Random matrices applications to signal processing

Romain Couillet and Mérouane Debbah

Random matrix theory deals with the study of *matrix-valued random variables*. It is conventionally considered that random matrix theory dates back to the work of Wishart in 1928 [1] on the properties of matrices of the type $\mathbf{X}\mathbf{X}^{\mathsf{H}}$ with $\mathbf{X} \in \mathbb{C}^{N \times n}$ a random matrix with independent Gaussian entries with zero mean and equal variance. Wishart and his followers were primarily interested in the joint distribution of the entries of such matrices and then on their eigenvalue distribution. It then dawned to mathematicians that, as the matrix dimensions N and n grow large with ratio converging to a positive value, its eigenvalue distribution converges weakly and almost surely to some deterministic distribution, which is somewhat similar to a law of large numbers for random matrices. This triggered a growing interest in particular among the signal processing community, as it is usually difficult to deal efficiently with large dimensional data because of the so-called curse of dimensionality. Other fields of research have been interested into large dimensional random matrices, among which the field of wireless communications, as the eigenvalue distribution of some random matrices is often a sufficient statistics for the performance evaluation of multi-dimensional wireless communication systems.

In the following, we introduce the main notions, results and details of classical as well as recent techniques to deal with large random matrices.

0.1 Probability notations

In this chapter, an *event* will be the element ω of some set Ω. Based on Ω, we will consider the probability space $(\Omega, \mathcal{F}, P)$, with $\mathcal{F}$ some σ-field on Ω and P a probability measure on $\mathcal{F}$. If X is a random variable on Ω, we will denote

$$\mu_X(A) \triangleq P\left(\{\omega, X(\omega) \in A\}\right)$$

the probability distribution of X.

When μ_X has a probability density function (p.d.f.), it will be denoted P_X, i.e., for X with image in $\mathbb{R}$ with Lebesgue measure and for all measurable f,

$$\int f(x) P_X(x) dx \triangleq \int f(x) \mu_X(dx).$$

To differentiate between multidimensional random variables and scalar random variables, we may denote $p_X(x) \triangleq P_X(x)$, in lowercase character, if X is scalar. The (cumulative) distribution function (d.f.) of a real random variable will often be denoted by the letter F, e.g., for $x \in \mathbb{R}$,

$$F(x) \triangleq p_X((-\infty, x])$$

denotes the d.f. of X.

We further denote, for X, Y two random variables with density, and for y such that $P_Y(y) > 0$,

$$P_{X|Y}(x, y) \triangleq \frac{P_{X,Y}(x, y)}{P_Y(y)}$$

the conditional probability density of X given Y.

0.2 Spectral distribution of random matrices

We start this section with a formal definition of a random matrix and the introduction of necessary notations.

Definition 0.1. *An $N \times n$ matrix $\mathbf{X}$ is said to be a random matrix if it is a matrix-valued random variable on some probability space $(\Omega, \mathcal{F}, P)$ with entries in some measurable space $(\mathcal{R}, \mathcal{G})$, where $\mathcal{F}$ is a σ-field on Ω with probability measure P and $\mathcal{G}$ is a σ-field on $\mathcal{R}$. As per conventional notations, we denote $\mathbf{X}(\omega)$ the realization of the variable $\mathbf{X}$ at point $\omega \in \Omega$.*

We shall in particular often consider the marginal probability distribution function of the eigenvalues of random Hermitian matrices $\mathbf{X}$. Unless otherwise stated, the d.f. of the *real* eigenvalues of $\mathbf{X}$ will be denoted $F^{\mathbf{X}}$.

We now discuss the properties of the so-called Wishart matrices and some known results on unitarily invariant random matrices. These properties are useful to the characterisation e.g. of Neyman-Pearson tests for signal sensing procedures [2], [3].

0.2.1 Wishart matrices

We start with the definition of a Wishart matrix.

Definition 0.2. *The $N \times N$ random matrix $\mathbf{X}\mathbf{X}^{\mathsf{H}}$ is a (real or complex) central Wishart matrix with n degrees of freedom and covariance matrix $\mathbf{R}$ if the columns of the $N \times n$ matrix $\mathbf{X}$ are zero mean independent (real or complex) Gaussian vectors with covariance matrix $\mathbf{R}$. This is denoted*

$$\mathbf{X}\mathbf{X}^{\mathsf{H}} \sim \mathcal{W}_N(n, \mathbf{R}).$$

Defining the *Gram matrix* associated to any matrix $\mathbf{X}$ as being the matrix $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, $\mathbf{X}\mathbf{X}^{\mathsf{H}} \sim \mathcal{W}_N(n, \mathbf{R})$ is by definition the Gram matrix of a matrix with Gaussian i.i.d. columns with zero mean and variance $\mathbf{R}$. When $\mathbf{R} = \mathbf{I}_N$, it is usual to refer to $\mathbf{X}$ as a *standard Gaussian matrix*.

One interest of Wishart matrices in signal processing applications lies in the following remark.

Remark 0.1. *Let* $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{C}^N$ *be* n *independent samples of the random process* $\mathbf{x}_1 \simeq \mathcal{CN}(0, \mathbf{R})$. *Then, denoting* $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$,

$$\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{H}} = \mathbf{X}\mathbf{X}^{\mathsf{H}}.$$

For this reason, the random matrix $\mathbf{R}_n = \frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}}$ *is often referred to as an (empirical) sample covariance matrix associated to the random process* $\mathbf{x}_1$. *This is to be contrasted with the population covariance matrix* $\mathrm{E}[\mathbf{x}_1\mathbf{x}_1^{\mathsf{H}}] = \mathbf{R}$. *Of particular importance is the case when* $\mathbf{R} = \mathbf{I}_N$. *In this situation,* $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, *sometimes referred to as a zero (or null) Wishart matrix, is proportional to the sample covariance matrix of a white Gaussian process. The zero (or null) terminology is due to the signal processing problem of hypothesis testing, in which one has to decide whether the observed* $\mathbf{X}$ *emerges from a white noise process or from an information plus noise process.*

Wishart provides us with the joint probability density function of the entries of Wishart matrices, as follows

Theorem 0.1 ([1])**.** *The p.d.f. of the complex Wishart matrix* $\mathbf{X}\mathbf{X}^{\mathsf{H}} \simeq \mathcal{W}_N(n, \mathbf{R})$, $\mathbf{X} \in \mathbb{C}^{N\times n}$, *for* $n \geq N$ *is*

$$P_{\mathbf{X}\mathbf{X}^{\mathsf{H}}}(\mathbf{B}) = \frac{\pi^{N(N-1)/2}}{\det \mathbf{R}^n \prod_{i=1}^{N}(n-i)!} e^{-\mathrm{tr}(\mathbf{R}^{-1}\mathbf{B})} \det \mathbf{B}^{n-N}. \tag{1}$$

Note in particular that for $N = 1$, this is a conventional chi-square distribution with n degrees of freedom.

For null Wishart matrices, notice that $P_{\mathbf{X}\mathbf{X}^{\mathsf{H}}}(\mathbf{B}) = P_{\mathbf{X}\mathbf{X}^{\mathsf{H}}}(\mathbf{U}\mathbf{B}\mathbf{U}^{\mathsf{H}})$, for any unitary $N \times N$ matrix $\mathbf{U}$.[1] Otherwise stated, the eigenvectors of the random variable $\mathbf{X}\mathbf{X}^{\mathsf{H}}$ are uniformly distributed over the space $\mathcal{U}(N)$ of unitary $N \times N$ matrices. As such, the eigenvectors do not carry relevant information, and $P_{\mathbf{X}\mathbf{X}^{\mathsf{H}}}(\mathbf{B})$ is only a function of the eigenvalues of $\mathbf{B}$. This property will turn out essential to the derivation of further properties of Wishart matrices.

The joint p.d.f. of the eigenvalues of zero Wishart matrices were studied simultaneously in 1939 by different authors [4], [5], [6], [7]. The two main results are summarized in the following,

Theorem 0.2. *Let the entries of* $\mathbf{X} \in \mathbb{C}^{N\times n}$, $n > N$, *be i.i.d. Gaussian with zero mean and unit variance. The joint p.d.f.* $P_{(\lambda_i)}$ *of the ordered eigenvalues* $\lambda_1 \geq \ldots \geq \lambda_N$ *of the zero Wishart matrix* $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, *is given by*

$$P_{(\lambda_i)}(\lambda_1, \ldots, \lambda_N) = e^{-\sum_{i=1}^{N}\lambda_i} \prod_{i=1}^{N} \frac{\lambda_i^{n-N}}{(n-i)!(N-i)!} \Delta(\mathbf{\Lambda})^2,$$

where, for a Hermitian nonnegative $N \times N$ *matrix* $\mathbf{\Lambda}$,[2] $\Delta(\mathbf{\Lambda})$ *denotes the Vandermonde determinant of its eigenvalues* $\lambda_1, \ldots, \lambda_N$,

$$\Delta(\mathbf{\Lambda}) \triangleq \prod_{1\leq i<j\leq N} (\lambda_j - \lambda_i).$$

[1] we remind that a unitary matrix $\mathbf{U} \in \mathbb{C}^{N\times N}$ is such that $\mathbf{U}\mathbf{U}^{\mathsf{H}} = \mathbf{U}^{\mathsf{H}}\mathbf{U} = \mathbf{I}_N$.

