Large System Spectral Analysis of Covariance Matrix Estimation

Husheng Li, Member, IEEE, and H. Vincent Poor, Fellow, IEEE

Abstract—Eigendecomposition of estimated covariance matrices is a basic signal processing technique arising in a number of applications, including direction-of-arrival estimation, power allocation in multiple-input/multiple-output (MIMO) transmission systems, and adaptive multiuser detection. This paper uses the theory of non-crossing partitions to develop explicit asymptotic expressions for the moments of the eigenvalues of estimated covariance matrices, in the large system asymptote as the vector dimension and the dimension of signal space both increase without bound, while their ratio remains finite and nonzero. The asymptotic eigenvalue distribution is also obtained from these eigenvalue moments and the Stieltjes transform, and is extended to first-order approximation in the large sample-size limit. Numerical simulations are used to demonstrate that these asymptotic results provide good approximations for finite systems of moderate size.

Index Terms—Covariance matrix, free cumulants, non-crossing partition, spectrum analysis.

I. INTRODUCTION

EIGENDECOMPOSITION is widely used in the context of vector signal processing. In eigendecomposition (also known as the subspace method), signal and noise subspaces are separated by carrying out eigendecomposition of the covariance matrix of the received signal. Once signal and noise subspaces are known, the information conveyed by vector signals is inferred by further processing based on the eigenvalues and eigenvectors associated with these subspaces.

A few examples of applications employing eigendecomposition are given as follows. Note that eigendecomposition is applied in many other domains, such as principal component analysis (PCA), as well.

- Direction of arrival (DOA) estimation, in which the vector signals are determined by the locations of the transmitters. Both the MUltiple Signal Classification (MUSIC) [20] and Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) [24] algorithms are based on knowledge of the signal subspace, which is obtained from eigendecomposition of a covariance matrix.
- Multiple-input/multiple-output (MIMO) systems, in which the vector signals represent the channel gains from a specific transmit antenna to multiple receive antennas. In MIMO systems, the covariance matrix of channel gains is eigendecomposed in order to facilitate water-filling power allocation when the channel state information is available at the transmitter [30].
- Subspace-based multiuser detection, in which spreading codes are assigned to different users. In code-division multiple access (CDMA) systems, multiuser detection can be carried out by projecting the received signal into the signal subspace, thus reducing the dimensionality of signals and eliminating the necessity of knowing the spreading codes of interfering users since the signal subspace can be obtained blindly via eigendecomposition [35].

All eigendecomposition based signal processing techniques are based on knowledge of the covariance matrix of a received signal. However, this covariance matrix is unknown to the receiver in most applications, and thus it must be estimated from the received signal. The precision of this covariance matrix estimation directly impacts the performance of eigendecomposition and further processing.

Therefore, it is of considerable interest to analyze the impact of covariance matrix estimation error on the eigendecomposition. However, it is difficult to obtain explicit error expressions for finite systems and finite sample sizes, in which cases only some bounds (e.g., the Davis–Kahan sinθ Theorem characterizing the perturbation of characteristic subspaces [5]) can be obtained. Hence, many studies (e.g., [9], [10], [36]) assume that the sample size is sufficiently large and the covariance matrix estimate is sufficiently close to the true value to allow asymptotic analysis. However, such large sample analysis is not realistic in some applications. For example, in the uplink of cdma2000 systems, a typical spreading gain, namely the dimension of the signal, is 64 and the chip rate is 1.2288 Mcps. So, for collecting 6400 samples (100 times the dimension) for estimating the covariance matrix, the receiver must wait for $\frac{64 \times 100 \times 63}{1.2288 \times 10^6} = 0.33$ seconds, which is larger than the typical channel coherence time for a terminal moving at pedestrian speed of 1 m/s and using a carrier frequency of 800 MHz1 (a typical value of carrier frequency in cdma2000 systems). Hence, the channel has changed before sufficiently many samples can be collected for a large sample analysis to apply in this case. To address such situations, in this paper, we focus on the analysis of systems with relatively

1From the formula $T_c = \frac{v}{fr_c}$, where $T_c$ is the coherence time, $v$ is the speed of the transmitter, $c$ is the speed of light, and $f_c$ is the carrier frequency, the coherence time is approximately 0.0665 s.
small numbers of samples\(^2\) and in the large system limit, namely, the limit in which the system size (the number of transmitters and the dimension of vector signals) and the number of samples tend to infinity while keeping their ratio constant. We will obtain explicit expressions for estimation error in this large system regime, which we will demonstrate to be good approximations for finite systems via numerical simulations. Note that similar system analyses have been widely used in many domains, such as multiuser detection [3], [31], [32] and MIMO systems [14]. Note that [6] and [22] (the journal version of the latter is [23]) address problems similar to that of this paper via different approaches. While [6] tackles the problem using the Stieltjes transform and [22] applies the theory of free probability, the approach in this paper is to consider eigenvalue moments and the theory of non-crossing partitions. An important by-product of our analysis is that we have obtained closed-form expressions for eigenvalue moments.

In this paper, we thus focus on the asymptotic analysis of the spectrum (i.e., the set of eigenvalues) of covariance matrix estimates by applying the method of moments [4] via the following three steps.

1) Obtain explicit expressions for the eigenvalue moments by exploiting the structure of non-crossing partitions [28], [29]. The main results are given in Theorem 1 (for systems without noise) and Theorem 5 (for systems with noise). Particularly, we extended previous results on non-crossing partition based moment analysis of covariance matrices [16], [29] to the analysis of covariance matrix estimates.

2) Obtain the eigenvalue distribution by solving an equation characterizing the Stieltjes transform of the eigenvalue distribution, which is obtained from the eigenvalue moments. The main results are given in Theorem 3 (for all systems), Corollary 6 (for systems without noise), and Corollary 13 (for systems with noise). Moreover, Theorem 2 assures the uniqueness and weak convergence of the eigenvalue distribution. The results are also extended to the large sample limit, which characterizes the perturbation in both large system and large sample limits.

3) Obtain qualitative properties of the eigenvalue distribution and the corresponding impact on the practical algorithms mentioned before. The main results are given in Theorem 4 (for systems without noise) and Theorem 6 (for systems with noise).

The remainder of this paper is organized as follows. A general model for vector channels, which includes single-user channel as a special case, is given in Section II. A preliminary introduction to non-crossing partitions and their relationship to eigenvalue moments is given in Section III. Section IV is focused on the spectral analysis of systems without noise, which is extended to systems with noise in Section V and to the large sample case in Section VI. Simulation results and conclusions are given in Sections VII and VIII, respectively.

\(^2\)Here, “relatively small number” means that the number of samples tends to infinity while the ratio between the number of samples and the signal dimension is bounded.

The following defines some mathematical notation used in this paper.

- \([A]\) is the cardinality of the set \(A\); \(A/B = \{x \mid x \in A, x \notin B\} \) denotes the complementary set of \(A\).
- \(X^{H}\) is the conjugate transpose of the matrix \(X\); \(X^{\dagger}\) is the transpose of the matrix \(X\).
- For an \(n \times n\) matrix, \(\text{Trace}(A) = \sum_{i=1}^{n} A_{ii}\) and \(\text{det}(A)\) denotes the determinant of \(A\).
- \(\lfloor x \rfloor\) denotes the smallest integer not less than \(x\), and \(\text{mod}(i, j)\) denotes the modulus of \(i\) with respect to \(j\), with the convention that \(\text{mod}(i, i) = i\).
- For \(x \in \mathbb{C}\), \(\text{Re}(x)\) denotes the real part of \(x\) and \(\text{Im}(x)\) denotes the imaginary part of \(x\).
- \(\delta(\cdot)\) denotes the Dirac delta function.
- \(\mathbb{C}^+ \triangleq \{z \in \mathbb{C} \mid \text{Im}(z) > 0\}\) and \(\mathbb{R}^+ \triangleq \{x \in \mathbb{R} \mid x > 0\}\).

II. MODEL OF VECTOR CHANNELS

A. Signal Model

We consider a discrete-time complex-valued \(K\)-user \(N\)-dimensional vector channel with \(M\) channel uses. For simplicity, we define the normalized sample size \(\alpha \triangleq \frac{N}{K}\) and the system load \(\beta \triangleq \frac{M}{K}\) (for single-user channels, we simply let \(\beta = 0\)). We assume that the system load \(\beta < 1\) \((K < N)\); otherwise, the signal subspace is simply the entire \(N\)-vector space. In the \(m\)th channel use, the signal at the receiver can be represented by an \(N\)-vector, which is given by

\[
r(m) = \sum_{k=1}^{K} x_{km} s_k + n(m) \tag{1}
\]

where \(x_{km}\) is the channel symbol of user \(k\), having unit power, \(s_k\) is the signature waveform of user \(k\) (note that \(s_k\) is independent of the sample index \(m\)), and \(n(m)\) is additive noise. On defining

\[
S_{N \times K} = (s_1, \ldots, s_K)
\]

and

\[
x_{K \times 1}(m) = (x_{1m}^*, \ldots, x_{Km}^*)^H
\]

(1) can be rewritten as

\[
r(m) = S x(m) + n(m). \tag{2}
\]

We make the following assumptions for the signal model. It should be noted that we do not assume specific distribution laws of the elements in \(S, x,\) and \(n\), thereby making the channel model more general.

- The elements of \(S\) are mutually independent random variables, each having zero expectation and variance \(\frac{1}{N}\).

Therefore, \(\forall k, \|s_k\| \rightarrow 1\) almost surely, as \(N \rightarrow \infty\).

- The elements of \(x(m)\) are mutually independent random variables. The random vectors \(\{x(m)\}_{m=1,\ldots,M}\) are mutually independent for different values of \(m\) and satisfy \(E[x(m)x^H(m)] = I_{K \times K}\) and \(E[x(m)x^T(m)] = 0_{K \times K}\).

- The elements of \(n(m)\) are mutually independent random variables. The random vectors \(\{n(m)\}_{m=1,\ldots,M}\) are mutually dependent for different values of \(m\).
and satisfy $E\{n(m)n^H(m)\} = \sigma_n^2 I_{N\times N}$ and $E\{n(m)n^T(m)\} = 0_{N\times N}$.

- $\mathbf{S}, \mathbf{x}(m)$, and $n(m)$ are jointly independent. Such a signal model is useful in many applications. Two examples are given as follows.

- When the column vectors of $\mathbf{S}$ are considered to be spreading codes, the signal model represents a synchronous short-code CDMA system with $K$ active users and spreading gain $N$.

- When the elements of $\mathbf{S}$ are considered to be channel gains from $K$ transmit antennas to $N$ receive antennas, the signal model represents a single-user MIMO system with an $N \times K$ channel matrix.

### B. Covariance Matrix Estimation

The covariance matrix of the received signal (1) is given as follows: (Note that the expectation here is conditioned on the realization of the spreading code matrix $\mathbf{S}$)

$$
\mathbf{R} \triangleq E\{\mathbf{r}(m)\mathbf{r}^H(m)\} = \mathbf{SS}^H + \sigma_n^2 I_{N\times N}.
$$

(3)

As explained in the Introduction, when no a priori information is available, the covariance matrix must be estimated from the received signal. In this paper, we adopt the unbiased sample covariance matrix estimate [1], which is given by (based on (2) and the assumption that $E\{\mathbf{x}(m)\mathbf{x}^H(m)\} = I_{K\times K}$)

$$
\hat{\mathbf{R}} = \frac{1}{M-1} \sum_{m=1}^{M} \mathbf{r}(m)\mathbf{r}^H(m)
$$

$$
= \frac{1}{M-1} (\mathbf{SX} + \mathbf{N})(\mathbf{SX} + \mathbf{N})^H
$$

where $\mathbf{X} \triangleq (x_1, \ldots, x_M)$ and $\mathbf{N} \triangleq (n_1, \ldots, n_M)$. Note that the denominator $M-1$ can be replaced by $M$ in the large system limit $M \to \infty$. Therefore, for notational simplicity, we use $M$ here instead of $M-1$ throughout this paper.

### III. PRELIMINARIES ON NON-CROSSING PARTITION BASED SPECTRAL ANALYSIS

The eigenvalue distribution of large random matrices is deeply related to the lattice structure of non-crossing partitions, as disclosed in [28] and [29]. In this section, we provide some preliminaries concerning non-crossing partitions and their relationship to the eigenvalue distribution of large random covariance matrices.

#### A. Non-Crossing Partitions

1) **Partitions**: A partition on a set $\{1, \ldots, p\}$ (note that the set has a complete ordering $1 < 2 < \cdots < p$) is defined as a division of the elements into a group of disjoint subsets, which we call blocks (a block is called an $i$-block when the block size is $i$; we keep the ordering of elements within the blocks). We denote by $i \rightarrow j$ that elements $i$ and $j$ are within the same block. A partition is called an $r$-partition when the number of blocks is $r$. For example, $\{1, 3, 4, 6\}, \{2, 5\}, \{7\}, \{8\}$ is a 4-partition of $\{1, 2, 3, 4, 5, 6, 7, 8\}$. We denote by $0_p$ the finest partition $\{\{1\}, \ldots, \{p\}\}$ and by $1_p$ the coarsest partition $\{1, \ldots, p\}$. For a partition $\pi$, we denote by $B(\pi)$ the number of blocks in $\pi$.

2) **Non-Crossing Partitions**: We call a partition of a $p$-set non-crossing if, for any two blocks $\{u_1, \ldots, u_s\}$ and $\{v_1, \ldots, v_t\}$, we have, $\forall k = 1, \ldots, s$

$$
v_k < u_1 < u_{k+1} \Leftrightarrow u_k < v_1 < v_{k+1}
$$

with the convention that $u_{s+1} = u_1$. For example, for the set $\{1, 2, 3, 4, 5, 6, 7, 8\}$, $\{1, 4, 5, 6\}, \{2, 3\}, \{7\}, \{8\}$ is non-crossing, while $\{1, 3, 4, 6\}, \{2, 5\}, \{7\}, \{8\}$ is not. Note that, in the context of non-crossing partitions, addition of the index means modular addition with respect to the cardinality of the corresponding set. A non-crossing partition can be illustrated by the circle graph, as shown in Fig. 1, in which the elements are represented by circular points on a circle in clockwise order, and the elements within the same block are connected with solid lines. Intuitively, a partition is non-crossing if and only if the lines belonging to different blocks do not intersect. Note that $0_p$ and $1_p$ are trivial examples of non-crossing partitions.

We denote different subsets of all non-crossing partitions on the set $\{1, 2, \ldots, p\}$ using the following notation.

- $\mathbb{N}_{C_{p}}$: the set of all non-crossing partitions. It is shown in [28] that

$$
|\mathbb{N}_{C_{p}}| = \frac{1}{p} \binom{2p}{p-1}.
$$

(6)

- $\mathbb{N}_{C_{p,r}}$: the set of all non-crossing $r$-partitions. It is shown in [8] that

$$
|\mathbb{N}_{C_{p,r}}| = \frac{1}{p} \binom{p}{r} \binom{p}{r-1}.
$$

(7)

- $\mathbb{N}_{C_{p,l}}(i_1, \ldots, i_r)$: the set of non-crossing $r$-partitions with blocks having cardinalities $i_1, \ldots, i_r$. It is shown in [16], [17] that

$$
|\mathbb{N}_{C_{p,l}}(i_1, \ldots, i_r)| = \frac{p!}{(p-r+1)!M(i_1, \ldots, i_r)}
$$

(8)

where $M(i_1, \ldots, i_r) \triangleq \prod_{k\geq 1} n_k!$ and $n_k$ denotes the number of elements equaling $k$ in $\{i_j\}_{j=1, \ldots, r}$ with the

![Circle graph of the non-crossing partition](image-url)
3) Dual Partition: The dual partition of a non-crossing partition, first defined in [13], can be defined intuitively via the circle graph. In the circle graph of the non-crossing partition \( \pi \) on \( \{1, \ldots, p\} \), we plot one square point on the arc connecting elements \( i \) and \( i+1 \), \( \forall i \in \{1, \ldots, p\} \) and label it by \((i+1)\)' then, we connect as many square points as possible using dashed lines under the constraint that no dashed line intersects the solid lines. Such a procedure results in a partition on the square points \( \{1', \ldots, p'\} \) if any two square points connected by a dashed line belong to the same block, which can be shown to be non-crossing [27] and is called the dual partition, denoted by \( \pi^* \), of the prime non-crossing partition \( \pi \). It can be shown that if \( \pi \) has \( r \) blocks, then \( \pi^* \) has \( p - r + 1 \) blocks [27]. We illustrate the dual partition of \( \{\{1, 4, 5, 6\}, \{2, 3\}, \{7\}, \{8\}\} \) in Fig. 2 (such a circle graph is called a complementation map in [13]).

4) Isomorphic Decomposition and Multiplicative Functions: The set of non-crossing partitions in \( \mathbb{N}C_p \) has a partial order structure, in which \( \pi \leq \sigma \) if each block of \( \pi \) is a subset of a corresponding block of \( \sigma \). For instance, we have \( \pi \leq \sigma \) if \( \pi = \{\{1, 4\}, \{5, 6\}, \{2, 3\}, \{7\}, \{8\}\} \) and \( \sigma = \{\{1, 4, 5, 6\}, \{2, 3\}, \{7\}, \{8\}\} \). Then, for any \( \pi \leq \sigma \in \mathbb{N}C_p \), we define the interval between \( \pi \) and \( \sigma \) as

\[
[\pi, \sigma] \triangleq \{\psi \in \mathbb{N}C_p | \pi \leq \psi \leq \sigma\}.
\]

It is shown in [28] that, for all \( \pi \leq \sigma \in \mathbb{N}C_p \), there exists a canonical sequence of positive integers \( \{k_j\}_{j \in \mathbb{N}} \) such that

\[
[\pi, \sigma] \cong \prod_{j \in \mathbb{N}} \mathbb{N}^{C_{k_j}}
\]

where \( \cong \) denotes isomorphism (the detailed mapping can be found in [28, proof of Proposition 1]), the product is the Cartesian product, and \( \{k_j\}_{j \in \mathbb{N}} \) is called the class of \( [\pi, \sigma] \). For a \( \pi \in \mathbb{N}C_p \), we are particularly interested in the intervals \([0, \pi] \) and \([\pi, 1_p] \), which have the following intuitive structures:

- As shown in [28], the canonical number \( k_i \) of \([0, \pi] \) is equal to the number of \( i \)-blocks in \( \pi \).
- As shown in [29], the canonical number \( k_i \) of \([\pi, 1_p] \) is equal to the number of \( i \)-blocks in \( \pi^* \).

