n \times N$ random matrix independent of X_n. Assume, almost surely, as $n \to \infty$, the empirical distribution function (e.d.f.) of the eigenvalues of $\frac{1}{N} R_n R_n^*$ converges in distribution to a nonrandom probability distribution function (p.d.f.), and the ratio $\frac{n}{N}$ tends to a positive number. Then it is shown that, almost surely, the e.d.f. of the eigenvalues of $\frac{1}{N}(R_n + \sigma X_n)(R_n + \sigma X_n)^*$ converges in distribution. The limit is nonrandom and is characterized in terms of its Stieltjes transform, which satisfies a certain equation.

1. Introduction

For any square matrix A with only real eigenvalues, let F^A denote the empirical distribution function (e.d.f.) of the eigenvalues of A (that is, $F^A(x)$ is the proportion of eigenvalues of $A \le x$). The focus of this paper is on the limiting e.d.f. of the eigenvalues of matrices of the form $C_n = \frac{1}{N}(R_n + \sigma X_n)(R_n + \sigma X_n)^*$ where X_n is $n \times N$ containing i.i.d. complex entries and unit variance (sum of variances of real and imaginary parts equals 1), $\sigma > 0$ is constant, R_n is an $n \times N$ random matrix independent of X_n, and n and N both converge to infinity but their ratio $\frac{n}{N}$ converges to a positive quantity c, and $F^{\frac{1}{N} R_n R_n^*}$ converges, almost surely, in distribution to a nonrandom probability distribution function (p.d.f.) H. The aim of this paper is to show that, almost surely, F^{C_n} converges in distribution to a nonrandom p.d.f. F.

The matrix C_n can be viewed as the sample correlation matrix of the columns of $R_n + \sigma X_n$, which models situations where relevant information is contained in the $R_{\cdot i}$'s and can be extracted from $\frac{1}{N} R_n R_n^*$. However, the creation of this matrix is hindered due to the fact that each $R_{\cdot i}$ is corrupted by additive noise $\sigma X_{\cdot i}$. If the number of samples N is sufficiently large and if the noise is centered ($EX_{11} = 0$), then C_n would be a reasonable estimate of $\frac{1}{N} R_n R_n^* + \sigma^2 I$ (I denoting the $n \times n$ identity matrix), which could also yield significant (if not all) information. Under the assumption $\frac{n}{N} \to c > 0$, C_n models situations where, due to the size of n, the number of samples needed to adequately approximate $\frac{1}{N} R_n R_n^* + \sigma^2 I$ is unattainable, but is on the same order of magnitude as n.

[1]North Carolina State University, Raleigh, NC 27695
Email: rbdozier@math.ncsu.edu
[2]North Carolina State University, Raleigh, NC 27695
Email: jack@math.ncsu.edu Homepage: www.math.ncsu.edu/~jack
Key words and phrases: random matrix, empirical distribution function of eigenvalues, Stieltjes transform.
AMS 1991 subject classifications: primary 60F15; secondary 62H99

One example of this is in the area of array signal processing with regard to the so-called detection problem. The model is described by a matrix $Y_n = R_n + \sigma X_n$ where the N columns of Y_n represent N "snapshots" (samples) of the data received at n sensors from signals transmitted by an unknown number $q < n$ of sources with unknown locations. The matrix R_n contains the signal information as transmitted, and the matrix σX_n is additive noise (variance σ^2 unknown) that contaminates the signal during transmission and processing. The contents of R_n include information on the unknown direction of arrival of the signals, values detailing sensor orientation, and the signal values at the sources. The entries of X_n are i.i.d. and standardized. The goal is to identify the number of sources and their direction of arrival (DOA), which could be achieved if the population matrix $\frac{1}{N}R_nR_n^* + \sigma^2 I$ were known (see Schmidt [2]). This matrix is estimated by the sample covariance matrix $C_n = \frac{1}{N}Y_nY_n^*$. However, as stated above, for large n, it may not be possible to collect a sufficient number of samples for estimation. Limiting results on the eigenvalues of these matrices will aid in the detection problem: determining the number of sources. Details of the importance of such limiting results are given in Silverstein and Combettes [5], where a less general case is presented, namely, independence across samples is assumed. In the present work, we only require that the e.d.f. of the eigenvalues of $\frac{1}{N}R_nR_n^*$ converges in distribution to a nonrandom p.d.f. H, thus relieving the matrix R_n of such independence assumptions and allowing for a more general approach to signal detection.

The methods used in this paper are similar to those used in Silverstein and Bai [4] with the main tool being the Stieltjes transform. For any p.d.f. G, the Stieltjes transform of G is defined as the analytic function

$$m_G(z) = \int \frac{1}{\lambda - z} dG(\lambda), \quad z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C} : \mathcal{I}m\, z > 0\},$$

and G can be retrieved by the inversion formula

$$G\{[a,b]\} = \frac{1}{\pi} \lim_{\eta \to 0^+} \int_a^b \mathcal{I}m\, m_G(\xi + i\eta) d\xi,$$

where a, b are continuity points of G. Due to the inversion formula, convergence of a tight sequence of p.d.f.'s is guaranteed by showing convergence of the corresponding Stieltjes transforms on a countable subset of $\mathbb{C}^+$ possessing at least one accumulation point in $\mathbb{C}^+$.

A property of Stieltjes transforms that will be needed later is that if G is any p.d.f. with nonnegative support, then

$$\mathcal{I}m\, z m_G(z) = \int \frac{\lambda z_2}{|\lambda - z|^2} dG(\lambda) \geq 0, \qquad (S.1)$$

for any $z = z_1 + iz_2 \in \mathbb{C}^+$.

For $p \times p$ matrix A with real eigenvalues $\lambda_1, \lambda_2, ..., \lambda_p$ the Stieltjes transform of F^A,

$$m_{F^A}(z) = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{\lambda_i - z} = \frac{1}{p} tr(A - zI)^{-1},$$

involves the resolvent of A and is well-suited for our analysis (tr denoting trace).

The following theorem will be proven.

THEOREM 1.1. Assume on a common probability space:

(a) For n=1,2,..., $X_n = (X_{ij}^n)$, $n \times N$, $X_{ij}^n \in \mathbb{C}^+$, i.d. for all n, i, j, independent across i, j for each n, $E|X_{11}^1 - EX_{11}^1|^2 = 1$

(b) R_n is $n \times N$ independent of X_n with $F^{\frac{1}{N}R_nR_n^*} \xrightarrow{\mathcal{D}} H$, a.s., ($\mathcal{D}$ denoting weak convergence) where H is a nonrandom p.d.f.

(c) $N = N(n)$ and $c_n = \frac{n}{N} \to c > 0$ as $n \to \infty$

(d) $C_n = \frac{1}{N}(R_n + \sigma X_n)(R_n + \sigma X_n)^*$ where $\sigma > 0$.

Then $F^{C_n} \xrightarrow{\mathcal{D}} F$, a.s., where F is a nonrandom p.d.f. whose Stieltjes transform $m = m(z)$ satisfies

$$m = \int \frac{dH(t)}{\frac{t}{1+\sigma^2 cm} - (1+\sigma^2 cm)z + \sigma^2(1-c)} \tag{1.1}$$

for any $z \in \mathbb{C}^+$.

From (S.1) we see that the imaginary part of the denominator of the integrand in (1.1) is less than or equal to $-z_2$, so that the integral is well-defined.