[2] all along this work, we will respect the convention that x (be it a scalar or an Hermitian matrix) is nonnegative if $x \geq 0$, while x is positive if $x > 0$.

The marginal p.d.f. p_λ ($\triangleq P_\lambda$) *of the unordered eigenvalues is*

$$p_\lambda(\lambda) = \frac{1}{M}\sum_{k=0}^{N-1}\frac{k!}{(k+n-N)!}[L_k^{n-N}(\lambda)]^2\lambda^{n-N}e^{-\lambda},$$

where $L_n^k(\lambda)$ *are the Laguerre polynomials defined as*

$$L_n^k(\lambda) = \frac{e^\lambda}{k!\lambda^n}\frac{d^k}{d\lambda^k}(e^{-\lambda}\lambda^{n+k}).$$

The generalized case of (non-zero) central Wishart matrices is more involved since it requires advanced tools of multivariate analysis, such as the fundamental Harish-Chandra integral [8]. We will mention the result of Harish-Chandra, which is at the core of the results in signal sensing presented later in Section 0.5.1.

Theorem 0.3. *For non singular* $N \times N$ *positive definite Hermitian matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *of respective eigenvalues* $a_1, \ldots, a_N$ *and* $b_1, \ldots, b_N$*, such that for all* $i \neq j$*,* $a_i \neq a_j$ *and* $b_i \neq b_j$*, we have*

$$\int_{\mathbf{U}\in\mathcal{U}(N)} e^{\kappa \mathrm{tr}(\mathbf{A}\mathbf{U}\mathbf{B}\mathbf{U}^\mathsf{H})}d\mathbf{U} = \left(\prod_{i=1}^{N-1} i!\right)\kappa^{\frac{1}{2}N(N-1)}\frac{\det\left(\{e^{-b_j a_i}\}_{1\leq i,j\leq N}\right)}{\Delta(\mathbf{A})\Delta(\mathbf{B})}$$

where, for any bivariate function f*,* $\{f(i,j)\}_{1\leq i,j\leq N}$ *denotes the* $N \times N$ *matrix of* (i,j) *entry* $f(i,j)$*, and* $\mathcal{U}(N)$ *is the space of* $N \times N$ *unitary matrices.*

This result enables the calculus of the marginal joint-eigenvalue distribution of (non-zero) central Wishart matrices [9], given as follows

Theorem 0.4. *Let the columns of* $\mathbf{X} \in \mathbb{C}^{N\times n}$ *be i.i.d. zero mean Gaussian with positive definite covariance* $\mathbf{R}$*. The joint p.d.f.* $P_{(\lambda_i)}$ *of the ordered positive eigenvalues* $\lambda_1 \geq \ldots \geq \lambda_N$ *of the central Wishart matrix* $\mathbf{X}\mathbf{X}^\mathsf{H}$*, reads*

$$P_{(\lambda_i)}(\lambda_1, \ldots, \lambda_N) = \frac{\det(\{e^{-r_j^{-1}\lambda_i}\}_{1\leq i,j\leq N})}{\Delta(\mathbf{R}^{-1})}\Delta(\mathbf{\Lambda})\prod_{j=1}^{N}\frac{\lambda_j^{n-N}}{r_j^n(n-j)!}$$

where $r_1 \geq \ldots \geq r_N$ *denote the ordered eigenvalues of* $\mathbf{R}$ *and* $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$*.*

This is obtained from the joint distribution of Wishart matrices $\mathbf{X}\mathbf{X}^\mathsf{H}$ which, up to a variables change, leads to the joint distribution of the couples $(\mathbf{U}, \mathbf{\Lambda})$ of unitary matrices and diagonal eigenvalue matrices such that $\mathbf{X}\mathbf{X}^\mathsf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\mathsf{H}$. In performing this variable change, the Jacobian $\Delta(\mathbf{\Lambda})^2$ arises. Integrating over $\mathbf{U}$ to obtain the marginal distribution of $\mathbf{\Lambda}$, we recognize the Harish-Chandra equality which finally leads to the result.

These results are of practical interest in several domains of research which have to deals with vectorial noise observations whose sample covariance matrix is given by such a matrix $\mathbf{X}\mathbf{X}^\mathsf{H}$. In the specific case of wireless communications, these expressions can be used to determine for instance the ergodic mutual information of a multi-antenna channel, modelled by a matrix $\mathbf{H} \in \mathbb{C}^{N\times n}$ with independent and identically distributed (i.i.d.) Gaussian entries or with correlated

columns. In this case, we indeed have that the ergodic mutual information $\mathcal{I}(\sigma^2)$ of a multi-antenna channel with additive Gaussian noise σ^2 reads [10]

$$\mathcal{I}(\sigma^2) = \mathrm{E}\left[\log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)\right],$$

where σ^{-2} denotes the signal-to-noise ratio (SNR) at the receiver and the expectation is taken over the realizations of the random channel $\mathbf{H}$, varying according to the (correlated or uncorrelated) Gaussian distribution. Denoting $\mathbf{H}\mathbf{H}^{\mathsf{H}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathsf{H}}$ the spectral decomposition of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$, this can be rewritten

$$\begin{aligned}\mathcal{I}(\sigma^2) &= \mathrm{E}\left[\log_2\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{\Lambda}\right)\right] \\ &= \mathrm{E}\left[\sum_{i=1}^{N}\log_2\left(1+\frac{\lambda_i}{\sigma^2}\right)\right] \\ &= \int\cdots\int\sum_{i=1}^{N}\log_2\left(1+\frac{\lambda_i}{\sigma^2}\right)dP_{(\lambda_i)}(\lambda_1,\ldots,\lambda_N),\end{aligned}$$

which can be evaluated from the theorem above.

These are the tools we need for the study of Wishart matrices. As it appears, the above properties hold due to the rotational invariance of Gaussian matrices. For more involved random matrix models, e.g., when the entries of the random matrices under study are no longer Gaussian, the study of the eigenvalue distribution is much more involved, if not unfeasible.

However, it turns out that, as the matrix dimensions grow large, nice properties arise that can be studied much more efficiently than when the matrix sizes are kept fixed. A short introduction to these large matrix considerations is described hereafter.

0.2.2 Limiting spectral distribution

Consider an $N\times N$ (non-necessarily random) Hermitian matrix $\mathbf{X}_N$. Define its *empirical spectral distribution* (e.s.d.) $F^{\mathbf{X}_N}$ to be the distribution function of the eigenvalues of $\mathbf{X}_N$, i.e., for $x\in\mathbb{R}$,

$$F^{\mathbf{X}_N}(x) = \frac{1}{N}\sum_{j=1}^{N}1_{\lambda_j\leq x}(x),$$

where $\lambda_1,\ldots,\lambda_N$ are the eigenvalues of $\mathbf{X}_N$.[3]

The relevant aspect of large $N\times N$ Hermitian matrices $\mathbf{X}_N$ is that their (random) e.s.d. $F^{\mathbf{X}_N}$ often converges, with $N\to\infty$, towards a non-random distribution F. This function F, if it exists, will be called the *limit spectral distribution* (l.s.d.) of $\mathbf{X}_N$. Weak convergence [12] of $F^{\mathbf{X}_N}$ to F, i.e., for all x where F is continuous, $F^{\mathbf{X}_N}(x)-F(x)\to 0$, is often sufficient to obtain relevant results; this is denoted

$$F^{\mathbf{X}_N}\Rightarrow F.$$

In most cases though, the weak convergence of $F^{\mathbf{X}_N}$ to F will only be true on a set of matrices $\mathbf{X}_N = \mathbf{X}_N(\omega)$ of measure one. This will be mentioned with the phrase $F^{\mathbf{X}_N}\Rightarrow F$ *almost surely.*

[3] the Hermitian property is fundamental to ensure that all eigenvalues of $\mathbf{X}_N$ belong to the real line. However, the extension of the e.s.d. to non-Hermitian matrices is sometimes requires; for a definition, see (1.2.2) of [11].

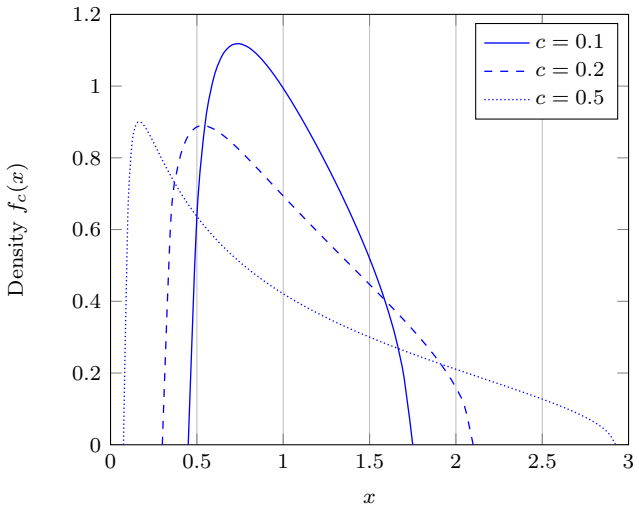

Figure 1: Marčenko-Pastur law for different limit ratios $c = \lim N/n$.