The incidence algebra on the partial order structure of \( \mathbb{N}C_p \) is defined as the set of all complex-valued functions \( f(\psi, \sigma) \) with the property that \( f(\psi, \sigma) = 0 \) if \( \psi \preceq \sigma \) [18]. An important function in the incidence algebra is the \( \zeta \) function, which is defined as

\[
\zeta(\pi, \sigma) \triangleq \begin{cases} 1, & \text{if } \psi \preceq \sigma \\ 0, & \text{else} \end{cases}
\]

The combinatorial convolution between two functions \( f \) and \( g \) in the incidence algebra is defined as

\[
f \ast g(\pi, \sigma) \triangleq \sum_{\pi \leq \psi \leq \sigma} f(\pi, \psi)g(\psi, \sigma), \quad \forall \pi \leq \sigma.
\]

As an important subset of the incidence algebra, a multiplicative function \( f_\lambda \) on \([\pi, \sigma] \) is defined as

\[
f_\lambda \triangleq f_\psi \ast \zeta
\]

B. Eigenvalues of Large Random Covariance Matrices

1) Free Cumulants: Denote generically the random eigenvalue of a large random covariance matrix \( SS^\dagger \) by \( \lambda \) and the \( p \)th asymptotic moment by \( \lambda_p = \lim_{p \to \infty} \frac{1}{\sqrt{p}} \mathbb{E}[\lambda^p] \). The series \( \{\lambda_k\}_{k \in \mathbb{N}} \) is related to the non-crossing partitions via the free cumulants \( \{c_j\}_{j \in \mathbb{N}} \), which satisfy [29]

\[
f_\lambda = f_c \ast \zeta
\]

The original definition of the free cumulants is as the coefficients of the R-transformation given in [34]. Here, we adopt the simple relationship with combinatorial convolution found in [29] for the definition of free cumulants.
which can be translated into the following recursive equation [28]:

\[
\lambda_p = \sum_{r=1}^{p} c_r \sum_{\bar{i}_1, \ldots, \bar{i}_r \geq 0} \prod_{j=1}^{r} \lambda_{i_j}
\]

(16)

where the indices \(i_1, \ldots, i_r\) are distinguishable.\(^4\)

We define the generating functions for the eigenvalue moments and free cumulants, respectively, by

\[
\Lambda(z) \triangleq 1 + \sum_{k=1}^{\infty} \lambda_k z^k, \quad \forall z \in \mathbb{C},
\]

(17)

and

\[
C(z) \triangleq 1 + \sum_{j=1}^{\infty} C_j z^j, \quad \forall z \in \mathbb{C}.
\]

(18)

It is shown in [29] that (15) implies the following simple relationship between \(\Lambda(z)\) and \(C(z)\):

\[
\Lambda(z) = C(z\Lambda(z)).
\]

(19)

2) Eigenvalue Moments and Non-Crossing Partitions: For a generic eigenvalue \(\lambda\) of \(SS^H\), it is easy to verify that [12], [16]

\[
E\{\lambda^p\} = \frac{1}{N^{p+1}} \sum_{k_1, \ldots, k_p=1}^{K} \sum_{n_1, \ldots, n_p=1}^{N} E\left\{v_{k_1 n_1}^* v_{k_2 n_2}^* \cdots v_{k_p n_p}^* v_{k_p n_p}^* \right\}
\]

(20)

where \(v_{kn}\) is the \(n\)th element of \(\sqrt{N} s_k\), which has unit variance.

As \(K, N \to \infty\), the computation of (20) becomes a combinatorial problem of how to partition the indices \(\{k_j\}_{j=1, \ldots, p}\) and \(\{n_j\}_{j=1, \ldots, p}\) into blocks, in each of which the indices assume the same value. For example, \(k_1 \sim k_2\) means that \(k_1 = k_2\) in the summation of (20). Then, the sum is determined by the partitions yielding positive and non-vanishing (which means that the expectation does not converge to 0 as \(K, N \to \infty\)) expectations. The following lemma [29], [16], [15] discloses the structure of all positive and non-vanishing partitions.

Lemma 1: A partition of \(\{k_j\}_{j=1, \ldots, p}\) and \(\{n_j\}_{j=1, \ldots, p}\) is positive and non-vanishing if and only if the partition of \(\{k_j\}_{j=1, \ldots, p}\) is non-crossing and the partition of \(\{n_j\}_{j=1, \ldots, p}\) is dual to \(\pi\).

We call the condition in the lemma the mutual dual condition. On combining (14) and (15), using the fact that \(\lambda_p = f_N(0, 1, \mathbb{P})\) and the non-crossing structure of indices \(\{k_j\}_{j=1, \ldots, p}\) and \(\{n_j\}_{j=1, \ldots, p}\), we obtain

\[
\lambda_p = \sum_{\pi \in \mathbb{P}(\pi)} \beta^{|\pi|}
\]

(21)

since each block of \(\{k_j\}_{j=1, \ldots, p}\) contributes \(\beta^{|\pi|}\) and each block of \(\{n_j\}_{j=1, \ldots, p}\) contributes 1 (recall that \(\mathbb{P}(\pi)\) denotes the number of blocks in partition \(\pi\)). By comparing (21) with (14), we obtain

\[
c_p = \beta, \quad \forall p \in \mathbb{N}
\]

(22)

from which we can obtain the following:

1) an explicit expression for \(\lambda_p\) [12], [16] by combining (7), (14), (15), and (22), which is given by

\[
\lambda_p = \sum_{k=1}^{p} \frac{\beta^k}{k!} \left( \frac{p}{k} \right) \left( \frac{p}{k-1} \right)
\]

(23)

2) a recursive expression for \(\lambda_p\) [16], [29] by combining (16) and (22), which is given by

\[
\lambda_p = \sum_{k=1}^{p} \beta \sum_{i_1 + \cdots + i_k = p} \prod_{j=1}^{k} \lambda_{i_j}
\]

(24)

where the indices \(i_1, \ldots, i_k\) are distinguishable;

3) an equation characterizing the corresponding Stieltjes transformation of the eigenvalue distribution (the Tse–Hanley equation characterizing the performance of the linear minimum mean-square error (MMSE) multiuser detector is based on this equation [32]) by combining (19) and (22), which is given by

\[
m(z) = \frac{1}{z - \frac{\beta}{1 + m(z)}}
\]

(25)

where \(m(z)\) is the Stieltjes transform, whose definition will be given later;

4) the probability density function (PDF) of \(\lambda\) (called the Marcenko–Pastur law) obtained from 1) by comparing with Wigner’s semicircle law [12] or from 3) by contour integration [25], which is given by (when \(\beta < 1\))

\[
f(x) = (1 - \beta) \delta(0) + \sqrt{(x - \lambda_{\min})(\lambda_{\max} - x)}
\]

(26)

where the minimum and maximum of the positive eigenvalues are given by

\[
\lambda_{\min} = (1 - \sqrt{\beta})^2 \quad \text{and} \quad \lambda_{\max} = (1 + \sqrt{\beta})^2.
\]

IV. SPECTRAL ANALYSIS OF THE NOISE-FREE CASE

In this section, we discuss spectral analysis when the noise variance \(\sigma_n^2 = 0\), which is useful in interference dominated applications. Note that, when \(\sigma_n^2 = 0\), an explicit expression for eigenvalue moments has been obtained in [37] using graph theory, and the combinatorial convolution relation in Theorem 1 has been obtained in [34] using the theory of free probability. However, here we obtain the same results and detailed combinatorial structure using the theory of non-crossing partitions, which provides tools for the general case \(\sigma_n^2 \geq 0\).
Similar to the analysis of the exact covariance matrix $\mathbf{R}$ introduced in Section III, we first convert the problem of eigenvalue moments into a problem of index partitioning, obtain the set of positive and non-vanishing partitions, and then obtain an explicit expression for the eigenvalue moments. The properties of the eigenvalue distribution will be derived from the latter expression.

A. Non-Crossing Partitions in Eigenvalue Moments

When $\sigma^2 = 0$, the sample covariance matrix is given by

$$\hat{\mathbf{R}} = \frac{1}{M} \mathbf{SXX}^H \mathbf{s}^H.$$

(26)

On denoting by $\hat{\lambda}$ a generic eigenvalue of $\hat{\mathbf{R}}$ and defining $\hat{\lambda}_p \triangleq \lim_{K,N,M \to \infty} E\{\mathcal{M}\}$, we have

$$E\{\hat{\lambda}^p\} = \frac{1}{N} \text{Trace}(E\{\hat{\mathbf{R}}^p\})$$

$$= \frac{1}{NMP} \text{Trace} \left( E \left\{ \sum_{m_1, \ldots, m_p = 1}^{K} \mathbf{s}_{k_1}^H (m_1) \mathbf{x}_{k_1 \cdot m_1} \mathbf{s}_{k_2}^H (m_2) \mathbf{x}_{k_2 \cdot m_2} \ldots \mathbf{s}_{k_p \cdot m_p} \mathbf{x}_{k_p \cdot m_p} \right\} \right)$$

$$= \frac{1}{NMP} \text{Trace} \left( E \left\{ \sum_{m_1, \ldots, m_p = 1}^{K} \mathbf{s}_{k_1}^H (m_1) \mathbf{s}_{k_2}^H (m_2) \ldots \mathbf{s}_{k_p}^H (m_p) \right\} \right)$$

(27)

$$= \frac{1}{N^{p+1}M^p} \text{Trace} \left( \sum_{m_1, \ldots, m_p = 1}^{K} \mathbf{s}_{k_1}^H (m_1) \mathbf{s}_{k_2}^H (m_2) \ldots \mathbf{s}_{k_p}^H (m_p) \right)$$

(28)

and

$$\begin{align*}
(k_1, n_1), & \quad (k_2, n_2), \quad \ldots, \quad (k_{2p-1}, n_{2p-1}), \quad (k_{2p-1}, n_{2p}) \\
(k_{2p}, n_1), & \quad (k_{2p}, n_2), \quad \ldots, \quad (k_{2p-1}, n_{2p-1}), \quad (k_{2p}, n_{2p})
\end{align*}$$

(29)

the partition assumes that, for every index pair $(k_{2p-1}, m_a)$ in the upper row of Table (28), there exists a unique index pair $(k_{2p}, m_b)$ in the lower row of Table (28) such that $k_{2p-1} \sim k_{2p}$ and $m_a \sim m_b$ (recall that $\sim$ signifies membership in the same block); and the same holds for Table (29).

- Non-vanishingness, which means that the expectation in the summation does not vanish as $p \to \infty$. This requires $k + m + n \geq 2p + 1$ since the summation in (27) is scaled by $\frac{1}{N^{p+1}M^p} = O\left(\frac{1}{M^{p+1}}\right)$. Note that we allow the possibility of infinite moments $(k + m + n > 2p + 1)$. As will be seen, this possibility does not exist and therefore the non-vanishing condition is equivalent to $k + m + n = 2p + 1$.

For example, for the case $p = 8$, the following partition $(m = 3, n = 6, k = 8)$ is positive and nonvanishing:

$$\begin{align*}
\{m_j\}_{j=1,\ldots,p} : \{[1, 4, 5, 6], [2, 3], [7, 8]\}
\{n_j\}_{j=1,\ldots,p} : \{[1, 7], [2, 4], [3, 5], [6, 8]\}
\{k_j\}_{j=1,\ldots,2p} : \{[1, 12], [2, 7], [3, 6], [4, 5], [8, 9], [10, 11], [14, 15], [13, 16]\}
\end{align*}$$

(30)

from which we have $m + n + k = 17 = 2 * 8 + 1$ and the following matching pairs:

$$\begin{align*}
(k_1, m_1) & \sim (k_2, m_4) \sim (k_3, n_1) \sim (k_4, n_7) \\
(k_3, m_2) & \sim (k_4, m_3) \sim (k_5, n_2) \sim (k_6, n_4) \\
(k_5, m_3) & \sim (k_4, m_2) \sim (k_5, n_3) \sim (k_6, n_5) \\
(k_7, m_4) & \sim (k_6, m_1) \sim (k_7, n_4) \sim (k_8, n_2) \\
(k_9, m_5) & \sim (k_8, m_5) \sim (k_9, n_5) \sim (k_{10}, n_5) \\
(k_{11}, m_6) & \sim (k_{10}, m_5) \sim (k_{11}, n_6) \sim (k_{10}, n_6) \\
(k_{13}, m_7) & \sim (k_{12}, m_8) \sim (k_{13}, n_7) \sim (k_{16}, n_1) \\
(k_{15}, m_8) & \sim (k_{14}, m_7) \sim (k_{15}, n_8) \sim (k_{14}, n_8)
\end{align*}$$

(31)

If the partition of $\{m_j\}_{j=1,\ldots,p}$ is changed to

$$\{[1, 4], [5, 6], [2, 3], [7, 8]\}$$

, then the partition is not positive since it is easy to check that the expectation is 0. If the partition of $\{k_j\}_{j=1,\ldots,2p}^{1,\ldots,2p}$ is changed to

$$\{[1, 2, 7, 12], [3, 4, 5, 6], [8, 9, 10, 11], [13, 14, 15, 16]\}$$

then the partition is positive but vanishing since $m + k + n > 17$. Thus, the asymptotic expression for $E\{\hat{\lambda}^p\}$ is determined by the positive and nonvanishing partitions. Similar to Section III, a circle graph can be constructed to represent the indices via the following procedure. When $p = 8$, the circle graph corresponding to the partition in (30) is illustrated in Fig. 3.

- Plot $p$ circle points, 1, $\ldots$, $p$, representing the sample indices $\{m_j\}_{j=1,\ldots,p}$ and $p$ square points, $1', \ldots, p'$, representing the dimension indices $\{n_j\}_{j=1,\ldots,p}$ on a circle in the same way as the circle graph in Section III.

- Plot $2p$ star points, 1*, $\ldots$, $2p*$, representing the user indices $\{k_j\}_{j=1,\ldots,2p}$ on the circle such that, for each $j$, $j^*$ lies on
the arc connecting \( (\frac{j}{2}) \) and \( (\frac{j+1}{2}) \) if \( j \) is even, and on the arc connecting \( (\frac{j+1}{2}) \) and \( (\frac{j}{2} + 1) \) if \( j \) is odd.

- The indices within the same block are connected by solid, dashed, and dotted lines for \( \{m_j\}_{j=1,...,p} \), \( \{n_j\}_{j=1,...,p} \), and \( \{k_j\}_{j=1,...,2p} \), respectively.

The following lemma discloses the structure of the positive partitions via the circle graph.

**Lemma 2:** Every positive and nonvanishing partition can be constructed via the following procedure.

1. Connect a set of circle points, determined by the partition, representing \( \{m_j\}_{j=1,...,p} \) using solid lines in the circle graph to form a non-crossing partition of \( \{m_j\}_{j=1,...,p} \).
2. Connect some square points representing \( \{n_j\}_{j=1,...,p} \) using dashed lines in the circle graph to form a non-crossing partition for \( \{n_j\}_{j=1,...,p} \), satisfying the condition of not intersecting the solid lines.
3. Connect star points representing \( \{k_j\}_{j=1,...,2p} \) using dotted lines as many as possible within the constraint of not intersecting solid lines or dashed lines.

**Proof:** First, we need to show that these steps are necessary for yielding a positive and nonvanishing partition.

For simplicity, we assume \( M = N (\alpha = 1) \) since positive and nonvanishing partitions are concerned with only the exponents of \( M, N, p \), and \( K \), thus ensuring that the simplification does not change the set of positive and nonvanishing partitions.

We can rearrange the subscripts of the indices, convert this case to the case of an exact covariance matrix discussed in Section III, and then apply Lemma 1, as follows.