Let $\mathbf{C}_n = \frac{1}{N}(R_n + \sigma X_n)^*(R_n + \sigma X_n)$. The spectra of C_n and $\mathbf{C}_n$ differ by $|n - N|$ zero eigenvalues and is expressed in

$$F^{\mathbf{C}_n} = \left(1 - \frac{n}{N}\right) 1_{[0,\infty)} + \frac{n}{N} F^{C_n} \tag{1.2}$$

(1_B denoting the indicator function over the set B). Because of this, information on the limit of $F^{\mathbf{C}_n}$ can be inferred from knowledge of F.

Notice also that the eigenvalues of C_n are directly related to those of the $N \times n$ matrix $\frac{1}{n}(R_n^* + \sigma X_n^*)(R_n^* + \sigma X_n^*)^* = \frac{N}{n}\mathbf{C}_n$. With this fact it is straightforward to show that if m satisfies (1.1) when $c \leq 1$, then m will also satisfy (1.1) when $c > 1$. We therefore assume, without loss of generality, that $0 < c \leq 1$.

Let $m_n(z) = m_{F^{C_n}}(z)$. Defining $\mathbf{m}_n(z) = m_{F^{\mathbf{C}_n}}(z)$ we get from (1.2)

$$\mathbf{m}_n = -\frac{1 - c_n}{z} + c_n m_n, \tag{1.3}$$

which will be used later for notational convenience.

It is noted here that the qualitative behavior of F is currently being investigated by the authors. Preliminary analysis indicates that much of this information can be retrieved from (1.1).

This paper is composed of four sections and an Appendix. The first section following the introduction mirrors Silverstein and Bai [4] in that justification is presented for restricting the assumptions on the matrices R_n, and X_n. Section three contains the bulk of the proof of Theorem 1.1, and section four is devoted to showing that solutions, m, to equation (1.1) are unique if $\mathcal{I}m\, m > 0$ and $\mathcal{I}m\, mz \geq 0$ (specifically, if m is the Stieltjes transform of a p.d.f. with nonnegative support). The Appendix contains the proof of a lemma from section two.

2. Truncation and Centralization

The first step in proving Theorem 1.1 is similar to that of Silverstein and Bai [4], in that, we truncate and centralize twice with regard to X_n, and, as in Silverstein [3], we truncate R_n. The reason for these truncations and centralizations is to justify our later replacing the matrices X_n and R_n with ones more suitable for analysis. We compare the e.d.f.'s of these matrices by the following metric presented in Silverstein and Bai [4]. Let $\{f_i\}$ be an enumeration of all continuous functions that take a constant $\frac{1}{m}$ value (m a positive integer) on $[a, b]$, where a, b are rational, 0 on $(-\infty, a - \frac{1}{m}] \bigcup [b + \frac{1}{m}, \infty)$, and linear of each of $[a - \frac{1}{m}, a]$, $[b, b + \frac{1}{m}]$. For probability measures F, G on $\mathbb{R}$ the metric

$$D(F, G) \equiv \sum_{i=1}^{\infty} \left| \int f_i dF - \int f_i dG \right| 2^{-i}$$

induces the topology of weak convergence, and, as noted in Silverstein and Bai [4], for sequences $\{F_n\}, \{G_n\}$ of probability measures on $\mathbb{R}$, we have

$$\lim_{n\to\infty} \|F_n - G_n\| = 0 \Rightarrow \lim_{n\to\infty} D(F_n, G_n) = 0 \tag{2.1}$$

where $\|\cdot\|$ denotes the sup-norm on bounded functions from $\mathbb{R}$ to $\mathbb{R}$.

Note that for $x, y \in \mathbb{R}$, $|f_i(x) - f_i(y)| \leq |x - y|$. Then, restating from Silverstein and Bai [4], we have for e.d.f.'s F, G on the (respective) sets $\{x_1, ..., x_n\}, \{y_1, ..., y_n\}$,

$$D^2(F, G) \leq \left(\frac{1}{n} \sum_{j=1}^{n} |x_j - y_j| \right)^2 \leq \frac{1}{n} \sum_{j=1}^{n} (x_j - y_j)^2. \tag{2.2}$$

Before continuing, some needed results are presented.

For $q \in \mathbb{C}^n$ and $n \times N$ matrix A, $\|q\|$ will denote the Euclidean norm, and $\|A\|$ the induced spectral norm on matrices, that is, the largest singular value of A. We also use the notation F^A_{sing} to denote the e.d.f. of the square root of the eigenvalues of AA^*, which are the n largest singular values of A. The constants, denoted by K, appearing henceforth in some of the expressions are nonrandom and may take on different values from one appearance to the next.

Lemma 2.1 [Corollary 7.3.8 of Horn and Johnson [1]]. For $r \times s$ matrices A and B with respective singular values $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_q$, $\tau_1 \geq \tau_2 \geq ... \geq \tau_q$, where $q = min\{r, s\}$, we have

$$\left(\sum_{i=1}^{q} (\sigma_i - \tau_i)^2 \right)^{\frac{1}{2}} \leq \|A - B\|_2,$$

where $\|\cdot\|_2$ is the Frobenius matrix norm.

Lemma 2.2 [Lemma 2.5 of Silverstein and Bai [4]]. For $n \times N$ matrices $Q, \overline{Q}$,

$$\|F^{QQ^*} - F^{\overline{Q}\,\overline{Q}^*}\| \leq \frac{2}{n} rank(Q - \overline{Q}).$$

The following are well-known properties of matrices.

Matrix Properties.

(MP1) For $n \times n$ A, B,

$$|trAB| \leq (trAA^* \, trBB^*)^{\frac{1}{2}} \leq n\|A\|\|B\|.$$

(MP2) For rectangular A, $rank(A) \leq$ the number of nonzero entries of A.

Proof of Theorem 1.1. Following Silverstein and Bai [4] we use the convention of occasionally suppressing the variables' dependence on n. All convergence statements are as $n \to \infty$. Let $\hat{X}_{ij} = X_{ij} 1_{(|X_{ij}| < \sqrt{n})}$ and $\hat{C}_n = \left(\frac{1}{\sqrt{N}} R + \sigma \hat{X}\right)\left(\frac{1}{\sqrt{N}} R + \sigma \hat{X}\right)^*$, where $\hat{X} = \left(\frac{1}{\sqrt{N}} \hat{X}_{ij}\right)$. It is shown in the Appendix that

$$\|F^{C_n} - F^{\hat{C}_n}\| \xrightarrow{a.s.} 0. \tag{2.3}$$

Let $\tilde{C}_n = \left(\frac{1}{\sqrt{N}} R + \sigma \tilde{X}\right)\left(\frac{1}{\sqrt{N}} R + \sigma \tilde{X}\right)^*$, where $\tilde{X} = \left(\frac{1}{\sqrt{N}} \tilde{X}_{ij}\right) = \left(\frac{1}{\sqrt{N}} \hat{X}_{ij} - E\frac{1}{\sqrt{N}} \hat{X}_{ij}\right)$. Since $rank(E\hat{X}) \leq 1$, we have from Lemma 2.2

$$\|F^{\hat{C}_n} - F^{\tilde{C}_n}\| \to 0. \tag{2.4}$$

Write $\frac{1}{\sqrt{N}} R$ in its singular value decomposition $\frac{1}{\sqrt{N}} R = U \Lambda V$. Let $R_\alpha = U \Lambda_\alpha V$ where Λ_α is the matrix Λ with each singular value s replaced by $s 1_{(s \leq \alpha)}$, for $\alpha > 0$.