The Marčenko-Pastur law

In signal processing, one is often interested in sample covariance matrices or even more general matrices such as i.i.d. matrices with left and right correlation, or i.i.d. matrices with a variance profile. One of the best known result with a large range of applications in signal processing is the convergence of the e.s.d. of the Gram matrix of a random matrix with i.i.d. entries of zero mean and normalized variance (not necessarily a Wishart matrix). This result is due to Marčenko and Pastur [13], so that the limiting e.s.d. of the Gram matrix is called the *Marčenko-Pastur law*. The result unfolds as follows.

Theorem 0.5. *Consider a matrix* $\mathbf{X} \in \mathbb{C}^{N\times n}$ *with i.i.d. entries* $\left(\frac{1}{\sqrt{n}}X_{ij}^{(N)}\right)$ *such that* $X_{11}^{(N)}$ *has zero-mean and variance* 1. *As* $n, N \to \infty$ *with* $\frac{N}{n} \to c \in (0,\infty)$, *the e.s.d. of* $\mathbf{R}_n = \mathbf{X}\mathbf{X}^{\mathsf{H}}$ *converges almost surely to a nonrandom distribution function* F_c *with density* f_c *given by*

$$f_c(x) = (1 - c^{-1})^+\delta(x) + \frac{1}{2\pi cx}\sqrt{(x-a)^+(b-x)^+}, \tag{2}$$

where $a = (1-\sqrt{c})^2$, $b = (1+\sqrt{c})^2$ *and* $\delta(x) = 1_{\{0\}}(x)$.

The d.f. F_c is named the Marčenko-Pastur law with limiting ratio c. This is depicted in Figure 1 for different values of the limiting ratio c. Notice in particular that, when c tends to be small and approaches zero, the Marčenko-Pastur law reduces to a single mass in 1, as the law of large numbers in classical probability theory requires.

Several approaches can be used to derive the Marčenko-Pastur law. However, the original technique proposed by Marčenko and Pastur is based on a fundamental tool, the *Stieltjes transform*, which will be constantly used in this document. In the following we present the Stieltjes transform, along with a few important lemmas, before we introduce several applications based on the Stieltjes transform method.

The Stieltjes transform and associated lemmas

Definition 0.3. *Let F be a real-valued bounded measurable function over $\mathbb{R}$. Then the Stieltjes transform $m_F(z)$,[4] for $z \in \mathrm{Supp}(F)^c$, the complex space complementary to the support of F,[5] is defined as*

$$m_F(z) \triangleq \int_{-\infty}^{\infty} \frac{1}{\lambda - z} dF(\lambda). \tag{3}$$

For all F that admit a Stieltjes transform, the inverse transformation exists and formulates as follows,

Theorem 0.6. *If x is a continuity points of F, then*

$$F(x) = \frac{1}{\pi} \lim_{y \to 0^+} \int_{-\infty}^{x} \Im \left[m_F(x + iy) \right] dx. \tag{4}$$

In practice here, F will be a distribution function. Therefore, there exists an intimate link between distribution functions and their Stieltjes transforms. More precisely, if F_1 and F_2 are two distribution functions (therefore right-continuous by definition, see e.g., Section 14 of [14]) that have the same Stieltjes transform, then F_1 and F_2 coincide everywhere and the converse is true. As a consequence, m_F uniquely determines F and vice-versa. It will turn out that, while working on the distribution functions of the empirical eigenvalues of large random matrices is often a tedious task, the approach via Stieltjes transforms greatly simplifies the study. The initial intuition behind the Stieltjes transform approach for random matrices lies in the following remark: for an Hermitian matrix $\mathbf{X} \in \mathbb{C}^{N \times N}$,

$$\begin{aligned} m_{F^{\mathbf{X}}}(z) &= \int \frac{1}{\lambda - z} dF^{\mathbf{X}}(\lambda) \\ &= \frac{1}{N} \mathrm{tr} \left(\mathbf{\Lambda} - z \mathbf{I}_N \right)^{-1} \\ &= \frac{1}{N} \mathrm{tr} \left(\mathbf{X} - z \mathbf{I}_N \right)^{-1}, \end{aligned}$$

in which we denoted $\mathbf{\Lambda}$ the diagonal matrix of eigenvalues of $\mathbf{X}$. Working with the Stieltjes transform of $F^{\mathbf{X}}$ therefore boils down to working with the matrix $(\mathbf{X} - z\mathbf{I}_N)^{-1}$, and more specifically with the sum of its diagonal entries. From matrix inversion lemmas and several fundamental matrix identities, it is then rather simple to derive limits of traces $\frac{1}{N}\mathrm{tr}\left(\mathbf{X} - z\mathbf{I}_N\right)^{-1}$, as N grows large, hence the Stieltjes transform of the weak limit of $F^{\mathbf{X}}$. For notational simplicity, we may denote $m_{\mathbf{X}} \triangleq m_{F^{\mathbf{X}}}$ the Stieltjes transform of the e.s.d. of the Hermitian matrix $\mathbf{X}$, and call $m_{\mathbf{X}}$ the *Stieltjes transform of* $\mathbf{X}$.

An identity of particular interest is the relation between the Stieltjes transform of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$ and $\mathbf{X}^{\mathsf{H}}\mathbf{X}$, for $\mathbf{X} \in \mathbb{C}^{N \times n}$. Note that both matrices are Hermitian, and actually nonnegative definite, so that the Stieltjes transform of both is well defined.

Lemma 0.1. *For $z \in \mathbb{C} \setminus \mathbb{R}^+$, we have*

$$\frac{n}{N} m_{F^{\mathbf{X}^{\mathsf{H}}\mathbf{X}}}(z) = m_{F^{\mathbf{X}\mathbf{X}^{\mathsf{H}}}}(z) + \frac{N-n}{N} \frac{1}{z}.$$

[4] we borrow here the notation m to a large number of contributions from Bai, Silverstein et al. In other works, the notation s or S for the Stieltjes transform is used.

[5] we recall that the support $\mathrm{Supp}(F)$ of a real function F is the set $\{x \in \mathbb{R}, |F(x)| > 0\}$.

On the wireless communication side, it turns out that the Stieltjes transform is directly connected to the expression of the mutual information, through the so-called *Shannon transform*, initially coined by Tulino and Verdù, see Section 2.3.3 of [15].

Definition 0.4. *Let F be a probability distribution defined on $\mathbb{R}^+$. The Shannon-transform $\mathcal{V}_F$ of F is defined, for $x \in \mathbb{R}^+$, as*

$$\mathcal{V}_F(x) \triangleq \int_0^\infty \log(1 + x\lambda) dF(\lambda). \tag{5}$$

The Shannon-transform of F is related to its Stieltjes transform m_F through the expression

$$\mathcal{V}_F(x) = \int_{\frac{1}{x}}^\infty \left(\frac{1}{t} - m_F(-t) \right) dt. \tag{6}$$

This last relation is fundamental to derive a link between the l.s.d. of a random matrix and the mutual information of a multi-dimensional channel, whose model is based on this random matrix.

We complete this section by the introduction of fundamental lemmas, required to derive the l.s.d. of random matrix models with independent entries, among which the Marčenko-Pastur law, and that will be necessary to the derivation of deterministic equivalents. These are recalled briefly below.

The first lemma is called the *trace lemma*, introduced in [16] (and extended in [17] under the form of a central limit theorem), that we formulate in the following theorem,

Theorem 0.7. *Let $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_N \in \mathbb{C}^{N \times N}$, be a series of matrices with uniformly bounded spectral norm. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots$ be random vectors of i.i.d. entries such that $\mathbf{x}_N \in \mathbb{C}^N$ has zero mean, variance $1/N$ and finite eighth order moment, independent of $\mathbf{A}_N$. Then*

$$\mathbf{x}_N^{\mathsf{H}} \mathbf{A}_N \mathbf{x}_N - \frac{1}{N} \mathrm{tr} \mathbf{A}_N \xrightarrow{\text{a.s.}} 0, \tag{7}$$

as $N \to \infty$.

Many versions of this result exist in the literature, that can be adapted to different application needs. We mention in particular that,

- in [18], it is shown that, when restricting the entries of $\mathbf{x}_N$ to be bounded by $\log N$, the convergence holds true without the need of the existence of an eighth order moment. This observation will be needed, along with the so-called truncation, centralization and rescaling steps, to alleviate all moment assumptions on $\mathbf{x}_N$, when deriving deterministic equivalents later in this chapter.
- in [19], we show that the above result also holds true when $\mathbf{A}_N$ is not uniformly bounded in spectral norm but is such that its largest eigenvalue is almost surely bounded for all large N; the bound in that case does not need to be uniform over the probability space generating the random $\mathbf{A}_N$ matrices.