First, (27) can be rewritten as

\[
E\{\hat{\Lambda}^p\} = \frac{1}{N^{2p+1}} \sum_{n_1 = 1}^{N} \sum_{n_2 = 1}^{K} \sum_{n_3 = 1}^{N} \sum_{n_4 = 1}^{N} \times E \left\{ \tilde{v}_{k_2n_4} \tilde{v}_{k_1n_3} \cdots \tilde{v}_{k_2n_1} \tilde{v}_{k_1n_2} \right\}.
\]

Then, we define random variables as follows:

\[
y_{r^j,j^j} = \begin{cases} 
\tilde{v}_{k_2n_1} \tilde{v}_{k_1n_2} \cdots \tilde{v}_{k_2n_4} \tilde{v}_{k_1n_3} & \text{if } j \text{ is odd} \\
\tilde{v}_{k_2n_1} \tilde{v}_{k_1n_2} \cdots \tilde{v}_{k_2n_4} \tilde{v}_{k_1n_3} & \text{if } j \text{ is even}
\end{cases}
\]

and

\[
y_{r^j,j^j+1} = \begin{cases} 
\tilde{v}_{k_2n_1} \tilde{v}_{k_1n_2} \cdots \tilde{v}_{k_2n_4} \tilde{v}_{k_1n_3} & \text{if } j \text{ is odd} \\
\tilde{v}_{k_2n_1} \tilde{v}_{k_1n_2} \cdots \tilde{v}_{k_2n_4} \tilde{v}_{k_1n_3} & \text{if } j \text{ is even}
\end{cases}
\]

For example, when \( p = 4 \), the above definitions result in

\[
y_{r_1,s_1} = \tilde{v}_{k_2n_1} \tilde{v}_{k_1n_2} \tilde{v}_{k_2n_3} \tilde{v}_{k_1n_4} \quad y_{r_2,s_2} = \tilde{v}_{k_2n_1} \tilde{v}_{k_1n_2} \tilde{v}_{k_2n_4} \tilde{v}_{k_1n_3} \quad y_{r_3,s_3} = \tilde{v}_{k_2n_1} \tilde{v}_{k_1n_2} \tilde{v}_{k_2n_4} \tilde{v}_{k_1n_3} \quad y_{r_4,s_4} = \tilde{v}_{k_2n_1} \tilde{v}_{k_1n_2} \tilde{v}_{k_2n_3} \tilde{v}_{k_1n_4} \quad y_{r_5,s_5} = \tilde{v}_{k_2n_1} \tilde{v}_{k_1n_2} \tilde{v}_{k_2n_4} \tilde{v}_{k_1n_3} \quad y_{r_6,s_6} = \tilde{v}_{k_2n_1} \tilde{v}_{k_1n_2} \tilde{v}_{k_2n_3} \tilde{v}_{k_1n_4} \quad y_{r_7,s_7} = \tilde{v}_{k_2n_1} \tilde{v}_{k_1n_2} \tilde{v}_{k_2n_4} \tilde{v}_{k_1n_3} \quad y_{r_8,s_8} = \tilde{v}_{k_2n_1} \tilde{v}_{k_1n_2} \tilde{v}_{k_2n_3} \tilde{v}_{k_1n_4} \quad y_{r_9,s_9} = \tilde{v}_{k_2n_1} \tilde{v}_{k_1n_2} \tilde{v}_{k_2n_4} \tilde{v}_{k_1n_3} \quad y_{r_{10},s_{10}} = \tilde{v}_{k_2n_1} \tilde{v}_{k_1n_2} \tilde{v}_{k_2n_3} \tilde{v}_{k_1n_4} .
\]

Based on the above definitions, it is easy to check that (32) can be rewritten as

\[
E\{\hat{\Lambda}^p\} = \frac{1}{N^{2p+1}} \sum_{r_1,n_1}^{K} \sum_{r_2,n_2}^{K} \sum_{n_3 = 1}^{N} \sum_{n_4 = 1}^{N} \times E \left\{ \tilde{v}_{r_1n_1} \tilde{v}_{r_2n_2} \cdots \tilde{v}_{r_kn_k} \tilde{v}_{r_{1+p}n_{1+p}} \right\}.
\]

which degenerates to the expression for the exact covariance matrix case treated in Section III.

Note that the set of partitions of \( \{r_j\}_{j=1,...,2p} \) and \( \{s_j\}_{j=1,...,2p} \) induced by the positive and nonvanishing partitions of \( \{m_j\}_{j=1,...,p} \), \( \{n_j\}_{j=1,...,p} \), and \( \{k_j\}_{j=1,...,2p} \) is a subset of the positive and nonvanishing partitions of \( \{r_j\}_{j=1,...,2p} \) and \( \{s_j\}_{j=1,...,2p} \) arising in the problem of an exact covariance matrix. (The reason this is a subset is that, in (35), \( s_i \) and \( s_j \) may not be able to be partitioned into the same block since they may represent sample and dimension indices, respectively.) Then, by applying Lemma 1, we can see that the procedures in the current lemma are necessary.

Now, we need to show that the proposed procedure results in positive and nonvanishing partitions. The first step results in a non-crossing partition of the indices \( \{s_{2j}\}_{j=1,...,p} \) having even subscripts since each index \( s_{2j} \) corresponds to sample index \( m_{p+1-j} \). The second step results in a non-crossing partition of the indices \( \{s_{2j+1}\}_{j=1,...,p} \) having odd subscripts since each index \( s_{2j+1} \) corresponds to dimension index \( n_{p+1-j} \) under the constraint of not intersecting the lines generated in the first step. Therefore, the first two steps in the procedure yield a non-crossing partition \( \pi \) of the indices \( \{s_j\}_{j=1,...,2p} \). The third step yields a dual partition \( \pi^* \) on the indices \( \{r_j\}_{j=1,...,2p} \) since each index \( r_j \) corresponds to user index \( k_{2j+1} \). According to Lemma 1, these three steps result in a partition of indices \( \{s_j\}_{j=1,...,2p} \) and \( \{r_j\}_{j=1,...,2p} \) satisfying the mutual dual condition; therefore, the expectation is positive and nonvanishing. This completes the proof.

It is easy to check that (30) can be generated by the three steps in the above lemma. We still call the circle graph obtained in Lemma 2 the complementation map. Based on Lemma 2, we can obtain the following theorem, which provides an explicit
expression for the eigenvalue moments via combinatorial convolution. Note that we discuss the quantity $\alpha^p \lambda_p$ instead of $\lambda_p$ in order to simplify the analysis, as will be seen.

**Theorem 1:** When $\sigma_n^2 = 0$, denoting by $f_{\lambda}$ the multiplicative function associated with series $\{\alpha^p \lambda_p\}_{p \in \mathbb{N}_0}$, we have

$$f_{\lambda} = f_{\alpha} \ast f_{\beta} \ast \zeta$$

(36)

where $f_{\alpha}$ is associated with $(\alpha, \alpha, \ldots)$, $f_{\beta}$ is associated with $(\beta, \beta, \ldots)$, and the zeta function $\zeta$ is defined in (11).

**Proof:** From (27), we have

$$E\{(\alpha \lambda) P\} = \frac{1}{N^{2p+1}} \sum_{m_1, \ldots, m_p = 1}^M \sum_{k_1, \ldots, k_{2p} = 1}^K \sum_{m_{p+1} = 1}^N$$

$$\times E\{a_{k_2m_1}^* a_{k_2m_1} \cdots a_{k_{2p}m_{p+1}}^* a_{k_{2p}m_{p+1}} \} \times E\{x_{k_2m_1}^* x_{k_2m_1} \cdots x_{k_{2p}m_{p+1}}^* x_{k_{2p}m_{p+1}} \}$$

(37)

which is determined by the positive and nonvanishing partitions of $\{m_j\}_{j=1, \ldots, p}$ and $\{n_j\}_{j=1, \ldots, p}$ and $\{k_j\}_{j=1, \ldots, 2p}$ as $K, N \to \infty$. We follow the procedure in Lemma 2 to count all positive and nonvanishing partitions and then evaluate the asymptotic summation in (37).

For the first step, we fix a non-crossing partition with $m (1 \leq m \leq p)$ blocks of the sample indices $\{m_j\}_{j=1, \ldots, p}$ and denote by $\pi$. Each block contributes a factor $M$ to the product corresponding to the partition since all indices within the same block assume the same value ranging from 1 to $M$. Therefore, the non-crossing partition $\pi$ contributes $f_{\alpha}(0_p, \pi) = \alpha^m$ to $\lambda_p$.

Then, in the second step, we partition the dimension indices $\{n_j\}_{j=1, \ldots, p}$ into $n (1 \leq n \leq p)$ non-crossing blocks under the constraint of not intersecting lines generated in the first step. Denote one of such partitions of $\{n_j\}_{j=1, \ldots, p}$ by $\delta$. If $\delta \notin \pi^*$ (recall the definition of partial order in non-crossing partitions), then there exists a line connecting two dimension indices, which does not belong to the lines corresponding to $\pi^*$, and which intersects the lines in the first step and thus contradicts step two. Therefore, $\forall \delta$, we have $\delta \leq \pi^*$.

Recall that the number of blocks in $\pi^*$ equals $p - m + 1$ and suppose

$$\pi^* \cong \prod_{j=1}^{p-m+1} [0_{n_j}, 1_{n_j}]$$

which satisfies the condition that $\sum_{j=1}^{p-m+1} n_j = p$. For example, for the partition $\pi$ of $\{m_j\}_{j=1, \ldots, p}$ in Fig. 3, we have

$$\pi^* \cong [0_2, 1_2]^2 \times [0_1, 1_1]^4$$

where the two lattices $[0_2, 1_2]$ correspond to $\{n_3, n_7\}$ and $\{n_2, n_4\}$, and the four lattices $[0_1, 1_1]$ correspond to $\{n_3\}$, $\{n_5\}$, $\{n_6\}$, and $\{n_8\}$.

Take an arbitrary $1 \leq j \leq p - m + 1$ and dimension indices $\{n_j\}_{j=1, \ldots, n_j}$ corresponding to $[0_{n_j}, 1_{n_j}]$. For any two consecutive indices $m_j$ and $m_{j+1}$, their neighboring sample indices $m_j$ and $m_{j+1}$ are the same index ($l_j = l_{j+1} - 1$) or belong to the same block ($m_j \sim m_{j+1}$), since $\pi^*$ is dual to $\pi$. It is obvious that a line connecting the user index $k_{2j-1}$ between $n_{j}$ and $m_{j}$ and the user index $k_{2j+1}$ between $n_{j+1}$ and $m_{j+1}$ does not intersect any existing lines. Therefore, we can connect the user indices $k_{2j-1}$ and $k_{2j+1}$ and merge them into one point in the complementation map. For example, in the partition of Fig. 3, $k_2$ and $k_7$ can be merged into one point since $m_1 \sim m_4$.

By merging the user indices in the way described above, we obtain $n_2$ dimension indices and $n_2$ user indices for the indices corresponding to $[0_{n_j}, 1_{n_j}]$ (for example, in Fig. 3, $n_2$, $n_4$, $k_3$ (merged with $k_6$) and $k_1$ (merged with $k_7$)), which cannot be connected to indices other than themselves; otherwise, a line will intersect the line connecting $m_{j}$ and $m_{j+1}$ for some $j$. Therefore, similar to the case of an exact covariance matrix, the partition of the dimension and user indices satisfies the mutual dual condition and therefore the contribution to $\lambda_p$ is given by

$$\sum_{\omega \in \mathbb{N}_{n_j}^*} f_{\beta}(0_{n_j}, \omega) \zeta(\omega, 1_{n_j}) = f_{\beta} \ast \zeta(0_{n_j}, 1_{n_j}).$$

Thus, the contribution of partitions of all dimension and user indices is given by

$$\prod_{i=1}^{p-m+1} f_{\beta} \ast \zeta(0_{n_i}, 1_{n_i}) = f_{\beta} \ast \zeta(0_p, \pi^*)$$

$$= f_{\alpha} \ast f_{\beta} \ast \zeta(0_p, 1_p)$$

(38)

which is equivalent to (36). \hfill \Box

A direct corollary of Theorem 1 is that the eigenvalue moments are bounded, which excludes the possibility of $k + m + n > 2p + 1$ for positive and nonvanishing partitions and implies that the nonvanishing property is equivalent to $k + m + n = 2p + 1$.

Note that (36) can be obtained directly from the asymptotic freeness of $SS^H$ and $XX^H$ [29], [34]. However, our discussion discloses the detailed structure of the non-crossing partitions and provides a starting point for the more general case of $\sigma_n^2 \geq 0$.

**B. Asymptotic Eigenvalue Moments**

In the previous subsection, we have obtained the structure of the non-crossing partitions of indices in the computation of $\lambda_p$. Based on Theorem 1, we obtain the following corollaries concerning the properties of eigenvalue moments.

1) **Free Cumulants:** The free cumulants of eigenvalues in the covariance matrix estimate are given in the following corollary.

**Corollary 1:** When $\sigma_n^2 = 0$, for the random variable $c \lambda$, the free cumulants are given by

$$\hat{c}_n = \sum_{k=1}^{n} \frac{\alpha^k \beta^{n-k} - 1}{n} \binom{n}{k} \binom{n}{k-1}, \quad n \in \mathbb{N}_0$$

(39)
Proof: Comparing (15) and (36), we obtain that
\[
\hat{\ell}_n = f_\alpha * f_\beta(0_n, 1_n) = \sum_{\pi \in \mathcal{N}_n} f_\alpha(0_n, \pi)f_\beta(\pi, 1_n) = \sum_{\pi \in \mathcal{N}_n} \alpha^{B(\pi)}\beta^{B(\pi')} = \sum_{\pi \in \mathcal{N}_n} \alpha^{B(\pi)}\beta^{B(\pi) + 1} = \sum_{k=1}^{n} \sum_{\pi \in \mathcal{N}_n} \lambda_k^{\pi k}\beta^{n-k+1}
\]
which results in (39) by applying (7).

2) Explicit Expression for Eigenvalue Moments: The following explicit expression has been obtained in [37] via graph theory. We derive it from Theorem 1 directly (recall that \(M(i_1, \ldots, i_k)\) is defined in (8)).

Corollary 2: When \(\sigma^2_n = 0\), the \(p\)th asymptotic moment of the eigenvalues of the covariance matrix estimate \(\hat{R}\) is given by (note that the function \(M(i_1, \ldots, i_k)\) is defined in (8))
\[
\hat{\lambda}_p = \frac{\sum_{k=1}^{p} \alpha^{k-1} \sum_{i_1 + \ldots + i_k = p} \frac{p!}{(p-k+1)!M(i_1, \ldots, i_k)}}{\sum_{i_1 + \ldots + i_k = p} \frac{p!}{(p-k+1)!M(i_1, \ldots, i_k)}} \times \prod_{j=1}^{k} \left( \sum_{i_j} \frac{\beta_j}{i_j} \left( \frac{i_j}{i_j - 1} \right) \right)
\]
where the indices \(i_1, \ldots, i_k\) are indistinguishable.

Proof: On denoting by \(f_\lambda\) the multiplicative function associated with the eigenvalue moments \(\{\lambda_p\}_{p \in \mathbb{N}}\) of the exact covariance matrix, we have
\[
\alpha^p\hat{\lambda}_p = \sum_{\pi \in \mathcal{N}_p} f_\lambda(0_p, \pi)f_\alpha(\pi, 1_p) = \sum_{\pi \in \mathcal{N}_p} f_\lambda(0_p, \pi)\alpha^{B(\pi) + 1} = \sum_{k=1}^{p} \alpha^{p-k+1} \sum_{i_1 + \ldots + i_k = p} \frac{p!}{(p-k+1)!M(i_1, \ldots, i_k)} \times \prod_{j=1}^{k} \lambda_i^{i_j}
\]
where the indices \(i_1, \ldots, i_k\) are indistinguishable since the summation is concerned with only the cardinalities of blocks.

Thus, (40) is obtained by substituting (8) and (23) into (41).

We list the first four moments of eigenvalues obtained from (40) as follows:
\[
E\{\hat{\lambda}\} = \beta
\]

\[
E\{\hat{\lambda}^2\} = \left(\frac{1}{\alpha} + 1\right)\beta^2 + \beta
\]

\[
E\{\hat{\lambda}^3\} = \left(\frac{1}{\alpha^2} + \frac{3}{\alpha} + 1\right)\beta^3 + 3\left(\frac{1}{\alpha} + 1\right)\beta^2 + \beta
\]

\[
E\{\hat{\lambda}^4\} = \left(\frac{1}{\alpha^3} + \frac{6}{\alpha^2} + \frac{6}{\alpha} + 1\right)\beta^4
\]

An interesting observation is that the expectation of \(\hat{\lambda}\) is independent of \(\alpha\). From (40), we can see that \(\hat{\lambda}_p\) decreases in \(\alpha\) and converges to \(\lambda_p\) as \(\alpha \to \infty\) when \(p > 1\). We have the following corollary to Theorem 1.

Corollary 3: For the asymptotic eigenvalue moments \(\{\hat{\lambda}_p\}\), we have
\[
\hat{\lambda}_p \begin{cases} = \lambda_p & \text{if } p = 1 \\ > \lambda_p & \text{if } p > 1. \end{cases}
\]

3) Recursive Expression for the Eigenvalue Moments: We can obtain a recursive expression for the eigenvalue moments from (16) and (39), similar to (25). That is, we have the following result.

Corollary 4: When \(\sigma^2_n = 0\), the asymptotic eigenvalue moments of \(\hat{R}\) are given recursively by
\[
\hat{\lambda}_p = \frac{\sum_{k=1}^{p} \alpha^{p-k+1} \sum_{i_1 + \ldots + i_k = p} \frac{p!}{(p-k+1)!M(i_1, \ldots, i_k)}}{\sum_{i_1 + \ldots + i_k = p} \frac{p!}{(p-k+1)!M(i_1, \ldots, i_k)}} \times \prod_{j=1}^{k} \hat{\lambda}_{i_j}, \quad \forall p \in \mathbb{N}
\]
where the indices \(i_1, \ldots, i_k\) are distinguishable.

4) Eigenvalue Moment Equivalence: The following corollary to Theorem 1 shows the asymptotic equivalence of the eigenvalue moments of two different matrices, which will be extended to systems with noise in the next section.

Corollary 5: The eigenvalue moments of the matrix \(\frac{1}{M}YSS^HY^H\), where \(Y\) is an \(M \times N\) matrix with mutually independent elements having unit variance, are the same as those of the matrix \(\frac{1}{M}X^SH^SX\).