Let Q be any $n \times N$ matrix. If α^2 is a continuity point of H, we have by Lemma 2.2 and assumptions (b), (c)

$$\|F^{(\frac{1}{\sqrt{N}} R + Q)(\frac{1}{\sqrt{N}} R + Q)^*} - F^{(R_\alpha + Q)(R_\alpha + Q)^*}\| \leq \frac{2}{n} rank\left(\frac{1}{\sqrt{N}} R - R_\alpha\right)$$

$$= \frac{2}{n} \sum_{i=1}^{n} 1_{(s_i > \alpha)} = \frac{2}{n} \sum_{i=1}^{n} 1_{(\lambda_i > \alpha^2)} \xrightarrow{a.s.} 2H\{(\alpha^2, \infty)\},$$

where the s_i's are the n largest singular values of $\frac{1}{\sqrt{N}} R$ and the λ_i's are the eigenvalues of $\frac{1}{N} RR^*$, i.e., $\lambda_i = s_i^2$. Let $\alpha \equiv \alpha_n = \ln(n)$. It follows that as $n \to \infty$

$$\|F^{(\frac{1}{\sqrt{N}} R + Q)(\frac{1}{\sqrt{N}} R + Q)^*} - F^{(R_\alpha + Q)(R_\alpha + Q)^*}\| \xrightarrow{a.s.} 0. \tag{2.5}$$

Let $\overline{X}_{ij} = \tilde{X}_{ij} 1_{(|X_{ij}| < \ln(n))} - E\tilde{X}_{ij} 1_{(|X_{ij}| < \ln(n))}$, $\overline{X} = \left(\frac{1}{\sqrt{N}} \overline{X}_{ij}\right)$, $\overline{\overline{X}}_{ij} = \tilde{X}_{ij} - \overline{X}_{ij}$, and $\overline{\overline{X}} = \left(\frac{1}{\sqrt{N}} \overline{\overline{X}}_{ij}\right)$. Let $\tilde{s}_1 \geq \tilde{s}_2 \geq ... \geq \tilde{s}_n$ and $\overline{s}_1 \geq \overline{s}_2 \geq ... \geq \overline{s}_n$ be the n largest singular values of $R_\alpha + \sigma \tilde{X}$ and $R_\alpha + \sigma \overline{X}$, respectively. Then using (2.2),(MP1), and Lemma 2.1 we get

$$D^2\left(F_{sing}^{R_\alpha + \sigma \tilde{X}}, F_{sing}^{R_\alpha + \sigma \overline{X}}\right) \leq \frac{1}{n} \sum_{j=1}^{n} (\tilde{s}_j - \overline{s}_j)^2 \leq \frac{1}{n} \|\overline{\overline{X}}\|_2^2$$

$$= \frac{1}{n} tr \overline{\overline{X}}^* \overline{\overline{X}} \leq \left(\frac{1}{n} tr \left(\overline{\overline{X}}^* \overline{\overline{X}}\right)^2\right)^{\frac{1}{2}} \xrightarrow{a.s.} 0$$

by (3.6) of Silverstein and Bai [4]. It follows that

$$D\left(F^{(R_\alpha + \sigma \tilde{X})(R_\alpha + \sigma \tilde{X})^*}, F^{(R_\alpha + \sigma \overline{X})(R_\alpha + \sigma \overline{X})^*}\right) \xrightarrow{a.s.} 0. \tag{2.6}$$

Therefore, by (2.1), (2.3), (2.4), (2.5), and (2.6), in order to show that $F^{C_n} \xrightarrow{\mathcal{D}} F$, it is sufficient to show that for any $z \in \mathbb{C}^+$,

$$m_{F^{(R_\alpha+\sigma\overline{X})(R_\alpha+\sigma\overline{X})^*}}(z) \xrightarrow{a.s.} m_F(z).$$

We may therefore add to the conditions of Theorem 1.1 the following:

(1) $|X_{11}| \leq \ln(n)$,
(2) $EX_{11} = 0$, $E|X_{11}|^2 = 1$,
(3) $\|\frac{1}{N}RR^*\| \leq \ln(n)$.

3. Completing the Proof

Fix $z = z_1 + iz_2 \in \mathbb{C}^+$. The next four results are used to complete the proof of Theorem 1.1.

Lemma 3.1. For $n \times n$ A and $n \times 1$ vectors q, v where A and $A + vv^*$ are invertible, we have

$$q^*\left(A + vv^*\right)^{-1} = q^*A^{-1} - \frac{q^*A^{-1}v}{1 + v^*A^{-1}v}v^*A^{-1}.$$

Notice if $q = v$ then

$$v^*\left(A + vv^*\right)^{-1} = \frac{1}{1 + v^*A^{-1}v}v^*A^{-1}.$$

Proof. Let $q^*\left(A + vv^*\right)^{-1} \equiv r^*$ so that $q^* = r^*A + r^*vv^*$. Multiplying by A^{-1} on the right we get

$$q^*A^{-1} = r^* + r^*vv^*A^{-1}, \tag{3.1.1}$$

and then multiplying by v on the right we get

$$q^*A^{-1}v = r^*v + r^*vv^*A^{-1}v = r^*v\left(1 + v^*A^{-1}v\right).$$

Since q is arbitrary we must have $1 + v^*A^{-1}v \neq 0$. Then

$$r^*v = \frac{q^*A^{-1}v}{1 + v^*A^{-1}v},$$

and hence by (3.1.1) we have

$$r^* = q^*A^{-1} - \frac{q^*A^{-1}v}{1 + v^*A^{-1}v}v^*A^{-1},$$

and the proof is complete.

Lemma 3.2 [Lemma 3.1 of Silverstein and Bai [4]]. Let $C = (c_{ij})$, $c_{ij} \in \mathbb{C}$, be an $n \times n$ matrix with $\|C\| \leq 1$, and $Y = (X_1, ..., X_n)^T$, $X_i \in \mathbb{C}$, where the X_i's are i.i.d. satisfying conditions (1) and (2). Then

$$E|Y^*CY - trC|^6 \leq Kn^3(\ln(n))^{12},$$

where the constant K does not depend on n, C, nor on the distribution of X_1.

LEMMA 3.3 [Lemma 2.6 of Silverstein and Bai [4]]. Let $z = z_1 + iz_2 \in \mathbb{C}^+$ with A and B $n \times n$, B Hermitian and $r \in \mathbb{C}^n$. Then

$$\left| tr \left((B - zI)^{-1} - (B + rr^* - zI)^{-1} \right) A \right|$$
$$= \left| \frac{r^*(B - zI)^{-1}A(B - zI)^{-1}r}{1 + r^*(B - zI)^{-1}r} \right| \leq \frac{\|A\|}{z_2}.$$

LEMMA 3.4 [Lemma 2.3 of Silverstein and Bai [4]]. Let x, y be nonnegative numbers. For rectangular matrices A, B of the same size,

$$F_{sing}^{A+B}\{(x + y, \infty)\} \leq F_{sing}^{A}\{(x, \infty)\} + F_{sing}^{B}\{(y, \infty)\}.$$

Using Lemma 3.2 we get $\frac{1}{n} tr \frac{1}{N} XX^* \xrightarrow{a.s.} 1$ which yields the almost sure tightness of $\{F^{\frac{1}{N}XX^*}\}$. This together with Lemma 3.4 and assumption (b) gives us $\{F^{C_n}\}$ being almost surely tight, and therefore the quantity

$$\delta \equiv \inf_n \mathcal{I}m\left(m_{F^{C_n}}(z)\right) \geq \inf_n \int \frac{z_2 dF^{C_n}(\lambda)}{2(\lambda^2 + z_1^2) + z_2^2}$$

is positive almost surely.