The second important ingredient is the rank-1 perturbation lemma, given below

Theorem 0.8. *(i) Let $z \in \mathbb{C} \setminus \mathbb{R}$, $\mathbf{A} \in \mathbb{C}^{N\times N}$, $\mathbf{B} \in \mathbb{C}^{N\times N}$ with $\mathbf{B}$ Hermitian, and $\mathbf{v} \in \mathbb{C}^N$. Then*

$$\left| \frac{1}{N} \mathrm{tr}\mathbf{A} \left((\mathbf{B} - z\mathbf{I}_N)^{-1} - (\mathbf{B} + \mathbf{v}\mathbf{v}^{\mathsf{H}} - z\mathbf{I}_N)^{-1} \right) \right| \leq \frac{\|\mathbf{A}\|}{N|\Im[z]|},$$

with $\|\mathbf{A}\|$ the spectral norm of $\mathbf{A}$.

(ii) Moreover, if $\mathbf{B}$ is nonnegative definite, for $z \in \mathbb{R}^-$,

$$\left| \frac{1}{N} \mathrm{tr}\mathbf{A} \left((\mathbf{B} - z\mathbf{I}_N)^{-1} - (\mathbf{B} + \mathbf{v}\mathbf{v}^{\mathsf{H}} - z\mathbf{I}_N)^{-1} \right) \right| \leq \frac{\|\mathbf{A}\|}{N|z|}.$$

Again, generalizations of the above result can be found e.g., in [19], where we prove that

$$\frac{1}{N} \mathrm{tr}\mathbf{A}\mathbf{B}^{-1} - \frac{1}{N} \mathrm{tr}\mathbf{A}(\mathbf{B} + \mathbf{v}\mathbf{v}^{\mathsf{H}})^{-1} \xrightarrow{\text{a.s.}} 0,$$

as $N \to \infty$, whenever there exists $\varepsilon > 0$ such that the smallest eigenvalue of $\mathbf{B}$ is almost surely greater than ε for all large N (the existence of $\mathbf{B}^{-1}$ and $(\mathbf{B} + \mathbf{v}\mathbf{v}^{\mathsf{H}})^{-1}$ being almost sure in such a case).

Based on the above ingredients and classical results from probability theory, it is possible to prove the almost sure weak convergence of the e.s.d. of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, where $\mathbf{X} \in \mathbb{C}^{N\times n}$ has i.i.d. entries of zero mean and variance $1/n$, to the Marčenko-Pastur law, as well as the convergence of the e.s.d. of more involved random matrix models based on matrices with independent entries. In particular, we will be interested in Section 0.5.2 in limiting results on the e.s.d. of sample covariance matrices.

l.s.d. of sample covariance matrices

The limiting spectral distribution of the sample covariance matrix unfolds from the following result, originally provided by Bai and Silverstein in [18], and further extended in e.g., [11],

Theorem 0.9. *Consider the matrix $\mathbf{B}_N = \mathbf{A}_N + \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N \mathbf{X}_N \in \mathbb{C}^{n\times n}$, where $\mathbf{X}_N = \left(\frac{1}{\sqrt{n}} X_{ij}^N \right) \in \mathbb{C}^{N\times n}$ with entries X_{ij}^N independent with zero mean, variance 1 and finite order $2+\varepsilon$ moment for some $\varepsilon > 0$ (ε is independent of N, i, j), the e.s.d. $F^{\mathbf{T}_N}$ of $\mathbf{T}_N = \mathrm{diag}(t_1^N, \ldots, t_N^N) \in \mathbb{R}^{N\times N}$ converges weakly and almost surely to F^T, $\mathbf{A}_N$ is $n \times n$ Hermitian whose e.s.d. converges weakly and almost surely to F^A, N/n tends to c, with $0 < c < \infty$ as n, N grow large. Then, the e.s.d. $F^{\mathbf{B}_N}$ of $\mathbf{B}_N$ converges weakly and almost surely to F^B such that, for $z \in \mathbb{C}^+$, $m_{F^B}(z)$ satisfies*

$$m_{F^B}(z) = m_{F^A} \left(z - c \int \frac{t}{1 + t m_{F^B}(z)} dF^T(t) \right). \tag{8}$$

The solution of the implicit equation (8) in the dummy variable $m_{F^B}(z)$ is unique on the set $\{z \in \mathbb{C}^+, m_{F^B}(z) \in \mathbb{C}^+\}$. Moreover, if the $\mathbf{X}_N$ has identically distributed entries, then the result holds without requiring that a moment of order $2+\varepsilon$ exists.

In the following, using the tools from the previous sections, we give a sketch of the proof of Theorem 0.9.

Proof. The fundamental idea to infer the final formula of Theorem 0.9 is to first guess the form it should take. For this, write

$$m_{F^{\mathbf{B}_N}}(z) \triangleq \frac{1}{n}\mathrm{tr}(\mathbf{A}_N + \mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N - z\mathbf{I}_N)^{-1}$$

and take $\mathbf{D}_N \in \mathbb{C}^{N\times N}$ to be some deterministic matrix such that we would like

$$m_{F^{\mathbf{B}_N}}(z) - m_N(z) \xrightarrow{\text{a.s.}} 0$$

with

$$m_N(z) \triangleq \frac{1}{n}\mathrm{tr}(\mathbf{A}_N + \mathbf{D}_N - z\mathbf{I}_N)^{-1}$$

as $N, n \to \infty$ with $N/n \to c$. We then have, from the identity $\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{A}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{B}^{-1}$,

$$m_{F^{\mathbf{B}_N}}(z) - m_N(z) = \frac{1}{n}\mathrm{tr}\left[(\mathbf{B}_N - z\mathbf{I}_N)^{-1}(\mathbf{D}_N - \mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N)(\mathbf{A}_N + \mathbf{D}_N - z\mathbf{I}_N)^{-1}\right].$$

Taking $\mathbf{D}_N = a_N\mathbf{I}_N$, and writing

$$\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N = \sum_{k=1}^{N} t_k^N \mathbf{x}_k\mathbf{x}_k^{\mathsf{H}}$$

with $\mathbf{x}_k$ the k^{th} column of $\mathbf{X}_N^{\mathsf{H}}$, we further have

$$\begin{aligned} m_{F^{\mathbf{B}_N}}(z) - m_N(z) = {} & \frac{a_N}{n}\mathrm{tr}\left[(\mathbf{B}_N - z\mathbf{I}_N)^{-1}(\mathbf{A}_N + \mathbf{D}_N - z\mathbf{I}_N)^{-1}\right] \\ & - \frac{1}{n}\sum_{k=1}^{N} t_k^N \mathbf{x}_k^{\mathsf{H}}(\mathbf{A}_N + \mathbf{D}_N - z\mathbf{I}_N)^{-1}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{x}_k. \end{aligned}$$

Using the matrix inversion identity

$$(\mathbf{A} + \mathbf{v}\mathbf{v}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{v} = \frac{1}{1 + \mathbf{v}^{\mathsf{H}}(\mathbf{A} - z\mathbf{I}_N)^{-1}\mathbf{v}}(\mathbf{A} - z\mathbf{I}_N)^{-1}\mathbf{v},$$

each term in the sum of the right-hand side can further be expressed as

$$t_k^N \mathbf{x}_k^{\mathsf{H}}(\mathbf{A}_N + \mathbf{D}_N - z\mathbf{I}_N)^{-1}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{x}_k = \frac{t_k^N \mathbf{x}_k^{\mathsf{H}}(\mathbf{A}_N + \mathbf{D}_N - z\mathbf{I}_N)^{-1}(\mathbf{B}_{(k)} - z\mathbf{I}_N)^{-1}\mathbf{x}_k}{1 + t_k^N \mathbf{x}_k^{\mathsf{H}}(\mathbf{B}_{(k)} - z\mathbf{I}_N)^{-1}\mathbf{x}_k}$$

where $\mathbf{B}_{(k)} = \mathbf{B}_N - t_k^N \mathbf{x}_k\mathbf{x}_k^{\mathsf{H}}$ and where now $\mathbf{x}_k$ and $(\mathbf{A}_N + \mathbf{D}_N - z\mathbf{I}_N)^{-1}(\mathbf{B}_{(k)} - z\mathbf{I}_N)^{-1}$ are independent. But then, using the trace lemma, Theorem 0.7, we have that

$$\mathbf{x}_k^{\mathsf{H}}(\mathbf{A}_N + \mathbf{D}_N - z\mathbf{I}_N)^{-1}(\mathbf{B}_{(k)} - z\mathbf{I}_N)^{-1}\mathbf{x}_k - \frac{1}{n}\mathrm{tr}(\mathbf{A}_N + \mathbf{D}_N - z\mathbf{I}_N)^{-1}(\mathbf{B}_{(k)} - z\mathbf{I}_N)^{-1} \xrightarrow{\text{a.s.}} 0.$$

Replacing the quadratic form by the trace in the Stieltjes transform difference, we then have for all large N,

$$\begin{aligned} m_{F^{\mathbf{B}_N}}(z) - m_N(z) \simeq {} & \frac{a_N}{n}\mathrm{tr}\left[(\mathbf{B}_N - z\mathbf{I}_N)^{-1}(\mathbf{A}_N + \mathbf{D}_N - z\mathbf{I}_N)^{-1}\right] \\ & - \frac{1}{n}\sum_{k=1}^{N} \frac{t_k^N \frac{1}{n}\mathrm{tr}(\mathbf{A}_N + \mathbf{D}_N - z\mathbf{I}_N)^{-1}(\mathbf{B}_{(k)} - z\mathbf{I}_N)^{-1}}{1 + t_k^N \frac{1}{n}\mathrm{tr}(\mathbf{B}_{(k)} - z\mathbf{I}_N)^{-1}}. \end{aligned}$$