Proof: It is easy to check that \(\frac{1}{M}YSS^HY^H\) is the covariance matrix estimate of a system with \(\alpha' = \frac{\beta}{\alpha}\) and \(\beta' = \frac{\alpha}{\beta}\), whose asymptotic eigenvalue moments are denoted by \(\{\hat{\lambda}_{\alpha', \beta', \gamma, \delta}\}_{\gamma, \delta \in \mathbb{N}}\). Denote by \(\{\hat{\lambda}_p\}_{p \in \mathbb{N}}\) the eigenvalue moments of \(\frac{1}{M}YSS^HY^H\). Then, we have
\[
\hat{\lambda}_p = \beta' \hat{\lambda}_{\alpha', \beta', \gamma, \delta} = \alpha^p \left(\frac{\beta}{\alpha}\right)^p \hat{\lambda}_{\alpha', \beta', \gamma, \delta} = \alpha^p \alpha^{p-1} \hat{\lambda}_{\alpha', \beta', \gamma, \delta} = \alpha^p f_\alpha * f_\beta * \zeta(0_p, 1_p)
\]
\[ M = \alpha^p \sum_{\pi} \left( \frac{\beta}{\alpha} \right)^{n_1(\pi)} \left( \frac{1}{\alpha} \right)^{n_2(\pi)} 1^{n_3(\pi)} \]
\[ = \frac{1}{\alpha^{p+1}} \sum_{\pi} \beta^{n_1(\pi)} 1^{n_2(\pi)} \alpha^{n_3(\pi)} \]
\[ = \frac{1}{\alpha} \hat{\lambda}_p \] (48)

where \( \pi \) is a positive and nonvanishing partition and \( n_1(\pi), n_2(\pi), \) and \( n_3(\pi) \) are the numbers of blocks of the three types of indices, respectively, satisfying \( n_1(\pi) + n_2(\pi) + n_3(\pi) = 2p + 1 \).

It is easy to check that the asymptotic eigenvalue moments of \( \frac{1}{M} X^H S^H S X \) are given by \( \{ \frac{1}{\alpha} \hat{\lambda}_p \}_{p \in \mathbb{N}} \) since it has the same trace as \( \frac{1}{M} S X X^H S^H \) and dimension \( M \times M \). This concludes the proof. \( \square \)

5) Stieltjes Transform: The Stieltjes transform of a real-valued random variable \( X \) having probability distribution \( P \) is defined as

\[ m_X(z) = \int \frac{1}{x - z} dP(x), \quad z \in \mathbb{C}^+. \] (49)

The following corollary provides an equation characterizing the Stieltjes transform of \( \hat{\lambda} \), which is denoted by \( m_{\hat{\lambda}}(z) \). Note that the equation is also obtained in [26] in a different way, using the non-crossing partition and combinatorial convolution.

**Corollary 6:** When \( \sigma_n^2 = 0 \), the Stieltjes transform of \( \hat{\lambda} \) satisfies

\[ z^2 m_{\hat{\lambda}}^3(z) + (2 - \alpha - \beta)zm_{\hat{\lambda}}^2(z) - (\alpha z - (1 - \beta)(1 - \alpha)m_{\hat{\lambda}}(z) = 0. \] (50)

**Proof:** Similar to (17) and (18), we define the generating functions

\[ \hat{\Lambda}(z) = 1 + \sum_{j=1}^{\infty} \alpha^j \hat{\lambda}_j z^j \] (51)

and

\[ \hat{C}(z) = 1 + \sum_{j=1}^{\infty} c_j z^j \] (52)

which satisfy (due to (19))

\[ \hat{\Lambda}(z) = \hat{C}(z \hat{\Lambda}(z)). \] (53)

Applying (39), we have

\[ \hat{C}(z) = 1 + \sum_{n=1}^{\infty} \sum_{k=1}^{n} \alpha^k \beta^{n-k+1} z^n \binom{n}{k} \binom{n}{k-1} \]
\[ = 1 + \beta \sum_{n=1}^{\infty} \sum_{k=1}^{n} \left( \frac{\beta}{\alpha} \right)^k (\beta z)^n \binom{n}{k} \binom{n}{k-1} \]
\[ = 1 + \beta \sum_{n=1}^{\infty} \lambda_{\beta,n} (\beta z)^n \]
\[ = (1 - \beta) + \beta \Lambda_{\beta}(\beta z) \] (54)

where \( \lambda_{\beta,n} \) and \( \Lambda_{\beta} \) denote the eigenvalue moment \( \lambda_n \) and generating function \( \Lambda \) corresponding to \( S S^H \) when the system load of \( S \) is \( \beta \).

Applying the definition of the Stieltjes transform in (49), we have

\[ m_{\hat{\lambda}}(z) = E \left\{ \frac{1}{\lambda - z} \right\}, \]
\[ = -z^{-1} E \left\{ \frac{1}{1 - \lambda z^{-1}} \right\} \]
\[ = -z^{-1} E \left\{ \frac{1}{1 - \alpha \lambda(z^{-1})} \right\} \]
\[ = -z^{-1} \hat{\Lambda}((\alpha z)^{-1}), \] (55)

Since the free cumulants corresponding to \( \{ \lambda_{\beta,p} \}_{p \in \mathbb{N}} \) are all equal to \( \frac{\beta}{\alpha} \), we have

\[ C_{\beta}(z) = 1 + \sum_{k=1}^{\infty} \frac{\alpha z}{\beta z - 1} \]
\[ = 1 + \frac{\alpha z}{\beta z - 1} \] (56)

where \( C_{\beta}(z) \) corresponds to the free cumulant generating function of \( S S^H \) when the system load of \( S \) is \( \frac{\beta}{\alpha} \) and satisfies

\[ \Lambda_{\beta}(z) = C_{\beta} \left( z \Lambda_{\beta}(z) \right). \] (57)

Combining (53)–(57), we obtain (50) after some algebra. \( \square \)

**Remark 1:** Denoting by \( G(m, z) \) the left-hand side of (50), we can divide it into two parts

\[ G(m, z) = (z^2 m^3 + (2 - \beta)zm^2 + (1 - \beta)m) \]
\[ -\alpha(m^2z + (z + (1 - \beta))m + 1). \] (58)

Letting \( \alpha \to \infty \) in (58), we obtain the equation determining the Stieltjes transform of the eigenvalues in \( S S^H \), which is given by

\[ z m^2(z) + (z + (1 - \beta)m(z) + 1 = 0. \] (59)

It is easy to check that (59) is equivalent to (25).

**Remark 2:** Note that the Stieltjes transform \( m(z) \) is defined on \( \mathbb{C}^+ \). We extend it to the real field and, for \( x \in \mathbb{R}^+ \), define \( m(x) \) as the solution of \( G(m, x) = 0 \), \( \forall x \in \mathbb{R} \). It is easy to check that \( m(z) \) is continuous in \( z \) when \( \text{Re}(z) > 0 \). Thus, the definition of \( m(x) \) for \( x \in \mathbb{R}^+ \) is equivalent to \( m(x) = \lim_{z \to x^+} m(z) \).

**Remark 3:** Note that (50) does not hold when \( x = 0 \) since (55) no longer holds. Actually, when \( x = 0 \), the left-hand side of (50) degenerates to \( \frac{1}{1 - \alpha}(1 - \beta)m(0) = 1 \), which yields a solution \( m_{\beta}(0) = \frac{1}{(1 - \alpha)(1 - \beta)} \). However, since the distribution of \( \hat{\lambda} \) has a mass point at 0, we have

\[ m_\beta(z) = \frac{1 - \min(\beta, \alpha)}{-z} + \int_{x \neq 0} \frac{1}{x - z} dP(x) \]

For the second variable of the function \( G(m, z) \) we use \( z \) to denote complex numbers and \( x \) to denote real numbers.
which increases without bound as \( z \to 0 \), which contradicts the above solution.

Remark 4: When \( z = x \) is on the real line, the cubic (50) has real coefficients. On denoting the discriminant \(^6\) of (50) by \( \Delta(x) \), which is a rational function of \( x \) of finite order, we have
- one real root and a pair of conjugate complex roots, when \( \Delta(x) > 0 \);
- three real roots, two of which are repeated roots, when \( \Delta(x) = 0 \);
- three different real roots, when \( \Delta(x) < 0 \).

C. Asymptotic Eigenvalue Distribution

Note that the rank of \( SXX^H S^H \) is \( \min(K, M) < N \) (recall that we assume that \( \beta < 1 \), \( 0 \) is a mass point of \( \lambda \) with probability \( 1 - \min(\alpha, \beta) \)). Therefore, we consider only the distribution of \( \hat{\lambda} \) when \( \hat{\lambda} > 0 \).

1) Weak Convergence: We have obtained expressions for the asymptotic eigenvalue moments for the matrix \( \hat{R} = \frac{1}{\pi} SXX^H S^H \). However, there may exist more than one probability measure corresponding to these eigenvalue moments. A sufficient condition for the uniqueness of the probability measure corresponding to given moments is given in the following lemma (I, Theorem 30.1)).

Lemma 3: Let \( \mu \) be a probability measure on the real line having bounded moments \( \alpha_k = \int_0^\infty x^k \mu(dx) \) of all orders. If the power series \( \sum_{k=1}^\infty \frac{\alpha_k}{k!} x^k \) has a positive radius of convergence, then \( \mu \) is the only probability measure with the moments \( \{\alpha_k\}_{k \in \mathbb{N}} \).

For random variable \( \hat{\lambda} \), we have the following lemma, which provides an exponential upper bound for \( \hat{\lambda}_p \).

Lemma 4: There exist a constant \( C > 0 \) and \( p_0 \in \mathbb{N} \) such that \( \hat{\lambda}_p < C p, \forall p > p_0 \).

Proof: For simplicity, we assume that \( \alpha = \beta = 1 \) since we can always scale \( C \) to incorporate the values of \( \alpha \) and \( \beta \). Then, the eigenvalue moments are equal to the number of positive and nonvanishing partitions of the corresponding indices.

Since for a positive and nonvanishing partition, the partitions of the sample indices \( \{m_j\}_{j=1, \ldots, p} \) and dimension indices \( \{n_j\}_{j=1, \ldots, p} \) are non-crossing and the partition of the user indices \( \{k_j\}_{j=1, \ldots, 2p} \) are completely determined by those of the sample indices and dimension indices, the number of positive and non-crossing partitions is upper-bounded by \( |\mathbb{N}_C|^2 \).

Due to (6), we have
\[
|\mathbb{N}_C|^2 = \left( \frac{(2)p!}{p(p+1)(p-1)!} \right)^2 \\
\quad \quad = \left( \frac{(2p)!}{(2p)^2} \right)^2 \\
\quad \quad < 4^p, \quad \text{as} \quad p \to \infty
\]

where the second convergence is due to Stirling’s formula. Therefore, the eigenvalue moments (recall that \( \alpha = \beta = 1 \) are upper-bounded by \( 4^p \) for sufficiently large \( p \). This completes the proof.

The following lemma concerning the weak convergence of probability measures ([4, Theorem 30.2]) will be useful in the sequel.

Lemma 5: Suppose that the probability measure \( \mu \) governing the random variable \( X \) is determined by its moments, that random variables \( \{X_n\}_{n \in \mathbb{N}} \) with probability measure \( \\{\mu_n\}_{n \in \mathbb{N}} \) have moments of all orders, and that \( \lim_{n \to \infty} E[X_n^r] = E[X^r] \) for \( r = 1, 2, \ldots \). Then \( \mu_n \) converges weakly to \( \mu \).

Combining Lemmas 3–5, we obtain the following theorem.

Theorem 2: The distribution of \( \hat{\lambda} \) converges weakly to a unique distribution determined by the eigenvalue moments as \( K, N, M \to \infty \).

Remark 5: Note that we have shown only the weak convergence of a generic \( \hat{\lambda} \). A stronger conclusion is shown in [26] that the empirical distribution of \( \hat{\lambda} \) converges vaguely to the distribution determined by the Stieltjes transform. However, this vague convergence of the empirical distribution is beyond the purposes of this discussion.

2) Distribution Law: We analyze the distribution law of the eigenvalue \( \hat{\lambda} \) via the Stieltjes transform determined by (50). On denoting the cumulative distribution function (CDF) and PDF of \( \hat{\lambda} \) by \( \hat{F} \) and \( \hat{f} \), respectively, the inversion formula for the Stieltjes transform is given by
\[
\hat{F}(b) - \hat{F}(a) = \frac{1}{\pi} \lim_{b \to a} \int_a^b \text{Im}(m(x+iy))dx,
\]

\( \forall a < b \in \mathbb{R} \). (60)

The following lemma provides a sufficient condition for the uniform convergence of \( \lim_{m \to 0} \text{Im}(m(z)) = 0 \) in (60), which assures the commutativity of the integral and limit in (60). The proof is given in Appendix I.

Lemma 6: Suppose that a complex-valued function \( m(z) \) is determined by equation \( G(m, z) = 0 \), where \( G(m, z) = \sum_{k=0}^n a_k(z)m^k \) is a monic (namely, \( a_0 = 1 \)) polynomial of variable \( m \), with coefficients determined by \( z \). If there exists an interval \([a, b] \subset \mathbb{R} \), such that the following two conditions hold
\begin{enumerate}
\item \( m(x) \) is not a repeated root of \( G(m, x) = 0 \),
\item coefficients \( \{a_k(z)\}_{k=0, \ldots, n} \) are all analytic functions of \( z \) in a neighborhood of \( x \),
\end{enumerate}
then, \( \forall x \in [a, b] \), there exists a neighborhood \( N \) of \( x \), such that \( m(x+iy) \) converges uniformly on \( N \) to \( m(x) \), as \( y \to 0 \).

Since the discriminant \( \Delta \) is a rational function of \( x \) with finite order, the number of values of \( x \) for which \( m(x) \) is a repeated root of \( G(m, x) = 0 \) is finite, and therefore the set of values of \( x \) such that \( m(x) \) is not a repeated root of \( G(m, x) = 0 \) is dense in \( \mathbb{R}^+ \). By applying Lemma 6 and commuting the integral and limit in (60) for sufficiently small interval \([a, b] \) (the commutativity is

\[\text{Here, uniform convergence means that } \forall \epsilon > 0, \text{ there exists a } \delta > 0 \text{ such that } |m(x+iy) - m(x)| < \epsilon, \forall |y| < \delta \text{ and } x \in [a, b].\]
by the uniform convergence), we obtain that, for any \( x \in \mathbb{R}^+ \) such that \( m(x) \) is not a repeated root and a sufficiently small \( \varepsilon > 0 \)

\[
\hat{F}(x) - \hat{F}(x - \varepsilon) = \frac{1}{\pi} \int_{x-\varepsilon}^{x} \text{Im}(m(x')) dx
\]

(61)

which implies that the CDF \( \hat{F} \) is differentiable at \( x \) and its derivative, namely, the PDF, is given by \( \hat{f}(x) = \frac{1}{\pi} \text{Im}(m(x)) \).

By applying an argument similar to that used in the proof in Appendix I, it is easy to check that the PDF \( \hat{f}(x) \) is continuous for all positive \( x \) not yielding repeated roots of \( G(m, x) = 0 \). Therefore, there are at most finitely many positive mass points of the distribution, which can be located only at points yielding repeated root of \( G(m, x) = 0 \). However, the following lemma excludes the possibility of such positive mass points.

**Lemma 7:** There is no mass point for a positive eigenvalue \( \lambda \).

**Proof:** It is easy to check that \( m(x) \) is bounded for all \( x > 0 \). In (60), on any arbitrary interval \( [a, b] \), \( m(x + iy) \) is also bounded for sufficiently small \( y \), due to the continuity of \( m(z) \). Therefore, the integral converges to 0 as \( a \to b \), which implies that there is no mass point for any positive eigenvalue. \( \square \)

For any real number \( x > 0 \) such that \( m(x) \) is a repeated root of (50), we can define \( \hat{f}(x) \) as an arbitrary positive real number since \( x \) is not a mass point (due to Lemma 7). In particular, for simplicity, we can define \( \hat{f}(x) \) as the limit of \( \hat{f}(x') \) as \( x' \) belongs to a vanishingly small neighborhood of \( x \) containing no other repeated roots of \( G(m, x') = 0 \). This limit is given by

\[
f(x) \triangleq \lim_{x' \to x} \hat{f}(x') = \lim_{x' \to x} \text{Im}(m(x')) = \text{Im}(m(x)) = 0
\]

(62)

where the second equation is because \( G(m, x') \) has no repeated roots, the third equation is due to the continuity of \( \text{Im}(m(x')) \), and the last equation is due to the assumption that \( m(x) \) is a repeated root of \( G(m, x) = 0 \) and, therefore, \( m(x) \) is real and \( \text{Im}(m(x)) = 0 \) since a repeated root of cubic function is always real.

Due to the above argument and lemmas, we obtain the following theorem.

**Theorem 3:** When \( \sigma_n^2 = 0 \), \( \forall x > 0 \), the PDF \( \hat{f}(x) \) of the random variable \( \lambda \) is given by

\[
\hat{f}(x) = \frac{1}{\pi} \text{Im}(m(x)),
\]

(63)

By solving the equation \( G(m, z) = 0 \) and applying Lemma 6, we obtain the following corollary of Theorem 3.

**Corollary 7:** The PDF of \( \hat{\lambda} \) within its support is given by

\[
\hat{f}(x) = \frac{\sqrt{3} \alpha}{2} \left( -\frac{q(x)}{2} + \sqrt{\left(\frac{q(x)}{2}\right)^2 + \left(\frac{p(x)}{3}\right)^3} \right)
\]

\[
+ \frac{\sqrt{3} \alpha}{2} \left( -\frac{q(x)}{2} + \sqrt{\left(\frac{q(x)}{2}\right)^2 + \left(\frac{p(x)}{3}\right)^3} \right)
\]

(64)

where \( p(x) \) and \( q(x) \) are defined in the equations at the bottom of the page.

Note that \( (\frac{q(x)}{2})^2 + (\frac{p(x)}{3})^2 > 0 \) since this quantity equals the discriminant of a cubic equation, which is positive when there are two conjugate complex roots.