For $j = 1, 2, ..., N$ let $x_j (= x_j^n)$ and $r_j (= r_j^n)$ denote the $j\,th$ column of X and R respectively and define $y_j = \frac{1}{\sqrt{N}}(r_j + \sigma x_j)$ so that $C_n = \sum_{j=1}^{N} y_j y_j^*$.

Note that $RR^* = \sum_{j=1}^{N} r_j r_j^*$, and since for each $j = 1, 2, ..., N$ the matrix $RR^* - r_j r_j^* = \sum_{i \neq j}^{N} r_i r_i^*$ is positive semidefinite, then using condition (3) from the previous section we get

$$\|r_j\|^2 = \|r_j r_j^*\| \leq \|RR^*\| \leq N \ln(n). \tag{3.1}$$

Define $D = C_n - zI$, $B = A_n - zI$, where

$$A_n \equiv \left(\frac{1}{1 + \sigma^2 c_n m_n} \right) \frac{1}{N} RR^* - \sigma^2 z \mathbf{m}_n I,$$

and for $j = 1, 2, ..., N$ let $C_{(j)} = C_n - y_j y_j^*$ and $D_j = D - y_j y_j^* (= C_{(j)} - zI)$. Write

$$D + zI = \sum_{j=1}^{N} y_j y_j^*.$$

Multiplying by D^{-1} on the right on both sides and using Lemma 3.1 we get

$$I + zD^{-1} = \sum_{j=1}^{N} \frac{1}{1 + y_j^* D_j^{-1} y_j} y_j y_j^* D_j^{-1}.$$

Taking the trace on both sides and dividing by N we have

$$c_n + z c_n m_n = \frac{1}{N} \sum_{j=1}^{N} \frac{y_j^* D_j^{-1} y_j}{1 + y_j^* D_j^{-1} y_j} = 1 - \frac{1}{N} \sum_{j=1}^{N} \frac{1}{1 + y_j^* D_j^{-1} y_j}.$$

From our definition (1.3) of $\mathbf{m}_n$, we see that

$$\mathbf{m}_n = -\frac{1}{N}\sum_{j=1}^{N}\frac{1}{z(1+y_j^*D_j^{-1}y_j)}. \tag{3.2}$$

Following the steps leading up to (2.3) of Silverstein [3] we get

$$\frac{1}{\left|z\left(1+y_j^*D_j^{-1}y_j\right)\right|} \leq \frac{1}{z_2}. \tag{3.3}$$

For $j = 1, 2, ..., N$, we make the following scalar definitions:

$$\rho_j = \frac{1}{N}r_j^*D_j^{-1}r_j, \quad \omega_j = \frac{1}{N}\sigma^2 x_j^*D_j^{-1}x_j,$$

$$\beta_j = \frac{1}{N}r_j^*D_j^{-1}\sigma x_j, \quad \gamma_j = \frac{1}{N}\sigma x_j^*D_j^{-1}r_j,$$

$$\hat{\rho}_j = \frac{1}{N}r_j^*D_j^{-1}B^{-1}r_j, \quad \hat{\omega}_j = \frac{1}{N}\sigma^2 x_j^*D_j^{-1}B^{-1}x_j,$$

$$\hat{\beta}_j = \frac{1}{N}r_j^*D_j^{-1}B^{-1}\sigma x_j, \quad \hat{\gamma}_j = \frac{1}{N}\sigma x_j^*D_j^{-1}B^{-1}r_j.$$

We begin the next stage of the proof by factoring the difference of inverses and expanding the middle factor to get

$$B^{-1} - D^{-1} = B^{-1}(D-B)D^{-1} = B^{-1}(C_n - A_n)D^{-1}$$

$$= B^{-1}\Big(\frac{\sigma^2 c_n m_n}{1+\sigma^2 c_n m_n}\,\frac{1}{N}RR^* + \frac{1}{N}\sigma XR^* + \frac{1}{N}R\sigma X^* + \frac{1}{N}\sigma^2 XX^* + \sigma^2 z\mathbf{m}_n I\Big)D^{-1}$$

$$= \sum_{j=1}^{N} B^{-1}\Big[\frac{\sigma^2 c_n m_n}{1+\sigma^2 c_n m_n}\,\frac{1}{N}r_j r_j^* + \frac{1}{N}\sigma x_j r_j^* + \frac{1}{N}r_j\sigma x_j^* + \frac{1}{N}\sigma^2 x_j x_j^* + \frac{1}{N}\sigma^2 z\mathbf{m}_n I\Big]D^{-1}$$

$$= \sum_{j=1}^{N}\Big[\frac{\sigma^2 c_n m_n}{1+\sigma^2 c_n m_n}B^{-1}\frac{1}{N}r_j r_j^* D^{-1} + B^{-1}\frac{1}{N}\sigma x_j r_j^* D^{-1} + B^{-1}\frac{1}{N}r_j\sigma x_j^* D^{-1}$$
$$+ B^{-1}\frac{1}{N}\sigma^2 x_j x_j^* D^{-1} + \frac{1}{N}\sigma^2 z\mathbf{m}_n B^{-1}D^{-1}\Big].$$

While using (3.2), we take the trace of both sides and divide by n to get

$$\frac{1}{n}tr(A_n - zI)^{-1} - m_n = \frac{1}{n}\sum_{j=1}^{N}\Big[\frac{\sigma^2 c_n m_n}{1+\sigma^2 c_n m_n}\,\frac{1}{N}r_j^*D^{-1}B^{-1}r_j + \frac{1}{N}r_j^*D^{-1}B^{-1}\sigma x_j$$
$$+ \quad \frac{1}{N}\sigma x_j^*D^{-1}B^{-1}r_j + \frac{1}{N}\sigma^2 x_j^*D^{-1}B^{-1}x_j$$
$$- \frac{1}{1+\frac{1}{N}(r_j+\sigma x_j)^*D_j^{-1}(r_j+\sigma x_j)}\,\frac{1}{N}\sigma^2 trD^{-1}B^{-1}\Big]$$

$$\equiv \frac{1}{n}\sum_{j=1}^{N}\Big[W_1^{n,j}+W_2^{n,j}+W_3^{n,j}+W_4^{n,j}-W_5^{n,j}\Big]. \tag{3.4}$$