But then, from the rank-1 perturbation lemma, Theorem 0.8, this is further approximated, for all large N by

$$m_{F^{\mathbf{B}_N}}(z) - m_N(z) \simeq \frac{a_N}{n}\mathrm{tr}\left[(\mathbf{B}_N - z\mathbf{I}_N)^{-1}(\mathbf{A}_N + \mathbf{D}_N - z\mathbf{I}_N)^{-1}\right] - \frac{1}{n}\sum_{k=1}^{N}\frac{t_k^N \frac{1}{n}\mathrm{tr}(\mathbf{A}_N + \mathbf{D}_N - z\mathbf{I}_N)^{-1}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}}{1 + t_k^N \frac{1}{n}\mathrm{tr}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}}$$

where we recognize in the right-hand side the Stieltjes transform $m_{F^{\mathbf{B}_N}}(z) = \frac{1}{n}\mathrm{tr}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$. Taking

$$a_N = \frac{1}{n}\sum_{k=1}^{N} t_k^N \frac{1}{1 + t_k^N m_{F^{\mathbf{B}_N}}(z)} \simeq c\int \frac{t}{1 + t m_{F^{\mathbf{B}_N}}(z)} dF^T(t),$$

it is clear that the difference $m_{F^{\mathbf{B}_N}}(z) - m_N(z)$ becomes increasingly small for large N and therefore $m_{F^{\mathbf{B}_N}}(z)$ is asymptotically close to

$$\frac{1}{n}\mathrm{tr}\left(\mathbf{A}_N + c\int \frac{t dF^T(t)}{1 + t m_{F^{\mathbf{B}_N}}(z)}\mathbf{I}_N - z\mathbf{I}_N\right)^{-1}$$

which is exactly

$$m_{F^{\mathbf{A}_N}}\left(z - c\int \frac{t dF^T(t)}{1 + t m_{F^{\mathbf{B}_N}}(z)}\right).$$

Hence the result. □

The sample covariance matrix model corresponds to the particular case where $\mathbf{A}_N = 0$. In that case, (8) becomes

$$m_{\underline{F}}(z) = -\left(z - c\int \frac{t}{1 + t m_{\underline{F}}(z)} dF^T(t)\right)^{-1}, \tag{9}$$

where we denoted $\underline{F} \triangleq F^B$ in this special case. This special notation will often be used to differentiate the l.s.d. F of the matrix $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ from the l.s.d. $\underline{F}$ of the reversed Gram matrix $\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N$. Remark indeed from Lemma 0.1 that the Stieltjes transform $m_{\underline{F}}$ of the l.s.d. $\underline{F}$ of $\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N$ is linked to the Stieltjes transform m_F of the l.s.d. F of $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ through

$$m_{\underline{F}}(z) = c m_F(z) + (c-1)\frac{1}{z} \tag{10}$$

and then we also have access to a characterization of F, which is exactly the asymptotic eigenvalue distribution of the sample covariance matrix model, when the denormalized columns $\sqrt{n}\mathbf{x}_1, \ldots, \sqrt{n}\mathbf{x}_n$ of $\sqrt{n}\mathbf{X}_N$ form a sequence of independent vectors with zero mean and covariance matrix $n\mathrm{E}[\mathbf{x}_1\mathbf{x}_1^{\mathsf{H}}] = \mathbf{T}_N$.

Secondly, in addition to the uniqueness of the pair $(z, m_{\underline{F}}(z))$ in the set $\{z \in \mathbb{C}^+, m_{\underline{F}}(z) \in \mathbb{C}^+\}$ solution of (9), an inverse formula for the Stieltjes transform can be written in closed-form, i.e., we can define a function $z_{\underline{F}}(m)$ on $\{\underline{m} \in \mathbb{C}^+, z_{\underline{F}}(\underline{m}) \in \mathbb{C}^+\}$, such that

$$z_{\underline{F}}(\underline{m}) = -\frac{1}{\underline{m}} + c\int \frac{t}{1 + t\underline{m}} dF^T(t). \tag{11}$$

This will turn out to be extremely useful to characterize the spectrum of F. More on this topic is discussed in Section 0.3.

0.3 Spectral analysis

In this section, we summarize some important results regarding (i) the characterization of the support of the eigenvalues of a sample covariance matrix and (ii) the position of the individual eigenvalues of a sample covariance matrix. The point (i) is obviously a must-have on a pure mathematical viewpoint but is also fundamental to the study of estimators based on large dimensional random matrices. Typically, we will provide in Section 0.4 and in Section 0.5.2 estimators of functionals of the eigenvalues of a population covariance matrix based on the observation of a sample covariance matrix. We will in particular investigate large dimensional sample covariance matrix models with population covariance matrix composed of a few eigenvalues with large multiplicities. The validity of these estimators relies importantly on the fact that the support of the l.s.d. of the sample covariance matrix is formed of disjoint so-called *clusters*, each cluster being associated to one of the few eigenvalues of the population covariance matrix. Characterizing the limiting support is therefore paramount to the study of the estimator performance. The point (ii) is even more important for the estimators described above as knowing the position of the individual eigenvalues allows one to derive such estimators. This second point is also fundamental to the derivation of hypothesis tests based on large dimensional matrix analysis, that will be introduced in Section 0.5.1. What we will show in particular is that, under mild assumptions on the random matrix model, all eigenvalues are asymptotically contained *within* the limiting support. Also, when the limiting support is divided into disjoint clusters, the number of sample eigenvalues in each cluster corresponds exactly to the multiplicity of the population eigenvalue attached to this cluster. For signal sensing, this is fundamental as the observation of a sample eigenvalue outside the expected limiting support of the pure noise hypothesis (called hypothesis $\mathcal{H}_0$) suggests that a signal is present in the observed data.

We start with the point (ii).

0.3.1 Exact eigenvalue separation

The results of interest here are due to Bai and Silverstein and are summarized in the following theorems.

Theorem 0.10 ([16]). *Let $\mathbf{X}_N = \left(\frac{1}{\sqrt{n}} X_{ij}^N\right) \in \mathbb{C}^{N\times n}$ have i.i.d. entries, such that X_{11}^N has zero mean, variance 1 and finite fourth order moment. Let $\mathbf{T}_N \in \mathbb{C}^{N\times N}$ be nonrandom, whose e.s.d. $F^{\mathbf{T}_N}$ converge weakly to H. From Theorem 0.9, the e.s.d. of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}} \in \mathbb{C}^{N\times N}$ converges weakly and almost surely towards some distribution function F, as N, n go to infinity with ratio $c_N = N/n \to c$, $0 < c < \infty$. Similarly, the e.s.d. of $\underline{\mathbf{B}}_N = \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N \mathbf{X}_N \in \mathbb{C}^{n\times n}$ converges towards $\underline{F}$ given by*

$$\underline{F}(x) = cF(x) + (1-c)1_{[0,\infty)}(x).$$

Denote $\underline{F}_N$ the distribution of Stieltjes transform $m_{\underline{F}_N}(z)$, solution, for $z \in \mathbb{C}^+$, of the following equation in m

$$m = -\left(z - \frac{N}{n}\int \frac{\tau}{1+\tau m} dF^{\mathbf{T}_N}(\tau)\right)^{-1},$$

and define F_N the d.f. such that

$$\underline{F}_N(x) = \frac{N}{n} F_N(x) + \left(1 - \frac{N}{n}\right) 1_{[0,\infty)}(x).$$

Let $N_0 \in \mathbb{N}$, and choose an interval $[a,b]$, $a>0$, outside the union of the supports of F and F_N for all $N \geq N_0$. For $\omega \in \Omega$, the random space generating the series $\mathbf{X}_1, \mathbf{X}_2, \ldots$, denote $\mathcal{L}_N(\omega)$ the set of eigenvalues of $\mathbf{B}_N(\omega)$. Then,

$$P(\omega, \mathcal{L}_N(\omega) \cap [a,b] \neq \emptyset, \text{ i.o.}) = 0.$$

This means concretely that, given a segment $[a,b]$ outside the union of the supports of F and $F_{N_0}, F_{N_0+1}, \ldots$, for all series $\mathbf{B}_1(\omega), \mathbf{B}_2(\omega), \ldots$, with ω in some set of probability one, there exists $M(\omega)$ such that, for all $N \geq M(\omega)$, there will be no eigenvalue of $\mathbf{B}_N(\omega)$ in $[a,b]$.

As an immediate corollary of Theorems 0.5 and 0.10, we have the following results on the extreme eigenvalues of $\mathbf{B}_N$, with $\mathbf{T}_N = \mathbf{I}_N$.