3) **Distribution Support:** Note that there is a mass point of the eigenvalue distribution at \( x = 0 \). For simplicity, we consider only the distribution of the positive eigenvalues and define the support of the PDF \( f \) to be the set \( \{ x \mid x > 0, \hat{f}(x) > 0 \} \). The support of the PDF \( f \) of \( \lambda \) is determined by the locations of the roots of \( G(m, z) = 0 \) in the complex plane. One direct way to find this support is to analyze the discriminant \( \Delta(x) \). An alternative way to find this support is to explore the function \( G(m, x) \) with both variables \( m \) and \( x \) in the real field. From (50), \( G(m, x) \) is a cubic polynomial in the variable \( m \), which has at most two stationary points, at which \( \frac{\partial G(m, x)}{\partial m} = 0 \), since its derivative \( \frac{\partial G(m, x)}{\partial m} \) is a quadratic polynomial of \( m \). It is easy to verify that \( G(0, x) < 0, G(\infty, x) = \infty \) and \( G(-\infty, x) = -\infty \). Thus, there exists one positive real root for all cases, and possibly two real negative roots if there are three real roots or two conjugate complex roots if there is only one real root. Therefore, there are three typical types of curves for \( G(m, x) \) when \( m, x \in \mathbb{R} \) (note that \( m \) is a variable and \( x \) is a parameter):

- one real root and no stationary points;
- one real root and two stationary points;
- three real roots and two stationary points.

In the first two cases, there is only one real root of the equation \( G(m, x) = 0 \), and therefore there exists a complex root \( m(x) \) such that \( \text{Im}(m(x)) > 0 \) and \( \hat{f}(x) > 0 \); in the last case, there is no complex root of the equation \( G(m, x) = 0 \), and therefore, for all roots \( m(x) \), \( \text{Im}(m(x)) = 0 \) and \( \hat{f}(x) = 0 \). We summarize the above analysis in the following lemma.

**Lemma 8:** \( x > 0 \) is within the support of \( \hat{f} \) if and only if

- \( \Delta(x) > 0 \); or
- the curve of \( G(m, x) \) (\( m \in \mathbb{R} \)) is of the first two types given above.

\[
p(x) = \frac{(1 - \alpha)^2 + (1 - \beta)^2 - (1 - \alpha)(1 - \beta) + 3\alpha x}{3\alpha^2}
\]

and

\[
q(x) = \frac{(2 - \alpha - \beta)(2(1 - \alpha)^2 + 2(1 - \beta)^2 + 9\alpha x - 5(1 - \alpha)(1 - \beta) - 27\alpha x)}{27\alpha x^3}
\]
Based on Lemma 8, we obtain the following theorem characterizing the support of $\hat{f}(x)$. The proof is given in Appendix II.

**Theorem 4:** The following properties of $\hat{f}(x)$ hold:
1) the support of $\hat{f}(x)$ is given by $(\lambda_{\text{min}}, \lambda_{\text{max}})$, where $\lambda_{\text{min}} \triangleq \inf_{\lambda > 0} (\lambda)$ and $\lambda_{\text{max}} \triangleq \sup_{\lambda > 0} (\lambda)$;
2) $\lambda_{\text{min}} \leq \lambda_{\text{max}} \leq \lambda_{\text{max}}(1 + \min(1/\sqrt{3}, 1/\sqrt{5}))$;
3) for sufficiently large $\alpha$, $(\lambda_{\text{min}}, \lambda_{\text{max}}) \subset (\lambda_{\text{min}}, \lambda_{\text{max}})$; and
4) for sufficiently small $\alpha < \beta$, $\lambda_{\text{min}} \geq \lambda_{\text{max}}$.

**Remark 6:** Conclusion 1) shows that the support of $\hat{f}$ is a continuous interval (note that the mass point $x = 0$ is not included in the definition of support). Conclusion 2) provides lower and upper bounds for $\lambda_{\text{max}}$. Conclusion 3) shows that the range of $\lambda$ is larger than that of $\lambda$ when $\alpha$ is sufficiently large. This conclusion is of particular importance for subspace-based multiuser detection, where the output of the multiuser detector for user $k$ is given by [35]

$$z_k = s_k^H U_s \begin{pmatrix} \frac{1}{\lambda(1)} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda(2)} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda(K)} \end{pmatrix} U_s^H \tau$$  \hspace{1cm} (65)

in which the columns of the $N \times K$ matrix $U_s$ are eigenvectors of the signal subspace, $(\lambda(k))_{k=1,\ldots,K}$ are the $K$ largest eigenvalues, and $\tau$ is the received vector signal. If the eigenvalue estimates are substituted into (65), then the detector will overemphasize the terms corresponding to small eigenvalues when $\alpha$ is sufficiently large since $\lambda_{\text{min}} < \lambda_{\text{max}}$, thus impairing the robustness of detection. Similarly, in water-filling power allocation in MIMO systems, imperfect estimation of the eigenvalues implies that more power than necessary will be allocated to the channel having the largest channel gain. Conclusion 4) shows that it is possible that $\lambda_{\text{min}} \geq \lambda_{\text{max}}$, which means that $(\lambda_{\text{min}}, \lambda_{\text{max}}) \cap (\lambda_{\text{min}}, \lambda_{\text{max}})$ is empty.

V. SPECTRAL ANALYSIS IN THE PRESENCE OF NOISE

In this section, we extend the analysis to the general case of $\sigma_n^2 \geq 0$ based on the conclusions in the previous section. Note that when $\sigma_n^2 > 0$, the exact covariance matrix $R$ is of full rank and there is a mass point at $\lambda = \sigma_n^2$ with probability $1 - \beta$. We first obtain two equivalent procedures generating positive and non-crossing partitions in Lemmas 1 and 11. Based on Lemma 11, we obtain an explicit expression for the eigenvalues in Theorem 5 using combinatorial convolution. Then, we obtain the properties of the eigenvalue moments and eigenvalue distribution as counterparts to the results of Section IV.

A. Non-Crossing Partitions in Eigenvalue Moments

When noise is present in (2), the covariance matrix estimate is given by

$$\hat{R} = \frac{1}{M} \sum_{m=1}^{M} \tau(m) \nu^H(m) = \frac{1}{M} \sum_{m=1}^{M} \left( \sum_{k=1}^{K} s_k(m)x_k(m) + \nu(m) \right)$$

where $\nu(m)$ is the probability of an eigenvalue at point $0$. It is obvious that the rank of $\hat{R}$ is $\min(M, N)$ and the probability of an eigenvalue at point 0 is $1 - \min(\alpha, 1)$.

Then, using the convention that, for a vector $x$, $x^\dagger \triangleq x$ and $x^0 \triangleq 1$ (a scalar), the expectation of the $p$th power of $\hat{R}$ is given by

$$E\{\hat{R}^p\} = \frac{1}{M^p} \sum_{i_1,\ldots,i_{2p} \in [0,1]} \sum_{m_1,\ldots,m_p=1}^{M} \prod_{j=1}^{p} \left( \sum_{k=1}^{K} s_k^H(m_j)x_{km_j} \right)^{i_{2j-1}} \times \left( \sum_{k=1}^{K} s_k^H(m_j)x_{km_j} \right)^{i_{2j}}$$

where the binary indices $\{i_1,\ldots,i_{2p}\}$ range over all possible combinations.

With manipulation similar to that used to obtain (27), we obtain the contribution of terms with an arbitrary fixed $\{i_1,\ldots,i_{2p}\}$ to the $p$th moment $E\{\hat{R}^p\}$, which is given by (note that the superscripts denote exponents instead of indices)

$$\frac{1}{N^{\sum_{j=1}^{2p} i_j} + 1} \times \sum_{m_1,\ldots,m_p=1}^{M} \sum_{j \in \mathcal{J}(1,\ldots,2p)} \sum_{k \in \mathcal{K}(1,\ldots,2p)} \prod_{j=1}^{p} \left( \sum_{k=1}^{K} s_k^H(m_j)x_{km_j} \right)^{i_{2j-1}} \times \left( \sum_{k=1}^{K} s_k^H(m_j)x_{km_j} \right)^{i_{2j}}$$

where $J \triangleq \{j \mid i_j = 1, 1 \leq j \leq 2p\}$ (we denote the complement set of $J$, $\{1,\ldots,2p\} / J$, by $\mathcal{J}$) and $u_{mn}$ is the $n$th element of $\nu(m)$.

For applying the conclusions of systems without noise in Section IV, we extend the definition of complementation map (we call this extension the extended complementation map) when the expectation of (67) is positive. The points and lines in the extended complementation map are defined in the following way.

- For indices $\{m_j\}_{j=1,\ldots,p}$, $\{n_j\}_{j=1,\ldots,p}$, and $\{k_j\}_{j \in \mathcal{J}}$, the definitions of points and lines are the same as in the complementation map.
• For user indices \( \{ k_j \}_{j \in \mathcal{J}} \), a line between two points \( k_{2s} \) and \( k_{2s-1} \) (suppose that \( i_{2s} = 0 \) and \( i_{2s-1} = 0 \)) means that \( u_{n_{t(i) + s + 1}} = u_{n_{t(i) + s}} \). We use thick solid lines to distinguish the user indices associated with noise terms (the corresponding \( i_j = 0 \)).

In order for the expectation in (67) to be positive, each block of \( \{ k_j \}_{j \in \mathcal{J}} \) must have an even number of elements (otherwise the expectation is 0). In the extended complementation map, we remove some thick lines if necessary such that each \( k_{2s} \) with \( 2s \in \mathcal{J} \) is connected to one and only one \( k_{2s-1} \) with \( 2s-1 \in \mathcal{J} \), which results in a total of \( \lfloor \frac{|\mathcal{J}|}{2} \rfloor \) thick solid lines in the extended complementation map. Then, (67) can be rewritten as

\[
\mathbb{E}[\mathcal{J}] N^{p+1} M^p \times \sum_{m_1, \ldots, m_p = 1}^{M} \sum_{k_1, \ldots, k_{2p} = 1}^{K} \sum_{n_1, \ldots, n_{2p} = 1}^{N} \prod_{p=1}^{2p-1} \prod_{i=1}^{2p-1} \left( x_{k_{2p+1}}^{i_{k_{2p+1}} n_{1}} \right)^{i_{k_{2p+1}}} \left( x_{k_{2p+1}}^{i_{k_{2p+1}} n_{2}} \right)^{i_{k_{2p+1}}}
\]

It should be noted that there may be multiple choices of the \( \lfloor \frac{|\mathcal{J}|}{2} \rfloor \) thick lines if four or more user indices associated with noise terms share the same subscripts and the extended complementation map is therefore not unique. An example is given in Fig. 4, in which \( \mathcal{J} = \{ 2, 3, 6, 7 \} \), the partition of \( \{ m_j \}_{j=1, \ldots, p} \) is \( \{ 1, 2, 3, 4, 5, 6 \}, \{ 7, 8 \} \), and the partition of \( \{ n_j \}_{j=1, \ldots, p} \) is the same as in Fig. 3. It is easy to check that \( u_{m_1 n_2} = u_{m_2 n_2} = u_{m_3 n_2} = u_{m_3 n_3} \). Fig. 4 shows two possible extended complementation maps representing the same partition. However, as will be seen below, the choice is unique for positive and non-vanishing partitions.

We now have the following lemma which discloses the necessity of non-crossing structure for positive and nonvanishing partitions.

**Lemma 9:** For a fixed \( (i_1, \ldots, i_{2p}) \), if the partition of the indices \( \{ m_j \}_{j=1, \ldots, p} \), \( \{ n_j \}_{j=1, \ldots, p} \), and \( \{ k_j \}_{j=1, \ldots, 2p} \) in (68) results in a positive and nonvanishing expectation, then different types of lines do not intersect in the corresponding extended complementation map.

**Proof:** Suppose that there exist crossing lines of different types in the extended complementation map corresponding to the partition of \( \{ m_j \}_{j=1, \ldots, p} \), \( \{ n_j \}_{j=1, \ldots, p} \), and \( \{ k_j \}_{j=1, \ldots, 2p} \) and the expectation is positive and nonvanishing. The nonvanishing property requires \( k + m + n = 2p + 1 \). The positivity requires that for all \( k_{2s} \sim k_{2s-1} \) \((s,t \in \mathcal{J})\), \( m_s \sim m_t \) and \( n_{s+1} \sim n_t \). It is easy to check that these two properties also assure that the term in (27) corresponding to the complementation map having the same lines (the thick solid lines are replaced with dotted lines) and points as the extended complementation map, results in a positive and nonvanishing expectation in noise-free systems. This contradicts the fact that the non-crossing property of the complementation map assures positive and nonvanishing expectations in noise-free systems.

Based on Lemma 9, we can show the following result, which is key to the structure of positive and nonvanishing partitions when \( \sigma_n^2 > 0 \).

**Lemma 10:** For a fixed \( (i_1, \ldots, i_{2p}) \), if the partition of indices \( \{ m_j \}_{j=1, \ldots, p} \), \( \{ n_j \}_{j=1, \ldots, p} \), and \( \{ k_j \}_{j \in \mathcal{J}} \) in (68) results in a positive and nonvanishing expectation, then, for any \( 2s \in \mathcal{J} \), there exists one and only one \( 2t - 1 \in \mathcal{J} \) such that \( k_{2s} \sim k_{2t-1} \) with a line.
$k_{2l-1}$, namely, $u_{m_en_{e+1}} = u_{m_t n_t}$ (equivalently, $m_e \sim m_t$ and $n_{e+1} \sim n_t$).

Proof: The existence of $2l-1$ is due to the positivity of the expectation. So, we need to show only the uniqueness of $2l-1$.

Suppose that there exist distinct $s, t, g, h \in \mathcal{J}$ such that

$$k_{2s} \sim k_{2l-1} \sim k_{2g} \sim k_{2l-1}$$

or equivalently

$$u_{m_en_{e+1}} = u_{m_t n_t} = u_{m_g n_{g+1}} = u_{m_h n_h}$$

which implies that

$$\begin{cases} m_e \sim m_t \sim m_g \sim m_h \\ n_{e+1} \sim n_t \sim n_{g+1} \sim n_h \end{cases}$$

The two possible relative locations of the points \( \{m_e, m_t, m_g, m_h\} \) and \( \{n_{e+1}, n_t, n_{g+1}, n_h\} \) in the extended complementation map are shown in Fig. 5. In both cases, the lines representing equivalence of different types of indices intersect with each other, which contradicts Lemma 9. □

Lemma 10 excludes the possibility that more than two noise terms in (68) are identical in positive and nonvanishing partitions. Therefore, the corresponding extended complementation map is unique and we can obtain the following corollary, which implies that different noise distributions having the same variance cause the same perturbation on the eigenvalues in the large system limit.

Corollary 8: The asymptotic eigenvalue moments are dependent on only the variance of the noise, and not on its distribution.

Based on Lemma 10, the following proposition provides a procedure for constructing positive and nonvanishing partitions.

Proposition 1: All positive and nonvanishing partitions for $\lambda_p$ can be constructed using the following steps:

1) Choose an even number $0 \leq \ell' \leq p'$; draw $\ell'$ non-crossing and nonadjacent thick solid lines (nonadjacent means that no points are shared by two thick solid lines as their ends), each of which connects points $k_{2s}$ and $k_{2l-1}$ for some $s$ and $t$.

2) The $\ell'$ lines partition the points in the extended complementation map into $\ell' + 1$ separated point sets (the user indices in $\mathcal{J}$ are excluded). If $k_{2s}$ and $k_{2l-1}$ are connected by a thick solid line, then connect $m_e$ and $m_t$ with a solid line, consider them as one point in the separated point set and do the same for $n_{e+1}$ and $n_t$.

3) For each of the $\ell' + 1$ separated point sets, if there exists at least one user index $k_{ij}$, construct a non-crossing partition in the same way as for the noise-free case in Section IV; if there is only one sample index or dimension index in this separated point set, simply consider it as a single-element block.

Proof: First, we need to show that the procedure in the proposition results in a positive and nonvanishing partition. The expectation equals the product of the noise terms and the expectations in the separated point sets. Obviously, the procedure yields a positive expectation. Thus, we need to show only that this partition is nonvanishing. For all $j$ such that $1 \leq j \leq \frac{p}{2} + 1$ (note that $|\mathcal{J}| = 2p'$ and the number of separated point sets is $|\mathcal{J}| + 1$), we denote by $a_j$ the number of sample indices $k_i$ in the $j$th separated point set (note that the number of dimension indices is the same), which satisfies

$$\sum_{j=1}^{2l+1} 2a_j = 2p - |\mathcal{J}|,$$

(69)

Then, for $1 \leq j \leq \frac{|\mathcal{J}|}{2} + 1$, the contribution of the non-crossing partition within the $j$th separated point set to the product is of order $N^{a_j+1}$, according to the analysis in Section IV. Therefore, the total product is of order $N\sum_{j=1}^{2l+1}(2a_j+1)$, where

$$\sum_{j=1}^{2l+1} (2a_j + 1) = 2p - |\mathcal{J}| + \frac{|\mathcal{J}|}{2} + 1 = 2p + 1 - \frac{|\mathcal{J}|}{2}$$

which results in a nonvanishing expectation according to (68).