Let $\alpha^{n,j} = 1+\frac{1}{N}(r_j+\sigma x_j)^* D_j^{-1}(r_j+\sigma x_j) = 1+\rho_j+\beta_j+\gamma_j+\omega_j$.
Since $D^{-1} = (D_j + \frac{1}{\sqrt{N}}(r_j+\sigma x_j)\frac{1}{\sqrt{N}}(r_j+\sigma x_j)^*)^{-1}$ we can use Lemma 3.1 to get

$$\begin{aligned}
W_1^{n,j} &= \frac{1}{\alpha^{n,j}}\Big(\frac{\sigma^2 c_n m_n}{1+\sigma^2 c_n m_n}\Big)\Big[\alpha^{n,j}\frac{1}{N}r_j^* D_j^{-1}B^{-1}r_j \\
&\qquad -\frac{1}{N}(r_j^* D_j^{-1}(r_j+\sigma x_j))\frac{1}{N}(r_j+\sigma x_j)^* D_j^{-1}B^{-1}r_j\Big] \\
&= \frac{1}{\alpha^{n,j}}\Big(\frac{\sigma^2 c_n m_n}{1+\sigma^2 c_n m_n}\Big)\Big[(1+\gamma_j+\omega_j)\hat{\rho}_j-(\rho_j+\beta_j)\hat{\gamma}_j\Big], \\
W_2^{n,j} &= \frac{1}{\alpha^{n,j}}\Big[\alpha^{n,j}\frac{1}{N}r_j^* D_j^{-1}B^{-1}\sigma x_j \\
&\qquad -\frac{1}{N}(r_j^* D_j^{-1}(r_j+\sigma x_j))\frac{1}{N}(r_j+\sigma x_j)^* D_j^{-1}B^{-1}\sigma x_j\Big] \\
&= \frac{1}{\alpha^{n,j}}\Big[(1+\gamma_j+\omega_j)\hat{\beta}_j-(\rho_j+\beta_j)\hat{\omega}_j\Big], \\
W_3^{n,j} &= \frac{1}{\alpha^{n,j}}\Big[\alpha^{n,j}\frac{1}{N}\sigma x_j^* D_j^{-1}B^{-1}r_j \\
&\qquad -\frac{1}{N}(\sigma x_j^* D_j^{-1}(r_j+\sigma x_j))\frac{1}{N}(r_j+\sigma x_j)^* D_j^{-1}B^{-1}r_j\Big] \\
&= \frac{1}{\alpha^{n,j}}\Big[(1+\rho_j+\beta_j)\hat{\gamma}_j-(\gamma_j+\omega_j)\hat{\rho}_j\Big], \\
W_4^{n,j} &= \frac{1}{\alpha^{n,j}}\Big[\alpha^{n,j}\frac{1}{N}\sigma x_j^* D_j^{-1}B^{-1}\sigma x_j \\
&\qquad -\frac{1}{N}(\sigma x_j^* D_j^{-1}(r_j+\sigma x_j))\frac{1}{N}(r_j+\sigma x_j)^* D_j^{-1}B^{-1}\sigma x_j\Big] \\
&= \frac{1}{\alpha^{n,j}}\Big[(1+\rho_j+\beta_j)\hat{\omega}_j-(\gamma_j+\omega_j)\hat{\beta}_j\Big], \qquad \text{and} \\
W_5^{n,j} &= \frac{1}{\alpha^{n,j}}\frac{1}{N}\sigma^2 trD^{-1}B^{-1}.
\end{aligned}$$

Therefore, after simplification, we get

$$\begin{aligned}
(3.4) \;&= \frac{1}{n}\sum_{j=1}^{N}\frac{1}{\alpha^{n,j}}\Big[\frac{1}{1+\sigma^2 c_n m_n}(\sigma^2 c_n m_n-\omega_j-\gamma_j)\hat{\rho}_j+\hat{\beta}_j \\
&\qquad +\frac{1}{1+\sigma^2 c_n m_n}(\rho_j+\beta_j+1+\sigma^2 c_n m_n)\hat{\gamma}_j+\hat{\omega}_j-\frac{1}{N}\sigma^2 trD^{-1}B^{-1}\Big] \\
&\equiv \frac{1}{n}\sum_{j=1}^{N}\frac{1}{\alpha^{n,j}}d^{n,j}.
\end{aligned}$$

For $j = 1, 2, ..., N$ we make the following definitions

$$m_{(j)} \equiv m_{F^{C_{(j)}}}(z), \qquad \mathbf{m}_{(j)} \equiv -\frac{1-c_n}{z} + c_n m_{(j)},$$

$$B_j \equiv \left(\frac{1}{1+\sigma^2 c_n m_{(j)}}\right) \frac{1}{N} RR^* - \sigma^2 z \mathbf{m}_{(j)}.$$

As noted below (2.5) of Silverstein [3], $\mathbf{m}_{(j)}$ is the Stieltjes transform of a p.d.f. (on $[0, \infty)$).

The following expressions hold for any $j = 1, 2, ..., N$ and any n.

From (3.3) we get

$$\frac{1}{|\alpha^{n,j}|} \leq \frac{|z|}{z_2}, \tag{3.5}$$

and since for any Hermitian matrix A, $\|(A - zI)^{-1}\| \leq \frac{1}{z_2}$, we have

$$\|D_j^{-1}\| \leq \frac{1}{z_2}. \tag{3.6}$$

By (S.1) we get

$$\frac{1}{|1+\sigma^2 c_n m_n|} \leq \frac{|z|}{z_2 + \sigma^2 c_n \mathcal{I}m\, z m_n} \leq \frac{|z|}{z_2} \tag{3.7}$$

and similarly

$$\frac{1}{|1+\sigma^2 c_n m_{(j)}|} \leq \frac{|z|}{z_2}. \tag{3.8}$$

From Lemma 3.3 we have

$$\max_{j \leq N} |m_n - m_{(j)}| \leq \frac{1}{n z_2}. \tag{3.9}$$

Suppose λ is an eigenvalue of $\frac{1}{N} RR^*$ and $\lambda^B = \frac{1}{1+\sigma^2 c_n m_n}\lambda - \sigma^2 z \mathbf{m}_n - z$ is the corresponding eigenvalue of B. Then (S.1) gives

$$|\lambda^B| \geq |\mathcal{I}m\, \lambda^B| = \left| \frac{\sigma^2 c_n \mathcal{I}m\, m_n}{|1+\sigma^2 c_n m_n|^2} \lambda + \sigma^2 \mathcal{I}m\, z \mathbf{m}_n + z_2 \right| \geq z_2.$$

Therefore

$$\|B^{-1}\| = \frac{1}{|\lambda^B_{min}|} \leq \frac{1}{z_2}, \tag{3.10}$$

and similarly

$$\|B_j^{-1}\| \leq \frac{1}{z_2}. \tag{3.11}$$

Using (3.7), (3.8), (3.9), (3.10), (3.11), and condition (3) we get

$$\|B_j^{-1} - B^{-1}\| = \|B_j^{-1}(B - B_j)B^{-1}\| \leq \frac{1}{z_2^2}\|B - B_j\|$$

$$= \frac{\sigma^2 c_n |m_{(j)} - m_n|}{z_2^2} \left\| \frac{1}{(1+\sigma^2 c_n m_n)(1+\sigma^2 c_n m_{(j)})} \frac{1}{N} RR^* + zI \right\|$$

$$\leq \frac{\sigma^2 c_n}{n z_2^3} \left(\frac{1}{|1+\sigma^2 c_n m_n||1+\sigma^2 c_n m_{(j)}|} \left\| \frac{1}{N} RR^* \right\| + |z| \right)$$

$$\leq \frac{\sigma^2 c_n}{n z_2^3}\left(\frac{|z|^2}{z_2^2}\ln(n) + |z|\right) \leq K\frac{\ln(n)}{n}. \tag{3.12}$$