Corollary 0.1. *Let $\mathbf{B}_N \in \mathbb{C}^{N\times N}$ be defined as $\mathbf{B}_N = \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}}$, with $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ with i.i.d. entries of zero mean, variance $1/n$ and finite fourth order moment. Then, denoting $\lambda_{\min}^N$ and $\lambda_{\max}^N$ the smallest and largest eigenvalues of $\mathbf{B}_N$, respectively, we have*

$$\lambda_{\min}^N \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2$$
$$\lambda_{\max}^N \xrightarrow{\text{a.s.}} (1+\sqrt{c})^2$$

as $N, n \to \infty$ with $N/n \to c$.

This result further extends to the case when $\mathbf{B}_N = \mathbf{X}_N \mathbf{T}_N \mathbf{X}_N^{\mathsf{H}}$, with $\mathbf{T}_N$ diagonal with ones on the diagonal but for a few entries different from one. This model, often referred to as *spiked model* lets some eigenvalues escape the limiting support of $\mathbf{B}_N$ (which is still the support of the Marčenko-Pastur law). Note that this is not inconsistent with Theorem 0.10 since here, for all finite N_0, the distribution functions $F_{N_0}, F_{N_0+1}, \ldots$ may all have a non-zero mass outside the support of the Marčenko-Pastur law. The segments $[a,b]$ where no eigenvalues are found asymptotically must be away from these potential masses. The theorem, due to Baik, is given precisely as follows

Theorem 0.11 ([20])**.** *Let $\bar{\mathbf{B}}_N = \bar{\mathbf{T}}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \bar{\mathbf{T}}_N^{\frac{1}{2}}$, where $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ has i.i.d. entries of zero mean and variance $1/n$, and $\bar{\mathbf{T}}_N \in \mathbb{R}^{N\times N}$ is diagonal given by*

$$\bar{\mathbf{T}}_N = \text{diag}(\underbrace{\alpha_1, \ldots, \alpha_1}_{k_1}, \ldots, \underbrace{\alpha_M, \ldots, \alpha_M}_{k_M}, \underbrace{1, \ldots, 1}_{N-\sum_{i=1}^M k_i})$$

with $\alpha_1 > \ldots > \alpha_M > 0$ for some positive integer M. We denote here $c = \lim_N N/n$. Call $M_0 = \#\{j | \alpha_j > 1+\sqrt{c}\}$. For $c<1$, take also M_1 to be such that $M - M_1 = \#\{j | \alpha_j < 1-\sqrt{c}\}$. Denote additionally $\lambda_1, \ldots, \lambda_N$ the eigenvalues of $\bar{\mathbf{B}}_N$, ordered as $\lambda_1 \geq \ldots \geq \lambda_N$. We then have

- *for $1 \leq j \leq M_0$, $1 \leq i \leq k_j$,*

$$\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} \alpha_j + \frac{c\alpha_j}{\alpha_j - 1},$$

- *for the other eigenvalues, we must discriminate upon c,*
  - *if $c<1$,*

* *for* $M_1 + 1 \leq j \leq M$, $1 \leq i \leq k_j$,

$$\lambda_{N-k_j-\ldots-k_M+i} \xrightarrow{\text{a.s.}} \alpha_j + \frac{c\alpha_j}{\alpha_j - 1},$$

* *for the indexes of eigenvalues of* $\bar{\mathbf{T}}_N$ *inside* $[1-\sqrt{c}, 1+\sqrt{c}]$,

$$\lambda_{k_1+\ldots+k_{M_0}+1} \xrightarrow{\text{a.s.}} (1+\sqrt{c})^2,$$
$$\lambda_{N-k_{M_1+1}-\ldots-k_M} \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2,$$

– *if* $c > 1$,

$$\lambda_n \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2,$$
$$\lambda_{n+1} = \ldots = \lambda_N = 0,$$

– *if* $c = 1$,

$$\lambda_{\min(n,N)} \xrightarrow{\text{a.s.}} 0.$$

The important part of this result for us is that all α_j such that $\alpha_j > 1 + \sqrt{c}$ produces an eigenvalue of $\mathbf{B}_N$ outside the support of the Marčenko-Pastur, found asymptotically at the position $\alpha_j + \frac{c\alpha_j}{\alpha_j - 1}$.

Now Theorem 0.10 and Theorem 0.11 ensure that, for a given N_0, no eigenvalue of $\mathbf{B}_N$ is found outside the support of $F_{N_0}, F_{N_0+1}, \ldots$ for all large N, but do not say where the eigenvalues of $\mathbf{B}_N$ are approximately positioned. The answer to this question is provided by Bai and Silverstein in [21] in which the exact separation properties of the l.s.d. of such matrices $\mathbf{B}_N$ is discussed.

Theorem 0.12 ([21])**.** *Assume* $\mathbf{B}_N$ *is as in Theorem 0.10 with* $\mathbf{T}_N$ *nonnegative definite and* $F^{\mathbf{T}_N}$ *converging weakly to the distribution function* H*, and* $c_N = N/n$ *converging to* c*. Consider also* $0 < a < b < \infty$ *such that* $[a, b]$ *lies outside the support of* F*, the l.s.d. of* $\mathbf{B}_N$*. Denote additionally* λ_k *and* τ_k *the* k^{th} *eigenvalues of* $\mathbf{B}_N$ *and* $\mathbf{T}_N$ *in decreasing order, respectively. Then we have*

1. *If* $c(1 - H(0)) > 1$*, then the smallest eigenvalue* x_0 *of the support of* F *is positive and* $\lambda_N \to x_0$ *almost surely, as* $N \to \infty$*.*

2. *If* $c(1 - H(0)) \leq 1$*, or* $c(1 - H(0)) > 1$ *but* $[a, b]$ *is not contained in* $[0, x_0]$*, then*

$$P(\omega, \lambda_{i_N} > b, \lambda_{i_N+1} < a) = 1,$$

for all N *large, where* i_N *is the unique integer such that*

$$\tau_{i_N} > -1/m_F(b),$$
$$\tau_{i_N+1} < -1/m_F(a).$$

Theorem 0.12 states in particular that, when the limiting spectrum can be divided in disjoint clusters, then the index of the sample eigenvalue for which a jump from one cluster (right

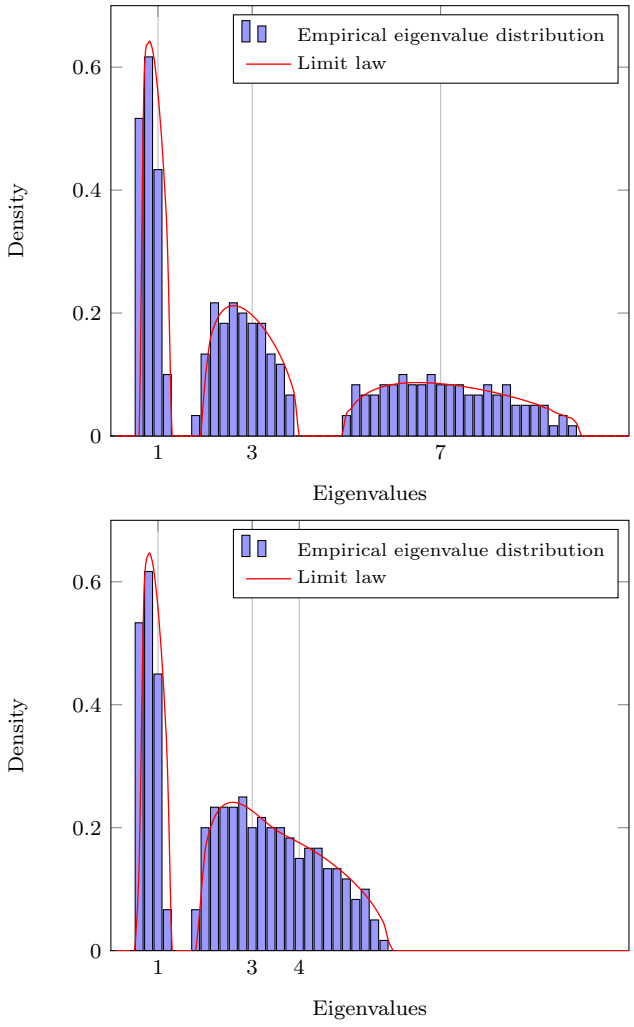

Figure 2: Histogram of the eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, $N = 300$, $n = 3000$, with $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in (i) 1, 3 and 7 on top, (ii) 1, 3 and 4 at bottom.

to b) to a subsequent cluster (left to a) arises corresponds exactly to the index of the population eigenvalue where a jump arises in the population eigenvalue spectrum (from $-1/m_F(b)$ to $-1/m_F(a)$). Therefore, the sample eigenvalues distribute as one would expect between the consecutive clusters. This result will be used in Section 0.4 and Section 0.5.2 to find which sample eigenvalues are present in which cluster. This is necessary because we will perform complex integration on contours surrounding specific clusters and that residue calculus will demand that we know exactly what eigenvalues are found inside these contours.

Nonetheless, this still does not exactly answer the question of the exact characterization of the limiting support, which we treat in the following.