Now, we need to show that all positive and nonvanishing partitions can be generated from the three steps. First, Lemmas 9 and 10 justify the $p'$ non-crossing and nonadjacent thick solid lines in step 1). Once the thick solid lines are determined, the remaining points can be partitioned only within the corresponding separated point set due to the non-crossing requirement in Lemma 9. Thus, if the partition within a separated point set having $2a_j$ user indices is crossing, then the corresponding contribution to the product is of order less than $N^{2a_j+1}$, which makes the expectation vanishing as $K, M, N \to \infty$ due to (69). Therefore, the non-crossing partition within each separated point set is necessary. This completes the proof. □

The following example illustrates the steps in Proposition 1. When $p = 8$ and $p' = 2$, we draw two thick solid lines between $k_1$ and $k_{32}$ and between $k_2$ and $k_{34}$ in step 1 (therefore, $\mathcal{J} = \{1, 2, 7, 12\}$). Then, due to step 2), we merge points $m_4$ and $m_5$, $m_5$ and $m_6$, $n_4$ and $n_7$, and $n_7$ and $n_2$. The remaining points in the circle graph are divided into three separated point sets, which are

$$\begin{align*}
\{m_7, m_8, n_7, n_8, k_{13}, k_{14}, k_{15}, k_{16}\}, \\
\{m_1, m_5, n_5, n_6, k_8, k_9, k_{10}, k_{11}\}, \\
\{m_2, m_3, n_2, n_3, k_3, k_4, k_5, k_6\}.
\end{align*}$$

Then, $a_1 = a_2 = a_3 = 2$ and the non-crossing partitions within each separated point set results in a contribution of order $N^{15}$, which cancels the scalar $\frac{1}{N^{15}}$ in (68). The corresponding extended complementation map is shown in Fig. 6. The procedure for generating positive and nonvanishing partitions in Proposition 1 results in an expression for the asymptotic eigenvalue moments given in the following corollary.

Corollary 9: The $p^\text{th}$ asymptotic eigenvalue moment of $\bar{R}$ is given by

$$\lambda_p = \sum_{i=0}^{p} \left(\frac{\sigma_n^2}{p}\right)^{r} \sum_{k=0}^{v-1} E_{r^a_j=0} \left\{\lambda_j^{x_j}, f_i\right\} \text{ (70)}$$

where $L_r$ denotes all possible selections of $r$ non-crossing and
nonadjacent thick solid lines in step 1) in Proposition 1, \( a_j(I) \) denotes the number of sample (or dimension) indices in the \( j \)th separated point set and \( E_{\sigma_j^2 = 0} \{ \hat{\lambda}_{j,I} \} \) is the asymptotic eigenvalue moment without noise, which is given in (40).
However, Corollary 9 does not provide an explicit expression for the eigenvalue moments since it is difficult to obtain explicit expressions for the cardinality of set $L_r$ and coefficients $\{a_j(D)\}_{j \in L_r}$. Therefore, in the following lemma, we provide an alternative procedure for generating positive and nonvanishing partitions by delaying the step of choosing thick solid lines after partitioning all other indices, based on which we can derive an explicit expression for the asymptotic eigenvalue moments. The steps are summarized in the following lemma.

**Lemma 11:** All positive and nonvanishing partitions can be constructed in the following steps.

1) Choose arbitrary $1 \leq m, n \leq p$ and $k = 2p + 1 - m - n$ and partition the indices $\{m_j\}_{j=1,\ldots,p}$, $\{n_j\}_{j=1,\ldots,p}$, and $\{k_j\}_{j=1,\ldots,2p}$, to form a complementation map in the same way as in the proof of Theorem 1.

2) For every pair of connected user indices $k_{2r}$ and $k_{2r-1}$ such that $m_{k_{2r+1}} \sim m_k$ and $n_{k_{2r-1}} \sim n_k$, label the corresponding line as either thick solid or dotted.

Take the partition in Fig. 3 for example. In step 1), we obtain a partition of indices $\{m_j\}_{j=1,\ldots,p}$, $\{n_j\}_{j=1,\ldots,p}$, and $\{k_j\}_{j=1,\ldots,2p}$ in Fig. 3. Then, we choose the lines between $k_1$ and $k_{12}$ and between $k_2$ and $k_7$ as thick solid lines, thus resulting in the extended complementation map shown in Fig. 6.

Finally, we obtain the following theorem providing an explicit expression for the asymptotic eigenvalue moments via combinatorial convolution, which is similar to that of the noise free case in Theorem 1.

**Theorem 5:** Denoting by $f_{\lambda}$ the multiplicative function associated with series $\{\alpha^p\lambda^p\}_{p \in \mathbb{N}}$, we have

$$ f_{\lambda} = f_\alpha \ast f_\beta \ast \sigma_n^2 \ast \zeta $$

(71)

where $f_\alpha$ is associated with $(\alpha, \alpha, \ldots)$, $f_\beta \sigma_n^2$ is associated with $(\beta + \sigma_n^2, \beta, \beta, \ldots)$, and $\zeta$ is the zeta function.

**Proof:** Due to Lemma 11, the extended complementation map of systems with noise is similar to the complementation map of systems without noise except that the connected user indices $k_{2r}$ and $k_{2r-1}$ satisfying that $m_{k_{2r+1}} \sim m_k$ and $n_{k_{2r-1}} \sim n_k$ can be selected as noise terms. Recall that such two user indices are merged into one user index in the proof of Theorem 1. Therefore, the corresponding single-element block in the non-crossing partitions of user indices $\{k_j\}_{j=1,\ldots,2p}$ contributes $\beta + \sigma_n^2$ to the expectation (contributes $\sigma_n^2$ when it is chosen as a noise term and $\beta$ otherwise), thus changing the first coefficient $\beta$ of $f_\beta$ in Theorem 1 to $\beta + \sigma_n^2$. \qed

**Remark 7:** An interesting observation on Theorem 5 is that the expression for the asymptotic eigenvalue moments with noise is very similar to that without noise. The noise changes only the first coefficient of $f_\beta$ in Theorem 1. Essentially, this is determined by Lemma 10.

Based on Theorem 5, we obtain the following corollary as an extension of Corollary 5.

**Corollary 10:** The distribution of eigenvalues of the matrix $\frac{1}{M} (S^T X + N^T Y)$ is the same as that of the matrix $\frac{1}{M} (S^T S^H + \sigma_n^2 I_{N \times N}) Y^H$, as $K, M, N \to \infty$, where $Y$ is an $M \times N$ matrix, whose elements are mutually independent random variables having unit variance.

**Proof:** Due to additivity of free cumulants [29], the free cumulant of the matrix $A + B$ is equal to the sum of those of matrices $A$ and $B$ if $A$ and $B$ are mutually free. Since the free cumulants of $SS^H$ are $\{\beta, \beta, \beta, \ldots\}$ and those of $\sigma_n^2 I_{N \times N}$ are $\{\sigma_n^2, 0, 0, \ldots\}$ (they are mutually free according to [34, Theorem 2.2]), the free cumulants of $SS^H + \sigma_n^2 I_{N \times N}$ are given by $\{\beta + \sigma_n^2, \beta, \beta, \ldots\}$. By applying the combinatorial convolution in [29] and Theorem 5, we obtain that the traces of powers of $\frac{1}{M} Y (SS^H + \sigma_n^2 I_{N \times N}) Y^H$ are the same as those of $\frac{1}{M} (S^T X + \bar{N}) (S^T Y + \bar{N})^H$. Notice that the dimension of $\frac{1}{M} Y (SS^H + \sigma_n^2 I_{N \times N}) Y^H$ is the same as that of $\frac{1}{M} (S^T X + \bar{N}) (S^T Y + \bar{N})$ (both are $N \times N$ matrices). Thus, their eigenvalue moments are identical since they have the same traces. \qed

### B. Asymptotic Eigenvalue Moments

In this subsection, we obtain the noisy counterparts of the conclusions on asymptotic eigenvalue moments in noise-free systems using approaches similar to those used in Section IV.

1) **Free Cumulants:** The following corollary gives an explicit expression for the free cumulants, which will be used when deriving the Stieltjes transform.

**Corollary 11:** The free cumulants of the random variable $\alpha \lambda$ are given by

$$ c_n = \sum_{j_1, \ldots, j_n = 0}^{n!} \frac{n! (\beta + \sigma_n^2)^j \sum_{k=1}^n j_k \lambda^m \sum_{h=1}^n j_h + 1)}{(\prod_{k=1}^n j_k!)} (n + 1 - \sum_{k=1}^n j_k) \lambda $$

(72)

where the indices $j_1, \ldots, j_n$ are indistinguishable. When $\alpha = 1$, (72) simplifies to

$$ c_n = \lambda $$

(73)
where $\lambda_n$ is the $n^{th}$ eigenvalue moment of the exact covariance matrix $R = SS^H + \sigma_n^2 I$.

Proof: Comparing (15) and (71), we obtain that

$$\hat{c}_n = f_{\alpha} * f_{\beta, \sigma_n^2} (0_n, 1_n)$$

$$= \sum_{\pi \in \mathcal{N}_C} f_{\beta, \sigma_n^2} (0_n, \pi) f_{\alpha} (\pi, 1_n)$$

$$= \sum_{\pi \in \mathcal{N}_C} \left[ \sum_{i=1}^{n} \sum_{k=1}^{n} f_{\beta, \sigma_n^2} (0_n, \pi) f_{\alpha} (0_n, \pi^*) \right]$$

where $j_k$ denotes the number of blocks having $k$ elements, the second equation is due to the commutativity of combinatorial convolution, and the third equation is due to the isomorphic decomposition in (10). Note that the indices $j_1, \ldots, j_n$ are indistinguishable because the summation is determined by only the cardinalities of blocks. For fixed $\{j_k\}_{k=1}^n$, the total number of blocks in $\pi$ is $\sum_{k=1}^{n} j_k$; therefore, we have

$$f_{\alpha} (0_n, \pi^*) = \alpha^{n - \sum_{k=1}^{n} j_k + 1}$$

since there are $n - \sum_{k=1}^{n} j_k + 1$ blocks in $\pi^*$ and each block contributes $\alpha$.

Since $f_{\beta, \sigma_n^2}$ is associated with $(\beta + \sigma_n^2, \beta, \ldots, \beta)$, we have

$$f_{\beta, \sigma_n^2} (0_n, \pi) = (\beta + \sigma_n^2)^{\sum_{j_k=1}^{n} j_k} \beta^{n - \sum_{k=1}^{n} j_k}.$$

Due to Corollary 1 in [28], the number of non-crossing partitions $[0, \pi] \cong \prod_{k=1}^{n} \mathcal{N}_C_{j_k}$ is given by

$$\frac{n!}{(\prod_{k=1}^{n} j_k)! (n + 1 - \sum_{k=1}^{n} j_k)!)}.$$

Equation (72) is obtained by combining the above equations. Equation (73) holds when $\alpha = 1$ because

$$\hat{c}_n = f_{\alpha} \ast f_{\beta, \sigma_n^2} (0_n, 1_n)$$

$$= \zeta \ast f_{\beta, \sigma_n^2} (0_n, 1_n)$$

$$= \lambda_n$$

where the second equation is because $f_{\alpha} = \zeta$ when $\alpha = 1$. □

2) An Explicit Expression for the Eigenvalue Moments: The following corollary gives an explicit expression for the asymptotic eigenvalue moments.

Corollary 12: See (74) at the bottom of the page, where $\lambda_{i_j}$ is the $i_j$th eigenvalue moment of $SS^H + \sigma_n^2 I$ and $M(i_1, \ldots, i_k)$ is defined in (8). Note that the indices $i_1, \ldots, i_k$ are indistinguishable.

Proof: We denote by $f_{\lambda}$ the multiplicative function associated with $\{\lambda_p\}_{p \in \mathbb{N}}$. Then, by applying Theorem 5, we have

$$\alpha^p \lambda_p = \sum_{\pi \in \mathcal{N}_C} f_{\lambda} (0_p, \pi) f_{\alpha} (\pi, 1_p)$$

$$= \sum_{\pi \in \mathcal{N}_C} f_{\lambda} (0_p, \pi) \alpha^{p - \sum_{j=1}^{p} i_j}$$

$$= \sum_{p} \sum_{\pi \in \mathcal{N}_C} \sum_{i_1, \ldots, i_k} f_{\lambda} (0_p, \pi) \alpha^{p - \sum_{j=1}^{p} i_j}$$

$$= \frac{1}{\alpha} \left( \frac{1}{\alpha} \right)^{p - k + 1} \sum_{\pi \in \mathcal{N}_C} \left( \alpha \sum_{i_1, \ldots, i_k} \lambda_{i_1, \ldots, i_k} \right)$$

$$= \frac{1}{\alpha} \left( \frac{1}{\alpha} \right)^{p - k + 1} \sum_{i_1, \ldots, i_k} \lambda_{i_1, \ldots, i_k} \prod_{j=1}^{k} \lambda_{i_j}$$

(75)

where the last equation is due to the fact that $f_{\lambda} (0_n, 1_n) = \lambda_n$. Again, the indices $i_1, \ldots, i_k$ are indistinguishable because the summation is determined by only the cardinalities of blocks.

Thus, (74) is obtained by substituting (8) into (75). □

From Corollary 12, we obtain the first four moments of the eigenvalues, which are given by

$$E\{\lambda\} = \sigma_n^2 + \beta$$

$$E\{\lambda^2\} = \left( \frac{1}{\alpha} + 1 \right) \left( \sigma_n^2 + \beta \right)^2 + \beta$$

$$E\{\lambda^3\} = \left( \frac{1}{\alpha^3} + \frac{3}{\alpha} + 1 \right) \left( \sigma_n^2 + \beta \right)^3 + 3 \left( \frac{1}{\alpha} + 1 \right) \beta \left( \sigma_n^2 + \beta \right) + \beta$$

$$E\{\lambda^4\} = \left( \frac{1}{\alpha^5} + \frac{6}{\alpha^2} + \frac{6}{\alpha} + 1 \right) \left( \sigma_n^2 + \beta \right)^4 + \left( \frac{6}{\alpha^2} + \frac{16}{\alpha} + 6 \right) \beta \left( \sigma_n^2 + \beta \right)^2 + \frac{1}{\alpha} \left( 4 \beta \sigma_n^2 + 6 \beta^2 \right) + 6 \beta^2 + 4 \beta \sigma_n^2 + \beta$$

(76)  (77)  (78)  (79)

Using an argument similar to that for the noise-free case, the conclusion that the asymptotic eigenvalue moments of the estimated covariance matrix are larger than those of the exact covariance matrix (except for the expectation) in Corollary 3 also holds in systems with noise.

3) Stieltjes Transform: The Stieltjes transform of the eigenvalue $\lambda$, denoted by $m_{\lambda} (z)$, is given in the following corollary.
Corollary 13: The Stieltjes transform \( m(z) \) satisfies
\[
\sigma_n z m(z) + \left( \alpha z^2 + 2(1 - \alpha)\sigma_n^2 z^2 \right) m''(z) \\
+ \left( (1 - \alpha)^2 \sigma_n^2 + \alpha (2 - \alpha - \beta - \sigma_n^2) z^2 \right) m'(z) \\
- \alpha \left( \alpha z - (1 - \alpha) (1 - \beta - \sigma_n^2) \right) m(z) - \alpha^2 = 0,
\]
which is the same as (59) except that \( z \) is replaced with \( z - \sigma_n^2 \). This then characterizes the Stieltjes transform of \( SS^H + \sigma_n^2 I_{N \times N} \).

C. Asymptotic Eigenvalue Distribution

The same argument as in Theorem 2 can be applied to show the weak convergence and uniqueness of the probability measure corresponding to the eigenvalue moments. By applying Lemma 6 and the same argument as in Section IV, the PDF \( f \) of the eigenvalue distribution equals the imaginary part of the Stieltjes transform, which is obtained by solving the quartic equation (80). Since the product of the roots of (80) equals \( -\alpha^2 \) and is negative, (80) has at most two conjugate complex roots, and therefore at most one complex root having positive imaginary part, which excludes the possibility that we obtain two roots having positive imaginary parts (from (80)).

Based on the Stieltjes transform, the properties of the eigenvalue distribution are summarized in the following theorem, which is the counterpart of Theorem 4. The proof is given in Appendix III. We define \( \lambda_{\text{max}} \triangleq \sup_{\lambda > 0} \frac{1}{\alpha} \hat{\lambda} \), \( \hat{\lambda}_{\text{min}} \triangleq \inf_{\lambda > 0} \frac{1}{\alpha} \hat{\lambda} \), \( \hat{\lambda}_{\text{max}} \triangleq \inf_{\lambda > 0} \frac{1}{\alpha} \hat{\lambda} \), and \( \hat{\lambda}_{\text{min}} \) are given by \( (1 + \sqrt{\beta})^2 + \sigma_n^2, \sigma_n^2, \alpha \), and \( (1 - \sqrt{\beta})^2 + \sigma_n^2 \), respectively.

Theorem 6: There is no mass point for any positive eigenvalue \( \hat{\lambda} \). The support of \( f \) satisfies the following properties:

1) for sufficiently large \( \alpha \), the support of \( f \) is not a continuous interval when \( \sigma_n^2 > 0 \);
2) \( \lambda_{\text{max}} \leq \hat{\lambda}_{\text{max}} \leq \lambda_{\text{max}} (1 + \min(\sqrt{\frac{1}{\alpha}}, \sqrt{\alpha}))^2 \);
3) for sufficiently large \( \alpha \), \( [\hat{\lambda}_{\text{min}}, \hat{\lambda}_{\text{max}}] \subset [\lambda_{\text{min}}, \lambda_{\text{max}}] \); and
4) for sufficiently small \( \alpha < \beta, \hat{\lambda}_{\text{min}} \geq \lambda_{\text{max}} \).

Remark 10: Notice that the properties 3) and 4) are the same as in Theorem 4. However, property 1) is completely different. The essential reason is due to the existence of a mass point at \( \sigma_n^2 \). When \( \sigma_n^2 \) is zero and \( \alpha > 0 \), the mass point at 0 always exists with probability 1 - \( \beta \) and the support on positive eigenvalues is continuous. When \( \sigma_n^2 > 0 \) and \( 1 < \alpha < \infty \), the estimated covariance matrix is of full rank and there is no mass point. As \( \alpha \to \infty \), the support of positive eigenvalues has to be separated.
into at least two disjoint intervals such that the support around \( \sigma_n^2 \) shrinks to a point.