A simple application of Lemma 3.2 gives

$$E\|x_j\|^{12} \leq K n^6 (\ln(n))^{12}. \tag{3.13}$$

The combination of (3.1), (3.6), (3.10), and the Cauchy-Schwarz inequality yields

$$|\hat{\rho}_j| \leq K\ln(n) \quad \text{and} \quad |\rho_j| \leq K\ln(n).$$

The Cauchy-Schwarz inequality along with Lemma 3.2, Lemma 3.3, (3.6), (3.10), (3.11), (3.12), (3.13), and (MP1) gives

$$\begin{aligned}
E|\hat{\omega}_j - \frac{1}{N}\sigma^2 trD^{-1}B^{-1}|^6 &= \frac{\sigma^{12}}{N^6}E|x_j^* D_j^{-1}B^{-1}x_j - trD^{-1}B^{-1}|^6 \\
&\leq \frac{K}{N^6}E|x_j^* D_j^{-1}(B^{-1} - B_j^{-1})x_j|^6 + \frac{K}{N^6}E|x_j^* D_j^{-1}B_j^{-1}x_j - trD_j^{-1}B_j^{-1}|^6 \\
&\quad + \frac{K}{N^6}E|trD_j^{-1}(B_j^{-1} - B^{-1})|^6 + \frac{K}{N^6}E|tr(D_j^{-1} - D^{-1})B^{-1}|^6 \\
&\leq K\frac{(\ln(n))^6}{N^{12}}E\|x_j\|^{12} + K\frac{(\ln(n))^{12}}{N^3} + K\frac{(\ln(n))^6}{N^6} + \frac{K}{N^6} \\
&\leq K\frac{(\ln(n))^{18}}{N^3}.
\end{aligned}$$

Using (3.6), Lemma 3.2, and Lemma 3.3 we get

$$\begin{aligned}
E|\omega_j - \sigma^2 c_n m_n|^6 &= \frac{\sigma^{12}}{N^6}E|x_j D_j^{-1}x_j - trD^{-1}|^6 \\
&\leq \frac{K}{N^6}\left(E|x_j^* D_j^{-1}x_j - trD_j^{-1}|^6 + E|tr(D_j^{-1} - D^{-1})|^6\right) \\
&\leq \frac{K}{N^3}(\ln(n))^{12} + \frac{K}{N^6} \\
&\leq K\frac{(\ln(n))^{12}}{N^3}.
\end{aligned}$$

From (3.1), (3.6), (3.11), (3.12), (3.13) Lemma 3.2, and the Cauchy-Schwarz inequality we have

$$\begin{aligned}
E|\hat{\gamma}_j|^{12} &\leq \frac{K}{N^{12}}E|x_j^* D_j^{-1}(B^{-1} - B_j^{-1})r_j|^{12} + \frac{K}{N^{12}}E\left||x_j^* D_j^{-1}B_j^{-1}r_j|^2\right|^6 \\
&\leq \frac{K}{N^{12}}E\|x_j\|^{12}\|r_j\|^{12}\|D_j^{-1}\|^{12}\|B^{-1} - B_j^{-1}\|^{12} \\
&\qquad + \frac{K}{N^{12}}E|x_j^* D_j^{-1}B_j^{-1}r_j r_j^* B_j^{-1*}D_j^{-1*}x_j|^6 \\
&\leq K\frac{(\ln(n))^{18}}{N^{18}}E\|x_j\|^{12} + \frac{K}{N^{12}}E|x_j^* D_j^{-1}B_j^{-1}r_j r_j^* B_j^{-1*}D_j^{-1*}x_j
\end{aligned}$$

$$-trD_j^{-1}B_j^{-1}r_jr_j^*B_j^{-1*}D_j^{-1*}|^6$$

$$+\frac{K}{N^{12}}E|r_j^*B_j^{-1*}D_j^{-1*}D_j^{-1}B_j^{-1}r_j|^6$$

$$\leq K\frac{(\ln(n))^{30}}{N^{12}}+\frac{K}{N^9}(\ln(n))^{12}E\|D_j^{-1}B_j^{-1}r_jr_j^*B_j^{-1*}D_j^{-1*}\|^6$$

$$+\frac{K}{N^{12}}E\|B_j^{-1*}D_j^{-1*}D_j^{-1}B_j^{-1}\|^6\|r_j\|^{12}$$

$$\leq K\frac{(\ln(n))^{30}}{N^{12}}+K\frac{(\ln(n))^{18}}{N^3}+K\frac{(\ln(n))^6}{N^6}$$

$$\leq K\frac{(\ln(n))^{30}}{N^3},$$

and similarly

$$E|\hat{\beta}_j|^{12}\leq K\frac{(\ln(n))^{30}}{N^3}.$$

Using (3.1), (3.6), Lemma 3.2, and the Cauchy-Schwarz inequality we have

$$E|\gamma_j|^{12}=\frac{\sigma^{12}}{N^{12}}E\left||x_j^*D_j^{-1}r_j|^2\right|^6=\frac{\sigma^{12}}{N^{12}}E|x_j^*D_j^{-1}r_jr_j^*D_j^{-1*}x_j|^6$$

$$\leq\frac{K}{N^{12}}E|x_j^*D_j^{-1}r_jr_j^*D_j^{-1*}x_j-trD_j^{-1}r_jr_j^*D_j^{-1*}|^6$$

$$+\frac{K}{N^{12}}E|r_j^*D_j^{-1*}D_j^{-1}r_j|^6$$

$$\leq\frac{K}{N^9}(\ln(n))^{12}E\|D_j^{-1}r_jr_j^*D_j^{-1*}\|^6+\frac{K}{N^{12}}E\|D_j^{-1}\|^{12}\|r_j\|^{12}$$

$$\leq K\frac{(\ln(n))^{18}}{N^3}+K\frac{(\ln(n))^6}{N^6}$$

$$\leq K\frac{(\ln(n))^{18}}{N^3},$$

and similarly

$$E|\beta_j|^{12}\leq K\frac{(\ln(n))^{18}}{N^3}.$$

From the Cauchy-Schwarz inequality and the above bounds we get

$$E|\beta_j\hat{\gamma}_j|^6\leq K\frac{(\ln(n))^{24}}{N^3}.$$

Therefore, using (3.7) with the above, we have as $n\to\infty$

$$\max_{j\leq N}\max\left\{\left|\frac{(\sigma^2c_nm_n-\omega_j)\hat{\rho}_j}{1+\sigma^2c_nm_n}\right|,\left|\frac{\gamma_j\hat{\rho}_j}{1+\sigma^2c_nm_n}\right|,|\hat{\beta}_j|,|\hat{\gamma}_j|,\left|\frac{\rho_j\hat{\gamma}_j}{1+\sigma^2c_nm_n}\right|,\right.$$

$$\left.\left|\frac{\beta_j\hat{\gamma}_j}{1+\sigma^2c_nm_n}\right|,|\hat{\omega}_j-\frac{\sigma^2}{N}trD^{-1}B^{-1}|\right\}\xrightarrow{a.s.}0. \qquad (3.14)$$