0.3.2 Support of l.s.d.

Remember from the inverse Stieltjes transform formula (4) that it is possible to determine the support of the l.s.d. F of a random matrix once we know its limiting Stieltjes transform $m_F(z)$ for all $z \in \mathbb{C}^+$. Thanks to Theorem 0.9, we know in particular that we can determine the support of the l.s.d. of a sample covariance matrix. Nonetheless, (4) features a limit for the imaginary part y of the argument $z = x + iy$ of $m_F(z)$ going to zero, which has not been characterized to this point (even its existence everywhere is not ensured). Choi and Silverstein proved in [22] that this limit does exist for the case of sample covariance matrices and goes even further in characterizing exactly what this limit is. This uses the important Stieltjes transform composition inverse formula (11) and is summarized as follows.

Theorem 0.13 ([22])**.** *Denote S_X^c the complementary of S_X, the support of some d.f. X. Let $\underline{\mathbf{B}}_N = \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N \mathbf{X}_N \in \mathbb{C}^{n \times n}$ have l.s.d. $\underline{F}$, where $\mathbf{X}_N \in \mathbb{C}^{N \times n}$ has i.i.d. entries of zero mean and variance $1/n$, $\mathbf{T}_N$ has l.s.d. H and $N/n \to c$. Let $B = \{\underline{m} \mid \underline{m} \neq 0, -1/\underline{m} \in S_H^c\}$ and $x_{\underline{F}}$ be the function defined on B by*

$$x_{\underline{F}}(\underline{m}) = -\frac{1}{\underline{m}} + c \int \frac{t}{1 + t\underline{m}} dH(t). \tag{12}$$

For $x_0 \in \mathbb{R}^*$, we can then determine the limit of $m_{\underline{F}}(z)$ as $z \to x_0$, $z \in \mathbb{C}^+$, along the following rules,

1. *If $x_0 \in S_{\underline{F}}^c$, then the equation $x_0 = x_{\underline{F}}(\underline{m})$ in the dummy variable $\underline{m}$ has a unique real solution $m_0 \in B$ such that $x'_{\underline{F}}(m_0) > 0$; this m_0 is the limit of $m_{\underline{F}}(z)$ when $z \to x_0$, $z \in \mathbb{C}^+$. Conversely, for $m_0 \in B$ such that $x'_{\underline{F}}(m_0) > 0$, $x_0 = x_{\underline{F}}(m_0) \in S_{\underline{F}}^c$.*

2. *If $x_0 \in S_{\underline{F}}$, then the equation $x_0 = x_{\underline{F}}(\underline{m})$ in the dummy variable $\underline{m}$ has a unique complex solution $m_0 \in B$ with positive imaginary part; this m_0 is the limit of $m_{\underline{F}}(z)$ when $z \to x_0$, $z \in \mathbb{C}^+$.*

From rule 1, it is possible to determine the exact support of F. It indeed suffices to draw $x_{\underline{F}}(\underline{m})$ for $-1/\underline{m} \in \mathbb{R} \setminus S_H$. Whenever $x_{\underline{F}}$ is increasing on an interval I, $x_{\underline{F}}(I)$ is outside $S_{\underline{F}}$. The support $S_{\underline{F}}$ of $\underline{F}$, and therefore of F (modulo the mass in 0), is then defined exactly by

$$S_{\underline{F}} = \mathbb{R} \setminus \bigcup_{\substack{a,b \in \mathbb{R} \\ a<b}} \left\{ x_{\underline{F}}((a,b)) \mid \forall \underline{m} \in (a,b), x'_{\underline{F}}(\underline{m}) > 0 \right\}.$$

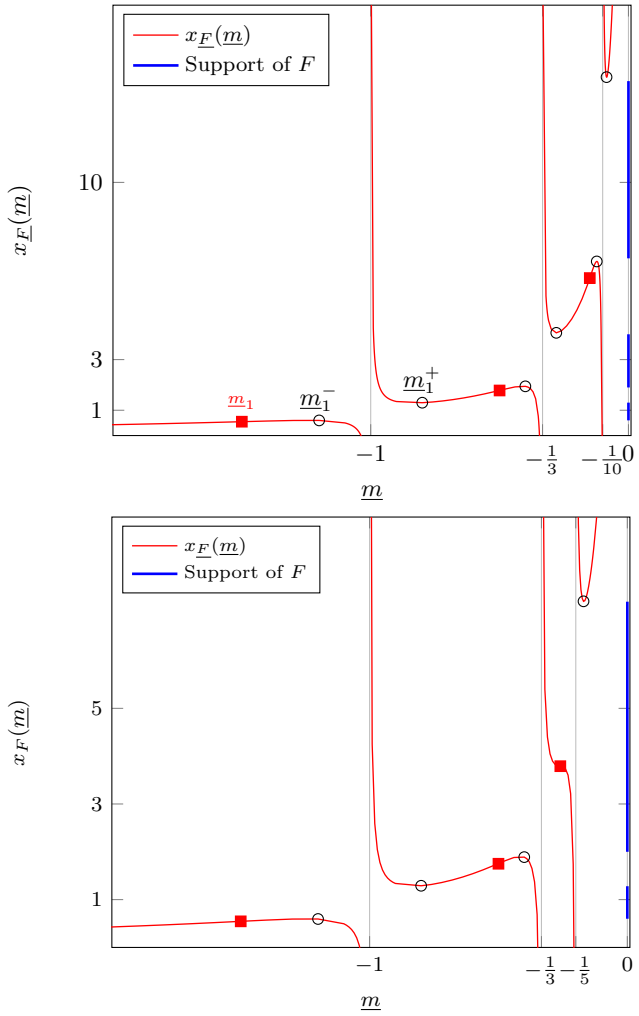

Figure 3: $x_{\underline{F}}(\underline{m})$ for $\underline{m}$ real, $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in 1, 3 and 10 (top) and 1, 3 and 5 (bottom), $c = 1/10$ in both cases. Local extrema are marked in circles, inflexion points are marked in squares. The support of F can be read on the right vertical axises.

This is depicted in Figure 3 in the case when H is composed of three evenly weighted masses t_1, t_2, t_3 in $\{1, 3, 5\}$ or $\{1, 3, 10\}$ and $c = 1/10$. Notice that, in the case where $t_3 = 10$, F is divided into three clusters while when $t_3 = 5$, F is divided into only two clusters, which is due to the fact that $x_{\underline{F}}$ is non-increasing in the interval $(-1/3, -1/5)$.

From Figure 3 and Theorem 0.13, we now observe that $x'_{\underline{F}}(\underline{m})$ has exactly $2K_F$ roots with K_F the number of clusters in F. Denote these roots $\underline{m}_1^- < \underline{m}_1^+ \leq \underline{m}_2^- < \underline{m}_2^+ < \ldots \leq \underline{m}_{K_F}^- < \underline{m}_{K_F}^+$. Each pair $(\underline{m}_j^-, \underline{m}_j^+)$ is such that $x_{\underline{F}}([\underline{m}_j^-, \underline{m}_j^+])$ is the j^{th} cluster in F. We therefore have a way to determine the support of the asymptotic spectrum through the function $x'_{\underline{F}}$. This is presented in the following result

Theorem 0.14 ([23],[24]). *Let $\mathbf{B}_N \in \mathbb{C}^{N \times N}$ be defined as in Theorem 0.15. Then the support*

S_F of the l.s.d. F of $\mathbf{B}_N$ is defined as

$$S_F = \bigcup_{j=1}^{K_F} [x_j^-, x_j^+],$$

where $x_1^-, x_1^+, \ldots, x_{K_F}^-, x_{K_F}^+$ are defined as

$$x_j^- = -\frac{1}{\underline{m}_j^-} + \sum_{r=1}^{K} c_r \frac{t_r}{1 + t_r \underline{m}_j^-},$$
$$x_j^+ = -\frac{1}{\underline{m}_j^-} + \sum_{r=1}^{K} c_r \frac{t_r}{1 + t_r \underline{m}_j^+},$$

with $\underline{m}_1^- < \underline{m}_1^+ \leq \underline{m}_2^- < \underline{m}_2^+ \leq \ldots \leq \underline{m}_{K_F}^- < \underline{m}_{K_F}^+$ the $2K_F$ (possibly counted with multiplicity) real roots of the equation in $\underline{m}$,

$$\sum_{r=1}^{K} c_r \frac{t_r^2 \underline{m}^2}{(1 + t_r \underline{m}^2)^2} = 1.$$

Notice further from Figure 3 that, while $x'_{\underline{F}}(\underline{m})$ might not have roots on some intervals $(-1/t_{k-1}, -1/t_k)$, it always has a unique inflexion point there. This is proved in [24] by observing that $x''_{\underline{F}}(\underline{m}) = 0$ is equivalent to

$$\sum_{r=1}^{K} c_r \frac{t_r^3 \underline{m}^3}{(1 + t_r \underline{m})^3} - 1 = 0,$$

the left-hand side of which has always positive derivative and shows asymptotes in the neighborhood of t_r; hence the existence of a unique inflexion point on every interval $(-1/t_{k-1}, -1/t_k)$, for $1 \leq k \leq K$, with convention $t_0 = 0+$. When $x_{\underline{F}}$ increases on an interval $(-1/t_{k-1}, -1/t_k)$, it must have its inflexion point in a point of positive derivative (from the concavity change induced by the asymptotes). Therefore, to verify that cluster k_F is disjoint from clusters $(k-1)_F$ and $(k+1)_F$ (when they exist), it suffices to verify that the $(k-1)^{th}$ and k^{th} roots $\underline{m}_{k-1}$ and $\underline{m}_k$ of $x''_{\underline{F}}(\underline{m})$ are such that $x'_{\underline{F}}(\underline{m}_{k-1}) > 0$ and $x'_{\underline{F}}(\underline{m}_k) > 0$. This is exactly what the following result states for the case of a sample covariance matrix whose population covariance matrix has few eigenvalues, each with a large multiplicity.