VI. LARGE SAMPLE ANALYSIS

In this section, we analyze the moments and distribution of eigenvalues in both large sample and large system limits, which means \( \alpha \to \infty \) while \( M \to \infty \), or intuitively, that estimates of moments and distribution are infinitesimally close to those of exact covariance matrix. Note that we consider the general case of \( \sigma_n^2 \geq 0 \).

A. Perturbation of Moments

The following proposition provides the first-order perturbation of eigenvalue moments. The proof is given in Appendix IV.

**Proposition 2:** When \( \alpha \) is sufficiently large, the \( p \)th-order eigenvalue moment is given by

\[
\hat{\lambda}_p = \lambda_p + \frac{1}{\alpha} \sum_{i_1 + i_2 = \nu; i_1 i_2 \geq 1} g(i_1, i_2) \lambda_{i_1} \lambda_{i_2} + o \left( \frac{1}{\alpha} \right) \tag{85}
\]

where

\[
g(i_1, i_2) = \begin{cases} \frac{\beta}{\nu}, & \text{if } i_1 = i_2 = \frac{\beta}{2} \\ \frac{1}{\nu}, & \text{otherwise}. \end{cases} \tag{86}
\]

Note that the indices \( i_1 \) and \( i_2 \) are indistinguishable.

B. Perturbation of Distribution

The following proposition provides the first-order perturbation of the eigenvalue distribution. The proof is given in Appendix V.

**Proposition 3:** For sufficiently large \( \alpha \), the PDF of \( \hat{\lambda} \) is given by

\[
f^{\dagger}(x) = \left( x - \frac{1}{\alpha} \frac{m g(x)}{(x^2 - \sigma_n^2) m^2 - 1} \right) + o \left( \frac{1}{\alpha} \right), \quad \forall x \neq \frac{\sigma_n^2}{m} \tag{87}
\]

where \( m \) is the solution of (80) and

\[
g(x) = \left( x^2 - 2 \sigma_n^2 x \right) m^3 + \left( \frac{2 - \beta - \sigma_n^2}{\sigma_n^2 m} \right)x - \frac{2 \sigma_n^2}{m^2} \] + \left( 1 - \beta - \sigma_n^2 \right) m. \]

We call the coefficient \( \text{Im} \left( \frac{mg(x)}{(x^2 - \sigma_n^2) m^2 - 1} \right) \) in (87) the *perturbation factor* since it determines the first-order perturbation on the eigenvalue distribution in the large sample limit.

VII. NUMERICAL RESULTS

We now illustrate the analytical results in this paper via simulation. For simplicity, we assume that the elements in both \( \mathbf{S} \) and \( \mathbf{X} \) take values in \{ -1, +1 \} equiprobably (this corresponds to binary spreading codes and binary phase-shift keying (BPSK))
modulation in CDMA systems) and the elements in $n$ are complex Gaussian distributed. All simulation results are based on 1000 realizations of random matrices.

A. Eigenvalue Moments

Fig. 7 shows a comparison between numerical simulation results and the explicit expressions in Corollary 12 for the asymptotic eigenvalue moments with various values of $K$ and the fixed configuration $\beta = 0.5$, $\alpha = 1$, and $\sigma_n^2 = 0.1$. The order of moments is denoted by $p$, ranging from 1 to 4. We observe that when $K$ is sufficiently large, the numerical results are close to those obtained from the asymptotic expressions. Figs. 8 and 9 show the same comparison except that $\sigma_n^2 = 0.5$ in Fig. 8 and $\alpha = 2$ in Fig. 9. We observe that the convergence rate increases in $\sigma_n^2$ and decreases in $\alpha$.

Fig. 10 shows the comparison of eigenvalue moments between the matrix $\frac{1}{N}(SX + N)^H(SX + N)$ (labeled as “matrix1”) and $\frac{1}{N}Y(\Sigma^H + \sigma_n^2 I_{N\times N})Y^H$ (labeled as “matrix2”), where the random matrix $Y$ is defined in Corollary 10, when $\alpha = 2$, $\beta = 0.5$, and $\sigma_n^2 = 0.1$. We observe that when $K$ is larger than 30, the eigenvalue moments of the two matrices are very close to each other, which demonstrates the conclusion in Corollary 10.

B. Eigenvalue Distribution

Fig. 11 shows a comparison of the eigenvalue distribution for finite systems ($K = 2,20$) and for large systems ($K \to \infty$), with the configuration $\alpha = 2$, $\beta = 0.5$, and $\sigma_n^2 = 0.2$. Note that the distribution is obtained over finitely many intervals and the distribution curves are normalized such that the corresponding sum of probability equals 1. We observe that, when $K$ is very small, there is a gap between the results for the finite system and the asymptotic results. When $K$ is sufficiently large, the curve coincides with the asymptotic results when $\lambda$ is large and there is still a considerable gap around the peak $\lambda = \sigma_n^2$. This implies that the convergence rate (with respect to $K$) around $\lambda = \sigma_n^2$ is smaller than for larger $\lambda$.

Fig. 12 shows the distribution of eigenvalues with the different values of $\alpha (\alpha = 2,20,\infty)$ and with the configuration $\beta = 0.5$ and $\sigma_n^2 = 0.2$. Note that the curve is normalized in the same way as Fig. 11. We observe that the gap between the curves for the estimated eigenvalue distribution ($\alpha = 2,20$) and that of exact eigenvalue ($\alpha = \infty$) is small for large $\lambda$ and becomes large around $\lambda = \sigma_n^2$.

In both simulation results, we see that the main error in the eigenvalue distribution lies in the region of small $\lambda$, particularly around $\lambda = \sigma_n^2$. The convergence rate of the distribution of large eigenvalues is substantially fast in both system size and sample size.

C. Perturbation Factor

Fig. 13 shows the perturbation factor $\text{Im} \left( \frac{m g(m)}{(m - \sigma_n^2)(m^2 - 1)} \right)$ in (87) with $\alpha = 2,5,10,20$, $\beta = 0.5$, and $\sigma_n^2 = 0.2$. We observe that the perturbation factor around $\lambda = \sigma_n^2$ is substantially large, which coincides with the observations in Figs. 12 and 13. Another interesting observation is that the perturbation factor around $\lambda = \sigma_n^2$ is not monotonic in $\alpha$. 

Fig. 8. Comparison between simulation results and the asymptotic expression for the eigenvalue moments when $\beta = 0.5$, $\alpha = 1$ and $\sigma_n^2 = 0.5$. 

Authorized licensed use limited to: UNIVERSITY OF TENNESSEE. Downloaded on February 25, 2009 at 08:45 from IEEE Xplore. Restrictions apply.
Fig. 9. Comparison between simulation results and the asymptotic expression for the eigenvalue moments when $\beta = 0.5$, $\alpha = 2$ and $\sigma_n^2 = 0.1$.

Fig. 10. Comparison between the eigenvalue moments of $\frac{1}{M}Y(\Sigma \Sigma^H + \sigma_n^2 I_{N \times N})Y^H$ and $\frac{1}{M}(\Sigma X + N)^H(\Sigma X + N)$. 
Fig. 11. Comparison of the eigenvalue distribution of a finite system and the corresponding asymptotic results.

Fig. 12. Eigenvalue distribution for different values of $\alpha$.
VIII. CONCLUSION

In this paper, we have carried out a large-system-limit based spectral analysis of covariance matrix estimation. By applying the theory of non-crossing partitions, we have obtained explicit expressions for the asymptotic eigenvalue moments of the covariance matrix estimate. Based on the eigenvalue moments and Stieltjes transform, we have derived equations determining the asymptotic eigenvalue distribution and extended the results to the large sample case. The corresponding impact of covariance matrix estimation error has been discussed for the application of subspace-based multiuser detection. Numerical results show that the asymptotic results can be applied in finite systems as good approximations and that the main difficulty of the eigen-decomposition lies at the lower end of spectrum.

APPENDIX I
PROOF OF LEMMA 6

Proof: Fix an $x_0 \in [a, b]$. Then we have

$$\frac{\partial G}{\partial m}
\bigg|_{m=m(x_0)} \neq 0$$

since $m(x_0)$ is not a repeated root of $G(m, x_0) = 0$ (note that $x$ is a repeated root of polynomial equation $f(x) = 0$ if and only if $f'(x) = 0$).

By applying the implicit function theorem, we obtain that there exist a neighborhood $D \in \mathbb{C}$ of $x_0$ and a neighborhood $D' \in \mathbb{C}$ of $m(x_0)$ such that $m(x)$ is a one-to-one mapping from $D$ onto $D'$ and $m(x)$ is not a repeated root of $G(m, x) = 0$, $\forall x \in D$, which is assured by the continuity of $\frac{\partial G}{\partial m}$. The neighborhoods $D$ and $D'$ can be selected sufficiently small such that

$$\left| \frac{\partial G}{\partial m} \bigg|_{m=m(x)} \right| > C, \quad \forall x \in D$$

where $C$ is a positive constant.

$$\left| \frac{\partial G}{\partial z} \bigg|_{z=x_0} \right|$$

is upper-bounded by a constant $C' > 0$ since $\{m_k(z)\}_k$ are analytic functions. Therefore, we have

$$\left| \frac{dm}{dz} \bigg|_{z=x_0} \right| = \left| \frac{\partial G}{\partial m} \bigg|_{z=x_0} \right| < C'$$

where $C'' = \frac{C'}{C}$. Note that $C''$ is common for all $x \in D$.

Therefore, for $x + iv \in D$, we have the following Taylor expansion:

$$m(x + iv) - m(x) = \frac{dm}{dz} \bigg|_{z=x_0} iv + o(v^2)$$

which implies that

$$\left| m(x + iv) - m(x) \right| < \left| \frac{dm}{dz} \bigg|_{z=x_0} \right| |v| + o(v^2)$$

$$< C'' |v| + o(v^2).$$

Therefore, the convergence of $\lim_{v \to 0} m(x + iv)$ is controlled by a term $C'' |v|$ independent of $x$, $\forall x \in D \cap [a, b]$, thus implying the uniform convergence in $D \cap [a, b]$, which is a neighborhood of $x_0$. \qed
APPENDIX II
PROOF OF THEOREM 4

Proof: For property 1), we apply the discriminant \( \Delta(x) \). From the definition of the discriminant of a cubic equation, we obtain

\[
\Delta(x) = \frac{-a^2b^2}{108} + \frac{c^2}{4} + \frac{b^3}{27} + \frac{a^2c}{27} - \frac{abc}{6}
\]

where

\[
\begin{cases}
a = -(2-\alpha-\beta) \\
b = -(\alpha\beta(1-\alpha)(1-\beta)) \\
c = -\frac{a^2}{27}.
\end{cases}
\]

Therefore, the numerator of \( \Delta(x) \) is a cubic polynomial in \( x \) and the denominator is \( x^6 \), which implies that there are at most three real roots of \( \Delta(x) = 0 \). When \( x \to \infty \) or \( x \to -\infty \), it is easy to check that \( \Delta(x) \) is dominated by the term \( \frac{b^3}{27} \), which is negative when \( x > 0 \) and positive when \( x < 0 \). When \( x \to 0 \), it is easy to check that \( \Delta(x) \) is dominated by the expression at the bottom of the page.

Therefore, we have

1) \( \Delta(x) < 0 \) for sufficiently large \( x > 0 \);  
2) \( \Delta(x) > 0 \) for sufficiently small \( x < 0 \); and  
3) \( \Delta(x) < 0 \) for \( x \) sufficiently close to 0.

Due to 1) and 3), there are an even number of roots for \( \Delta(x) = 0 \) when \( x > 0 \). The number cannot be zero (since the probability of a positive eigenvalue is not zero) or four (since \( \Delta(x) = 0 \) has at most three real roots). Therefore, \( \Delta(x) = 0 \) has two positive roots and one negative root. Between the two positive roots, \( \Delta(x) > 0 \), which implies that the equation \( G(m,x) = 0 \) has two conjugate complex roots and \( \hat{f}(x) > 0 \). For other positive \( x \), \( \Delta(x) < 0 \), which implies that the equation \( G(m,x) = 0 \) has only one real root and \( \hat{f}(x) = 0 \). This completes the proof of property 1).

In property 2), the fact that \( \lambda_{\max} \geq \lambda_{\max} \) is simply a corollary of Corollary 3. If \( \lambda_{\max} < \lambda_{\max} \), then \( \hat{f}(\lambda) > 0 \), \( \forall \lambda \in \lambda_{\max}, \lambda_{\max} \), which implies that there exists a sufficiently large \( p \in \mathbb{N} \) such that \( \lambda_p > \lambda_p \), thus contradicting Corollary 3. To prove the other inequality, we need the following lemma, which is a corollary of the Courant–Fischer theorem ([11, Theorem 4.2.11]).

Lemma 12: For an \( n \times n \) Hermitian matrix \( \mathbf{A} \), the maximal eigenvalue \( \lambda_{(n)} \) is given by

\[
\lambda_{(n)} = \max_{\mathbf{u} \in \mathbb{C}^n} \frac{\mathbf{u}^H \mathbf{A} \mathbf{u}}{\mathbf{u}^H \mathbf{u}}.
\]

Then, for finite \( K, N, M \), we have

\[
\lambda_{\max} = \max_{\mathbf{u} \in \mathbb{C}^K} \left( \frac{\mathbf{u}^H \mathbf{S} \mathbf{X}^H \mathbf{S}^H \mathbf{u}}{\mathbf{S} \mathbf{u}^H \mathbf{u}} \right)
\]

\[
= \max_{\mathbf{u} \in \mathbb{C}^K} \left( \frac{\mathbf{u}^H \mathbf{X}^H \mathbf{S}^2 \mathbf{u}}{\mathbf{S} \mathbf{u}^H \mathbf{u}} \right)
\]

\[
\leq \max_{\mathbf{v} \in \mathbb{C}^K} \left( \frac{\mathbf{v}^H \mathbf{X}^H \mathbf{S}^2 \mathbf{v}}{\mathbf{S} \mathbf{v}^H \mathbf{v}} \right)
\]

\[
= \lambda_{\max} \max_{\mathbf{v} \in \mathbb{C}^K} \left( \frac{\mathbf{v}^H \mathbf{S}^2 \mathbf{v}}{\mathbf{S} \mathbf{v}^H \mathbf{v}} \right)
\]

(88)

where \( \lambda_{\max} \) denotes the maximal eigenvalue of the Hermitian matrix \( \frac{1}{M} \mathbf{X}^H \mathbf{X} \). Letting \( K, N, M \to \infty \) in (88), we obtain the desired inequality by applying the fact that \( \lambda_{\max} \to (1 + \min(\sqrt{\frac{a}{\beta}}, \sqrt{\frac{b}{\alpha}}))^2 \).

For property 3), we need to show only that \( \lambda_{\min} \leq \lambda_{\min} \) for sufficiently large \( \alpha \) since \( \lambda_{\max} \geq \lambda_{\max} \) holds for all \( \alpha \), due to property 2). The inequality \( \lambda_{\min} \leq \lambda_{\min} \) is equivalent to \( \hat{f}(\lambda_{\min}) > 0 \). Therefore, we need to show only that, for sufficiently large \( \alpha \), the equation \( G(m, \lambda_{\min}) = 0 \) has only one real root. Due to (58), the function \( G(m, x) \) can be separated into two parts, namely, \( G(m, x) = G_1(m, x) - \alpha G_2(m, x) \), where

\[
G_1(m, x) = x^2m^3 + (2 - \beta)xm^2 + ((1 - \beta)m, \\
= m(xm + 1)(xm + (1 - \beta))
\]

and

\[
G_2(m, x) = x^2m^2 + (x + (1 - \beta)m + 1.
\]

Notice that \( G_1(m, x) = 0 \) has three real roots, namely, \( 0, \frac{-1}{\beta}, \) and \( \frac{-1}{\alpha} \). Therefore, it is easy to check that \( G_1(m, x) \) is positive only in the interval \( (\frac{-1}{\beta}, \frac{-1}{\alpha}) \) when \( m < 0 \) since \( G_1(m, x) \to -\infty \) as \( x \to -\infty \). For \( G_2(m, x) \), the minimum is achieved at \( m_0(x) = \frac{1}{(1 - \beta)m} \). Now we fix \( x = \lambda_{\min} \). It is easy to check that \( G_2(m, \lambda_{\min}) \geq 0 \) and the equality holds at only \( m_0(x) = \frac{-1}{\lambda_{\min}(1 - \beta)} \). So we have

\[
m_0(\lambda_{\min}) = \frac{1}{\lambda_{\min}(1 - \beta)} = \frac{\lambda_{\min} + (1 - \beta)}{2\lambda_{\min} - 1} \frac{1}{1 - \beta} + \frac{1}{\lambda_{\min} - 1} \frac{1}{1 - \beta} = \frac{1}{\lambda_{\min} - 1} \frac{1}{1 - \beta} + \frac{1}{\lambda_{\min} - 1} \frac{1}{1 - \beta} = \frac{\beta - \beta}{\lambda_{\min} - 1} > 0
\]

where the last inequality is due to the fact that \( \beta < 1 \). Therefore, we have \( m_0(\lambda_{\min}) > \frac{1}{\lambda_{\min}(1 - \beta)} \), which implies that \( G_2(m, \lambda_{\min}) \) > 0 in the interval \( (\frac{1}{\lambda_{\min}}, \frac{1}{\lambda_{\min}(1 - \beta)}) \). Thus,
when $\alpha$ is sufficiently large, $G(m, \lambda_{\min})$ is negative for all $m < 0$, which implies that there are no real roots for $m < 0$. Now, suppose that there are three positive real roots for $G_1(m, \lambda_{\min}) = 0$, denoted by $m_1, m_2,$ and $m_3$. Then $m_1 m_2 + m_2 m_3 + m_3 m_1 = -\frac{\alpha}{\sigma_n^2} \frac{1}{1-\beta}$ due to (20), which is negative when $\alpha$ is sufficiently large and contradicts the assumption that $m_1, m_2,$ and $m_3$ are all positive. Therefore, $G_1(m, \lambda_{\min})$ has one real positive root and two conjugate complex roots, which implies $f(\lambda_{\min}) > 0$. Due to the continuity of $f$, there exists a neighborhood of $\lambda_{\min}$ within which $f$ is positive. Therefore, $\lambda_{\min} < \lambda_{\min}$. This completes the proof of property 3.