We now concentrate on a realization for which (3.14) holds, $\{F^{C_n}\}$ is tight, and $F^{\frac{1}{N}RR^*}$ converges in distribution to H. From (3.14) we get that $\max_{j\leq N}|d^{n,j}|\to 0$ as $n\to\infty$. Therefore, using (3.5),

$$\frac{1}{n}tr(A_n-zI)^{-1}-m_n\to 0 \qquad \text{as} \quad n\to\infty.$$

Consider a subsequence $\{n_i\}$ on which m_{n_i} (bounded in absolute value by $\frac{1}{z_2}$) converges to a number m. We have $\mathcal{I}m\, m\geq\delta>0$. Let $b=1+\sigma^2cm$ and $b_n=1+\sigma^2c_nm_n$. Since $\mathcal{I}m\, zm_{n_i}\geq 0$ then $\mathcal{I}m\, zm\geq 0$. From this we find for all $t\geq 0$

$$\mathcal{I}m\,\left(\frac{t}{b}-bz+\sigma^2(1-c)\right)\leq -z_2<0,$$

and similarly

$$\mathcal{I}m\,\left(\frac{t}{b_{n_i}}-b_{n_i}z+\sigma^2(1-c_{n_i})\right)\leq -z_2<0.$$

Then

$$\left|\frac{1}{\frac{t}{b_{n_i}}-b_{n_i}z+\sigma^2(1-c_{n_i})}-\frac{1}{\frac{t}{b}-bz+\sigma^2(1-c)}\right|$$

$$=\left|\frac{\frac{t}{bb_{n_i}}(b_{n_i}-b)+z(b_{n_i}-b)+\sigma^2(c_{n_i}-c)}{\left(\frac{t}{b_{n_i}}-b_{n_i}z+\sigma^2(1-c_{n_i})\right)\left(\frac{t}{b}-bz+\sigma^2(1-c)\right)}\right|$$

$$\leq\frac{|z||b_{n_i}-b|+\sigma^2|c_{n_i}-c|}{z_2^2}+\frac{|b_{n_i}-b|}{z_2|b_{n_i}|}\left|\frac{\frac{t}{b}}{\frac{t}{b}-bz+\sigma^2(1-c)}\right|$$

$$=\frac{|z||b_{n_i}-b|+\sigma^2|c_{n_i}-c|}{z_2^2}+\frac{|b_{n_i}-b|}{z_2|b_{n_i}|}\left|1-\frac{-bz+\sigma^2(1-c)}{\frac{t}{b}-bz+\sigma^2(1-c)}\right|$$

$$\leq\frac{|z||b_{n_i}-b|+\sigma^2|c_{n_i}-c|}{z_2^2}+\frac{|b_{n_i}-b|}{z_2|b_{n_i}|}\left(1+\frac{|bz|+\sigma^2|1-c|}{|\frac{t}{b}-bz+\sigma^2(1-c)|}\right)$$

$$\leq\frac{|z||b_{n_i}-b|+\sigma^2|c_{n_i}-c|}{z_2^2}+\frac{|b_{n_i}-b|}{z_2\sigma^2c_{n_i}\delta}\left(1+\frac{|z|\left(1+\frac{\sigma^2c}{z_2}\right)+\sigma^2|1-c|}{z_2}\right),$$

which converges to zero uniformly in t. Therefore as $n_i\to\infty$

$$\frac{1}{n_i}tr(A_{n_i}-zI)^{-1}=\int\frac{1}{\frac{t}{b_{n_i}}-b_{n_i}z+\sigma^2(1-c_{n_i})}dF^{\frac{1}{N_i}RR^*}(t)$$

$$\longrightarrow\int\frac{1}{\frac{t}{b}-bz+\sigma^2(1-c)}dH(t).$$

Thus m satisfies (1.1).

Now, using the result from the next section we have that m is unique. Therefore, with probability one, F^{C_n} converges in distribution to the p.d.f. F having Stieltjes transform defined by (1.1), and the proof of Theorem 1.1 is complete.

4. Unique Solution to (1.1)

In this section it is shown that certain solutions to equation (1.1) are unique.

THEOREM 4.1. Let $z = z_1 + iz_2 \in \mathbb{C}^+$, $m = m_1 + im_2 \in \mathbb{C}^+$, and $\mathbf{m} = \mathbf{m}_1 + i\mathbf{m}_2 \in \mathbb{C}^+$ with $\mathcal{I}m\, mz \geq 0$, and $\mathcal{I}m\, \mathbf{m}z \geq 0$. If both m and $\mathbf{m}$ satisfy (1.1), then $m = \mathbf{m}$.

Proof. Define $b \equiv 1 + \sigma^2 cm = b_1 + ib_2$ and $\mathbf{b} \equiv 1 + \sigma^2 c\mathbf{m} = \mathbf{b}_1 + i\mathbf{b}_2$ and suppose that both m and $\mathbf{m}$ satisfy (1.1). We have that $m - \mathbf{m} = (m - \mathbf{m})\alpha$, where

$$\alpha = \sigma^2 c \int \frac{\frac{t}{b\mathbf{b}} + z}{(\frac{t}{b} - bz + \sigma^2(1-c))(\frac{t}{\mathbf{b}} - \mathbf{b}z + \sigma^2(1-c))} dH(t).$$

Using the triangle and Cauchy-Schwarz inequalities we get

$$\begin{aligned}
|\alpha| \;\leq\; & \sigma^2 c \int \frac{\frac{t}{|b||\mathbf{b}|} dH(t)}{\left|\frac{t}{b} - bz + \sigma^2(1-c)\right| \left|\frac{t}{\mathbf{b}} - \mathbf{b}z + \sigma^2(1-c)\right|} \\
& + \sigma^2 c|z| \int \frac{dH(t)}{\left|\frac{t}{b} - bz + \sigma^2(1-c)\right| \left|\frac{t}{\mathbf{b}} - \mathbf{b}z + \sigma^2(1-c)\right|} \\
\leq\; & \left(\int \frac{\sigma^2 c \frac{t}{|b|^2} dH(t)}{\left|\frac{t}{b} - bz + \sigma^2(1-c)\right|^2}\right)^{\frac{1}{2}} \left(\int \frac{\sigma^2 c \frac{t}{|\mathbf{b}|^2} dH(t)}{\left|\frac{t}{\mathbf{b}} - \mathbf{b}z + \sigma^2(1-c)\right|^2}\right)^{\frac{1}{2}} \\
& + |z| \left(\int \frac{\sigma^2 c\, dH(t)}{\left|\frac{t}{b} - bz + \sigma^2(1-c)\right|^2}\right)^{\frac{1}{2}} \left(\int \frac{\sigma^2 c\, dH(t)}{\left|\frac{t}{\mathbf{b}} - \mathbf{b}z + \sigma^2(1-c)\right|^2}\right)^{\frac{1}{2}} \\
\equiv\; & (g(b))^{\frac{1}{2}} (g(\mathbf{b}))^{\frac{1}{2}} + |z| (G(b))^{\frac{1}{2}} (G(\mathbf{b}))^{\frac{1}{2}}. \qquad (4.1)
\end{aligned}$$

Note that $g(b), g(\mathbf{b}) \geq 0$ and $G(b), G(\mathbf{b}) > 0$.

The following statements are valid for both b and $\mathbf{b}$.