The proof of property 4) is similar to that of property 3). Again, we separate $G(m, x)$ into two parts $G_1(m, x)$ and $G_2(m, x)$. When $x > \lambda_{\max}$, $f(x) = 0$, and $G_2(m, x) > 0$, $\forall m \in \mathbb{R}$. Then, for a sufficiently small $\epsilon > 0$, take a sufficiently small $\alpha$ such that

$$0 < \alpha < \min_{m \in [-\frac{1}{\lambda} + \epsilon, -\frac{1-\beta-\epsilon}{\sigma_n^2}]} G_1(m, x) \max_{m \in [-\frac{1}{\lambda} + \epsilon, -\frac{1-\beta-\epsilon}{\sigma_n^2}]} G_2(m, x).$$

Note that such an $\alpha$ exists since the numerator and denominator are both positive (the positivity of the numerator is shown in the proof of property 3) and that of the denominator is due to the property $f(\lambda_{\max}) = 0$. Then, for this $\alpha$, $G(m, x) > 0$ for all $m \in [-\frac{1}{\lambda} + \epsilon, -\frac{1-\beta-\epsilon}{\sigma_n^2}]$. Combining the facts that $G(0, x) < 0$ and $G(-\infty, x) < 0$, we obtain that there are two negative real roots for equation $G(m, x) = 0$, which implies that there are three real roots for equation $G(m, x) = 0$ and thus $f(\lambda_{\max}) = 0$. Since $\lambda_{\max} \geq \lambda_{\max}$, we have $\lambda_{\min} > \lambda_{\max}$, otherwise contradicting the continuity of the eigenvalue distribution support in property 1). Finally, we need to show that it is necessary that $\alpha < \beta$ for $\lambda_{\min} > \lambda_{\max}$. Suppose that $\alpha > \beta$ and $\lambda_{\min} > \lambda_{\max}$. Then, the probability at mass point $x = 0$ is $0$ for both $\lambda$ and $\lambda$. Since for all $\lambda > 0$ and $\lambda < 0$, $\lambda > \lambda$, we have $E[\lambda] > E[\lambda]$, which contradicts the fact $E[\lambda] = E[\lambda] = \beta$ in (42). This completes the proof.

**APPENDIX III**

**PROOF OF THEOREM 6**

We need the following lemma for the proof of Theorem 6, which states that the PDF $f(x)$ is positive at $x = \sigma_n^2$ for sufficiently large $\alpha$. This is intuitive since there is a mass point at $x = \sigma_n^2$ with probability $1 - \beta$ when $\alpha \rightarrow \infty$. However, the proof is nontrivial.

**Lemma 13:** When $\sigma_n^2 > 0$, we have $f(\sigma_n^2) > 0$ for sufficiently large $\alpha$.

**Proof:** For the properties of PDF support, we follow an approach similar to that used in the proof of Theorem 4, namely, rewriting the function $G(m, z)$ in (80) as $G(m, z)$ in (80) as $G(m, z) = G_1(m, z) + \sigma^2 G_2(m, z)$, where

$$G_1(m, z) = \sigma_n^2 z^2 m^4 + 2 \sigma_n^2 z m^3 + \sigma_n^2 m^2 \alpha \left((z^2 - 2 \sigma_n^2 z) m^3 + ((2 - \beta - \sigma_n^2) z) \frac{-2 \sigma_n^2}{m^2} + (1 - \beta - \sigma_n^2) m) \right)$$

$$+ \frac{(z^2 - 2 \sigma_n^2 z) m^3 + ((2 - \beta - \sigma_n^2) z)}{m^2} + (1 - \beta - \sigma_n^2) m \right)$$

and

$$G_2(m, z) = (z - \sigma_n^2) m^2 + (z - \sigma_n^2 + (1 - \beta) m + 1.$$  

Obviously, $G_1(m, \sigma_n^2) = 0$ has four real roots for sufficiently large $\alpha$, namely, $0$, and the two real roots of $(\sigma_n^2)^2 m^2 + (1 - \alpha) \sigma_n^2 m + (1 - \beta - \sigma_n^2) m = 0$ (denoted by $m_1$ and $m_2$ and suppose $m_1 < m_2$).

Suppose that $\sigma_n^2 > 1 - \beta$. Then $m_1 < 0$ and $m_2 > 0$ since $m_1 m_2 = (1 - \beta - \sigma_n^2) < 0$. Therefore, it is easy to check that $G_1(m, \sigma_n^2)$ is positive in the intervals $(-\infty, m_1)$, $(-\frac{1}{\lambda}, 0)$, and $(m_2, \infty)$ (note that $m_1 < -\frac{1}{\lambda}$ for sufficiently large $\alpha$), and is negative in both $(m_1, -\frac{1}{\lambda})$ and $(0, m_2)$. Also, we note that $G_2(m, \sigma_n^2) = (1 - \beta) m + 1 = 0$ has one root at $m = -\frac{1}{\beta}$, which lies within the interval $(m_1, -\frac{1}{\lambda})$.

Now, we begin to show that there are only two real roots for equation $G(m, \sigma_n^2) = 0$. First, we show that there is only real root for $m < m_2$. For sufficiently large $\alpha$, there exists an $\epsilon > 0$ such that when $m < -\frac{1}{\lambda} + \epsilon$, $G_1(m, \sigma_n^2) > \alpha^2 G_2(m, \sigma_n^2)$, and when $-\frac{1}{\lambda} + \epsilon < m \leq m_2$, $G_1(m, \sigma_n^2) < \alpha^2 G_2(m, \sigma_n^2)$. When $-\frac{1}{\lambda} - \epsilon < x < -\frac{1}{\lambda} + \epsilon$, sufficiently large $\alpha$ assures that $G(m, \sigma_n^2)$ decreases monotonically. Therefore, there is only one real root satisfying $x \leq m_2$, which is located within $(-\frac{1}{\lambda} - \epsilon, -\frac{1}{\lambda} + \epsilon)$.

Then, we show that there is only one real root for equation $G(m, \sigma_n^2) = 0$ when $m > m_2$ (note that $G(m_2, \sigma_n^2) < 0$ and therefore $m_2$ is not a root). Since $\alpha$ is finite, $G_1(m, \sigma_n^2)$ is larger than $G_2(m, \sigma_n^2)$ for sufficiently large $m$. Therefore, there is at least one positive root when $m > m_2$ since $G(m_2, \sigma_n^2) < 0$, among which the smallest positive root is denoted by $m_0'$. It is easy to check that the second-order derivative of $G(\alpha, \sigma_n^2)$ is given by

$$\frac{d^2 G(m, \sigma_n^2)}{dm^2} = 2 \sigma_n^2 C_1(m) + (4 \sigma_n^2 m + 2) C_2(m) + 2 (\sigma_n^2)^2 m (\sigma_n^2 m + 1)$$

where $C_1(m) = (\sigma_n^2)^2 m^2 + (1 - \alpha) \sigma_n^2 m + (1 - \beta - \sigma_n^2)$ and $C_2(m) = C_1(m) = (\sigma_n^2)^2 m^2 + (1 - \beta - \sigma_n^2)$. Note that $C_1(m_2) = 0$; therefore, $C_1(m) > 0$ for all $m > m_2$ and sufficiently large $\alpha$. Thus, $C_2(m) > 0$ for all $m > m_2$ and sufficiently large $\alpha$ since $C_2(m) = C_1(m) \sigma_n^2 m + (1 - \sigma_n^2) < 0$ (recall $\sigma_n^2 > 1 - \beta$). So, we obtain that $\frac{d^2 G(m, \sigma_n^2)}{dm^2} > 0$ when $m > m_2$ and $\alpha$ is sufficiently large, which implies that $G(m, \sigma_n^2)$ is convex when $m > m_2$. Suppose that there exists a real root larger than $m_2$. Then, there exists an $m_0 > m_2$ such that $\frac{d^2 G(m, \sigma_n^2)}{dm^2} = 0$ at $m_0$ since $G(m_0, \sigma_n^2) > 0$ for sufficiently small $\epsilon > 0$. Combining with the fact that $\frac{d^2 G(m, \sigma_n^2)}{dm^2} > 0$ at $m_2$ (which is easy to check from the expression for $\frac{d^2 G(m, \sigma_n^2)}{dm^2}$), we obtain that $G(m, \sigma_n^2)$ is not convex when $m > m_2$, which contradicts the convexity of $G(m, \sigma_n^2)$ when $m > m_2$. Thus, there exists only one positive root for the equation $G(m, \sigma_n^2) = 0$ when $\alpha$ is sufficiently large.
Combining the above results, we reach the conclusion that there exist only two real roots for the equation \( G(m, \sigma^2_{n}) = 0 \). Therefore, the PDF of \( \lambda \) is positive at \( \sigma^2_{n} \).

When \( \sigma^2_{n} \leq 1 - \beta \), then \( m_1 > 0 \) and \( m_2 > 0 \), in which case we can follow the same argument as for the case \( \sigma^2_{n} > 1 - \beta \). This completes the proof.

Based on Lemma 13, we can prove Theorem 6. Note that we still use the definitions of \( G_1(m, x) \) and \( G_2(m, x) \) in the proof of Lemma 13.

Proof: The proof that there are no mass points for positive eigenvalues is the same as that of Lemma 7.

Now we begin to prove property 1) for the support of \( f \).

First, we show that \( \forall \sigma^2_{n} < x < (1 - \sqrt{\beta})^2 + \sigma^2_{n} \), there exists a sufficiently large \( \alpha \) such that \( \hat{f}(x) = 0 \). Notice that for the exact covariance \( R_n, \hat{f}(x) = 0 \) since it lies between the mass point \( \sigma^2_{n} \) and the range of positive eigenvalues \((1 - \sqrt{\beta})^2 + \sigma^2_{n} (1 + \sqrt{\beta})^2 + \sigma^2_{n}\). Therefore, \( G_2(m, x) = 0 \) has two real roots, which is equivalent to the existence of real numbers \( m_1 < m_2 < m_3 \), all being independent of \( \alpha \) such that \( G_2(m_1, x) > 0, G_2(m_2, x) < 0, \) and \( G_2(m_3, x) > 0 \). Thus, there exists sufficiently large \( \alpha > 0 \) such that \( G_2(m, x) \) dominates \( G(m, x) \), which means that \( G(m_1, x) < 0, G(m_2, x) > 0, \) and \( G(m_3, x) < 0 \). By applying the fact that \( G(\infty, x) > 0 \) and \( G(-\infty, x) > 0 \), we come to the conclusion that the equation \( G(m, x) = 0 \) has four real roots and therefore \( \hat{f}(x) = 0 \).

From Lemma 13, we have that \( \hat{f}(\sigma^2_{n}) > 0 \) for sufficiently large \( \alpha \). So, due to the fact that \( \hat{f}(x) \) is continuous for positive \( x, \forall \alpha > 0 \), there exists an \( \epsilon > 0 \) such that \( \hat{f}(x) > 0, \forall \sigma^2_{n} < x < \sigma^2_{n} + \epsilon \). This concludes the proof of property 1).

The argument for the proof of property 2) is similar to that used in Theorem 4. The inequality \( \tilde{\lambda}_{\max} > \lambda_{\max} \) can be shown in the same way as in the proof of Theorem 4. Thus, we need to show only the second inequality. By applying Corollary 10, we need only consider the minimal eigenvalue of \( \frac{1}{\Lambda} Y (SS^H + \sigma^2_{n} I) Y^H \), where the matrix \( Y \) is an \( M \times N \) matrix with mutually independent random elements having unit variance. By applying Lemma 12, we have

\[
\tilde{\lambda}_{\max} = \max_{u \in \mathbb{C}^M} \left( \frac{u^H Y (SS^H + \sigma^2_{n} I) Y^H u}{Mu^H u} \right)
= \max_{u \in \mathbb{C}^M} \left( \frac{u^H Y (SS^H + \sigma^2_{n} I) Y^H u}{u^H Y Y^H u} \frac{u^H Y Y^H u}{Mu^H u} \right)
\leq \max_{u \in \mathbb{C}^N} \left( \frac{v^H (SS^H + \sigma^2_{n} I) v}{v^H v} \right) \max_{u \in \mathbb{C}^M} \left( \frac{u^H Y Y^H u}{Mu^H u} \right)
\rightarrow \left( (1 + \sqrt{\beta})^2 + \sigma^2_{n} \right) \left( 1 + \min \left( \frac{1}{\alpha}, \sqrt{\alpha} \right) \right)^2.
\]

Therefore, we obtain the second inequality in property 2) by applying the fact that \( \lambda_{\max} = (1 + \sqrt{\beta})^2 + \sigma^2_{n} \).

For property 3), the inequality \( \lambda_{\max} \geq \lambda_{\min} \) holds for all \( \alpha \). The inequality \( \lambda_{\min} \leq \lambda_{\min} = \sigma^2_{n} \) is obtained from the continuity and the fact that \( \hat{f}(\lambda_{\min}) > 0 \) in Lemma 13.

For property 4), the proof is similar to that of Theorem 4. This concludes the proof.

APPENDIX IV

PROOF OF PROPOSITION 2

Proof: The conclusion can be obtained from the explicit expression in Corollary 12.

Alternatively, we can derive it in a more direct way. Notice that the term scaled by \( \frac{1}{\alpha} \) corresponds to the partition with \( m = p - 1 \) (recall that \( m \) denotes the number of blocks in the partition of sample indices), namely, the sample indices \( \{m_j\}_{j=1,...,p} \) are partitioned into \( p - 1 \) blocks. Therefore, only two sample indices are partitioned into the same block and all other indices form single element blocks. This is equivalent to connecting two points corresponding to sample indices in the extended complementation map and thus dividing the sample indices into two continuous sets with cardinalities of \( i_1 \) and \( i_2 \) satisfying \( i_1 + i_2 = p \). Then, it is easy to check that the number of possible choices of lines is \( p \) when \( i_1 \neq i_2 \) and \( \frac{p}{2} \) when \( i_1 = i_2 = \frac{p}{2} \) (\( p \) is even). When the partition of sample indices is fixed, the corresponding expectation equals \( \lambda_{i_1} \lambda_{i_2} \) since the dimension and user indices are divided into two independent sets. By summing all possible partitions of sample indices, we obtain the conclusion. Again, the indices \( i_1 \) and \( i_2 \) are indistinguishable because the sum depends on only the sizes of the blocks.

APPENDIX V

PROOF OF PROPOSITION 3

Proof: For sufficiently large \( \alpha \) and to a first-order approximation, we can ignore the first part of (83) and rewrite it as

\[
l_h(m, \epsilon) \triangleq (x - \sigma^2_{n} m^2 + (x - \sigma^2_{n} + (1 - \beta)) m + 1 - \epsilon(x^2 - 2\sigma^2_{n}) m^2 + ((2 - \beta - \sigma^2_{n}) x - 2\sigma^2_{n}) m^2 + (1 - \beta - \sigma^2_{n}) m = 0 \]

where \( \epsilon = \frac{1}{\alpha} \).

Consider \( m \) as a function of \( \epsilon \). Then, by applying the implicit function theorem, we have

\[
\frac{dm}{d\epsilon} \bigg|_{\epsilon=0} = -\frac{\frac{\partial l_h}{\partial \epsilon}}{\frac{\partial l_h}{\partial m}} \bigg|_{\epsilon=0} = \frac{g_n(m)}{2(x - \sigma^2_{n}) m + (x - \sigma^2_{n} + (1 - \beta)) m + 1} \bigg|_{\epsilon=0} = \frac{m g_n(m)}{(x - \sigma^2_{n}) m^2 - 1} \bigg|_{\epsilon=0}
\]

where the last equation follows from the fact that \( (x - \sigma^2_{n}) m^2 + (x - \sigma^2_{n} + (1 - \beta)) m + 1 = 0 \) when \( \epsilon = 0 (\alpha \to \infty) \).
When $\epsilon = 0$, $m$ is given by the root of (50). Thus, (87) is obtained by first order expansion.

REFERENCES

[18] H. Vincent Poor (S’72–M’77–SM’82–F’87) received the Ph.D. degree in electrical engineering and computer science from Princeton University, Princeton, NJ, in 1977.
[19] From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990, he has been on the faculty at Princeton University, where he is the Dean of Engineering and Applied Science, and the Michael Henry Strater University Professor of Electrical Engineering. His research interests are in the areas of stochastic analysis, statistical signal processing and their applications in wireless networks, and related fields. Among his publications in these areas are the recent books MIMO Wireless Communications (Cambridge, U.K.: Cambridge University Press, 2007), coauthored with Ezio Biglieri et al., and Quickest Detection (Cambridge, U.K.: Cambridge University Press, 2009), coauthored with Olympia Hadjiliadis.
[20] Dr. Poor is a member of the National Academy of Engineering, a Fellow of the American Academy of Arts and Sciences, and a former Guggenheim Fellow. He is also a Fellow of the Institute of Mathematical Statistics, the Optical Society of America, and other organizations. In 1990, he served as President of the IEEE Information Theory Society, and during 2004–2007 as the Editor-in-Chief of the IEEE Transactions on Information Theory. He was the recipient of the 2007 IEEE Marconi Prize Paper Award, the 2007 Technical Achievement Award of the IEEE Signal Processing Society, and the 2008 Aaron D. Wyner Distinguished Service Award of the IEEE Information Theory Society.