From (1.1) we get

$$b_1 = 1 + b_1 g(b) + (\sigma^2(1-c) - \mathcal{R}e\, bz)G(b) \qquad (4.2)$$

$$b_2 = b_2 g(b) + (\mathcal{I}m\, bz)G(b), \qquad (4.3)$$

and (4.3) implies

$$b_1 = b_2 \frac{1 - g(b) - z_1 G(b)}{z_2 G(b)}. \qquad (4.4)$$

Since (4.2) can be written as

$$b_1(1 - g(b) + z_1G(b)) = (1 + \sigma^2(1-c)G(b)) + b_2z_2G(b)$$

we replace b_1 using (4.4) and get

$$b_2((1-g(b))^2 - |z|^2G^2(b)) = (1 + \sigma^2(1-c)G(b))z_2G(b) > 0$$

(recall $c \le 1$).
Therefore,

$$(1-g(b))^2 - |z|^2G^2(b) > 0. \tag{4.5}$$

Since $G(b) > 0$ and $\mathcal{I}m\, bz = z_2 + \sigma^2 c\mathcal{I}m\, mz > 0$, we have $g(b) < 1$ and hence (4.5) implies

$$0 < |z|G(b) < 1 - g(b).$$

We now have

$$g(b) < 1 - |z|G(b) \ \text{ and } \ g(\mathbf{b}) < 1 - |z|G(\mathbf{b}). \tag{4.6}$$

For real numbers x and y with $x, y \in [0,1]$ it is easy to show that

$$(1-x)^{\frac{1}{2}}(1-y)^{\frac{1}{2}} \le 1 - (xy)^{\frac{1}{2}} \tag{4.7}$$

with equality holding if and only if $x = y$.

To complete the theorem's proof we use (4.1), (4.6), and (4.7) to get

$$\begin{aligned} |\alpha| &\le (g(b))^{\frac{1}{2}}(g(\mathbf{b}))^{\frac{1}{2}} + |z|(G(b)G(\mathbf{b}))^{\frac{1}{2}} \\ &< (1-|z|G(b))^{\frac{1}{2}}(1-|z|G(\mathbf{b}))^{\frac{1}{2}} + |z|(G(b)G(\mathbf{b}))^{\frac{1}{2}} \\ &\le 1 - (|z|G(b))^{\frac{1}{2}}(|z|G(\mathbf{b}))^{\frac{1}{2}} + |z|(G(b)G(\mathbf{b}))^{\frac{1}{2}} = 1. \end{aligned}$$

Therefore $|\alpha| < 1$, and hence $m = \mathbf{m}$.

Appendix

Here we prove (2.3) which states that

$$\|F^{C_n} - F^{\hat{C}_n}\| \xrightarrow{a.s.} 0 \qquad \text{as} \quad n \to \infty.$$

First, we define $p_n \equiv P(|X_{11}| \ge \sqrt{n})$ and note that since $E|X_{11}|^2 < \infty$ we have $p_n = \frac{o(1)}{n}$.

Now, to prove (2.3) we will need the following theorem.

Theorem A.1. Let $X_1, X_2, ..., X_n$ be i.i.d. Bernoulli with $p \equiv P(X_1 = 1) < \frac{1}{2}$. Then for any $\epsilon > 0$ such that $p + \epsilon \le \frac{1}{2}$ we have

$$P\Big(\frac{1}{n}\sum_{i=1}^{n} X_i - p \ge \epsilon\Big) \le e^{-\frac{n\epsilon^2}{2(p+\epsilon)}}.$$

Proof. For $t > 0$

$$P\Big(\frac{1}{n}\sum_{i=1}^{n} X_i - p \geq \epsilon\Big) \leq e^{-tn(p+\epsilon)} E\left[e^{t\sum_{i=1}^{n} X_i}\right]$$

$$= \Big(pe^{t(1-(p+\epsilon))} + (1-p)e^{-t(p+\epsilon)}\Big)^n.$$

Minimizing over t we get

$$P\Big(\frac{1}{n}\sum_{i=1}^{n} X_i - p \geq \epsilon\Big) \leq \left[\left(\frac{1-p}{1-(p+\epsilon)}\right)^{1-(p+\epsilon)}\left(\frac{p}{p+\epsilon}\right)^{p+\epsilon}\right]^n \equiv e^{n\psi(p,\epsilon)},$$

where ψ is defined by

$$\psi(p,\epsilon) = (1-(p+\epsilon))\ln\left(1+\frac{\epsilon}{1-(p+\epsilon)}\right) + (p+\epsilon)\ln\left(1-\frac{\epsilon}{p+\epsilon}\right).$$

Now, using the Taylor series

$$\ln(1-x) = \sum_{k=1}^{\infty}\frac{x^k}{k} \quad \text{for} \quad |x| < 1,$$

we get

$$\psi(p,\epsilon) = -\sum_{k=2}^{\infty}\frac{\epsilon^k}{k}\left(\frac{1}{(p+\epsilon)^{k-1}} + \frac{(-1)^k}{(1-(p+\epsilon))^{k-1}}\right).$$

Since $p+\epsilon \leq \frac{1}{2}$, the terms in the sum are all nonnegative, and therefore, dropping all but the first term, we get

$$\psi(p,\epsilon) \leq -\frac{\epsilon^2}{2}\left(\frac{1}{p+\epsilon} + \frac{1}{1-(p+\epsilon)}\right) < -\frac{\epsilon^2}{2(p+\epsilon)}$$

and the theorem is proven.

We now prove (2.3) by first noting that for $\epsilon > 0$ we get from Lemma 2.2 and (MP2)

$$P\Big(\|F^{C_n} - F^{\hat{C}_n}\| \geq \epsilon\Big) \leq P\Big(\frac{2}{n}\sum_{i,j} 1_{(|X_{ij}| \geq \sqrt{n})} \geq \epsilon\Big)$$

$$= P\Big(\frac{1}{Nn}\sum_{i,j} 1_{(|X_{ij}| \geq \sqrt{n})} - p_n \geq \frac{\epsilon}{2n} - p_n\Big).$$

Since $p_n = \frac{o(1)}{n}$, for any $\epsilon \in (0, \frac{1}{2})$ we can apply Theorem A.1 to get for all n large

$$P\Big(\|F^{C_n} - F^{\hat{C}_n}\| \geq \epsilon\Big) \leq e^{-\frac{n\epsilon}{16}}$$

when $p_n < \frac{\epsilon}{4n}$. Therefore $P\Big(\|F^{C_n} - F^{\hat{C}_n}\| \geq \epsilon\Big)$ is summable, and hence $\|F^{C_n} - F^{\hat{C}_n}\| \xrightarrow{a.s.} 0$ as $n \to \infty$ which proves (2.3).

References

[1] Horn, R. A. and Johnson, C. R. (1985). *Matrix Analysis.* Cambridge Univ. Press.

[2] Schmidt, R. O. (1986). Multiple emitter location and signal parameter estimation, *IEEE Transactions on Antennas and Propagation* **AP-34**(3) 276-280.

[3] Silverstein, J. W. (1995). Strong convergence of the empirical distribution of the eigenvalues of large dimensional random matrices, *J. Multivariate Anal.* **5** 331-339.

[4] 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.

[5] Silverstein, J. W. and Combettes, P. L. (1992). Signal detection via spectral theory of large dimensional random matrices, *IEEE Transactions on Signal Processing* **40**(8) 2100-2105.