Theorem 0.15 ([25],[24])**.** *Let $\mathbf{B}_N$ be defined as in Theorem 0.10, with $\mathbf{T}_N = \mathrm{diag}(\tau_1, \ldots, \tau_N) \in \mathbb{R}^{N \times N}$, diagonal containing K distinct eigenvalues $0 < t_1 < \ldots < t_K$, for some fixed K. Denote N_k the multiplicity of the k^{th} largest eigenvalue, counted with multiplicity (assuming ordering of the τ_i, we may then have $\tau_1 = \ldots = \tau_{N_1} = t_1$, $\ldots$, $\tau_{N-N_K+1} = \ldots = \tau_N = t_K$). Assume also that for all $1 \leq r \leq K$, $N_r/n \to c_r > 0$, and $N/n \to c$, with $0 < c < \infty$. Then the cluster k_F associated to the eigenvalue t_k in the l.s.d. F of $\mathbf{B}_N$ is distinct from the clusters $(k-1)_F$ and $(k+1)_F$ (when they exist), associated to t_{k-1} and t_{k+1} in F, respectively, if and only if*

$$\sum_{r=1}^{K} c_r \frac{t_r^2 \underline{m}_k^2}{(1 + t_r \underline{m}_k^2)^2} < 1,$$
$$\sum_{r=1}^{K} c_r \frac{t_r^2 \underline{m}_{k+1}^2}{(1 + t_r \underline{m}_{k+1}^2)^2} < 1, \tag{13}$$

where $\underline{m}_1, \ldots, \underline{m}_K$ *are such that* $\underline{m}_{K+1} = 0$ *and* $\underline{m}_1 < \underline{m}_2 < \ldots < \underline{m}_K$ *are the* K *solutions of the equation in* $\underline{m}$*,*

$$\sum_{r=1}^{K} c_r \frac{t_r^3 \underline{m}^3}{(1 + t_r \underline{m})^3} = 1.$$

For $k = 1$*, this condition ensures* $1_F = 2_F - 1$*; for* $k = K$*, this ensures* $K_F = (K-1)_F + 1$ *and for* $1 < k < K$*, this ensures* $(k-1)_F + 1 = k_F = (k+1)_F - 1$*.*

This result is again fundamental in the sense that the separability of subsequent clusters in the support of the l.s.d. of $\mathbf{B}_N$ will play a fundamental role in the validity of statistical inference methods. In the subsequent section, we introduce the key ideas that allow statistical inference for sample covariance matrices.

0.4 Statistical inference

Statistical inference allows for the estimation of deterministic parameters present in a stochastic model based on observations of random realisations of the model. In the context of sample covariance matrices, statistical inference methods consist in providing estimates of functionals of the eigenvalue distribution of the population covariance matrix $\mathbf{T}_N \in \mathbb{C}^{N \times N}$ based on the observation $\mathbf{Y}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N$ with $\mathbf{X}_N \in \mathbb{C}^{N \times n}$ a random matrix of independent and identically distributed entries. Different methods exist that allow for statistical inference that mostly rely on the study of the l.s.d. of the sample covariance matrix $\mathbf{B}_N = \frac{1}{n} \mathbf{Y}_N \mathbf{Y}_N^{\mathsf{H}}$. One of these methods relates to free probability theory [26], and more specifically to free deconvolution approaches, see e.g., [27], [28]. The idea behind free deconvolution is based on the fact that the moments of the l.s.d. of some random matrix models can be written as a polynomial function of the moments of the l.s.d. of another (random) matrix in the model, under some proper conditions. Typically, the moments of the l.s.d. of $\mathbf{T}_N$ can be written as a polynomial of the moments of the (almost sure) l.s.d. of $\mathbf{B}_N$, if $\mathbf{X}_N$ has Gaussian entries and the e.s.d. of $\mathbf{T}_N$ has uniformly bounded support. Therefore, to put it simply, one can obtain all moments of $\mathbf{T}_N$ based on a sufficiently large observation of $\mathbf{B}_N$; this allows one to recover the l.s.d. of $\mathbf{T}_N$ (since Carleman condition is satisfied) and therefore any functional of the l.s.d. However natural, this method has some major drawbacks. From a practical point of view, a reliable estimation of moments of high order requires extremely large dimensional matrix observations. This is due to the fact that the estimate of the moment of order k of the l.s.d. is based on polynomial expressions of the estimates of moments of lower orders. A small error in the estimate in a low order moment therefore propagates as a large error for higher moments; it is therefore compelling to obtain accurate first order estimates, hence large dimensional observations.

We will not further investigate the moment-based approach above, which we discuss in more detail with a proper introduction to free probability theory in [29]. Instead, we introduce the methods based on the Stieltjes transform and those rely strongly on the results described in the previous section. We will introduce this method for the sample covariance matrix model discussed so far, because it will be instrumental to understanding the power estimator introduced in Section 0.5.2. Similar results have been provided for other models of interest to telecommunications, as for instance the so-called information-plus-noise model, studied in [30].

The central idea is based on a trivial application of the Cauchy complex integration formula

[31]. Consider f some complex holomorphic function on $U \subset \mathbb{C}$, H a distribution function and denote G the functional

$$G(f) = \int f(z) dH(z).$$

From the Cauchy integration formula, we have, for a *negatively* oriented closed path γ enclosing the support of H and with winding number one,

$$\begin{aligned} G(f) &= \frac{1}{2\pi i} \int \oint_\gamma \frac{f(\omega)}{z-\omega} d\omega dH(z) \\ &= \frac{1}{2\pi i} \oint_\gamma \int \frac{f(\omega)}{z-\omega} dH(z) d\omega \\ &= \frac{1}{2\pi i} \oint_\gamma f(\omega) m_H(\omega) d\omega, \end{aligned} \tag{14}$$

the integral inversion being valid since $f(\omega)/(z-\omega)$ is bounded for $\omega \in \gamma$. Note that the sign inversion due to the negative contour orientation is compensated by the sign reversal of $(\omega - z)$ in the denominator.

If dH is a sum of finite or countable masses and one is interested in evaluating $f(\lambda_k)$, with λ_k the value of the k^{th} mass with weight l_k, then on a negatively oriented contour γ_k enclosing λ_k and excluding λ_j, $j \neq k$,

$$l_k f(\lambda_k) = \frac{1}{2\pi i} \oint_{\gamma_k} f(\omega) m_H(\omega) d\omega. \tag{15}$$

This last expression is particularly convenient when one has access to H only through an expression of its Stieltjes transform.

Now, in terms of random matrices, for the sample covariance matrix $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, we already noticed that the l.s.d. F of $\mathbf{B}_N$ (or equivalently the l.s.d. $\underline{F}$ of $\underline{\mathbf{B}}_N = \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N \mathbf{X}_N$) can be rewritten under the form (9), which can further be rewritten

$$\frac{c}{m_{\underline{F}}(z)} m_H\left(-\frac{1}{m_{\underline{F}}(z)}\right) = -z m_{\underline{F}}(z) + (c-1), \tag{16}$$

where H is the l.s.d. of $\mathbf{T}_N$. Note that it is allowed to evaluate m_H in $-1/m_{\underline{F}}(z)$ for $z \in \mathbb{C}^+$ since $-1/m_{\underline{F}}(z) \in \mathbb{C}^+$.

As a consequence, if one only has access to $F^{\mathbf{B}_N}$ (from the observation $\mathbf{B}_N$), then the only link from the observation to H is obtained by (i) the fact that $F^{\underline{\mathbf{B}}_N} \Rightarrow \underline{F}$ almost surely and (ii) the fact that $\underline{F}$ and H are related through (16). Evaluating a functional f of the eigenvalue λ_k of $\mathbf{T}_N$ is then made possible by (15). The relations (15) and (16) are the essential ingredients behind the derivation of a consistent estimator for $f(\lambda_k)$.

We now concentrate specifically on the case of the sample covariance matrix $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N$ defined as in Theorem 0.10 with $\mathbf{T}_N$ composed of K distinct eigenvalues $t_1, \ldots, t_K$ of multiplicities $N_1, \ldots, N_K$, respectively. We further denote $c_k \triangleq \lim_n N_k/n$ and will discuss the question of estimating t_k itself. What follows summarizes the original ideas of Mestre in [23] and [25]. We have from Equation (15) that, for any continuous f and for any *negatively oriented* contour

$\mathcal{C}_k$ that encloses t_k and t_k only, $f(t_k)$ can be written under the form

$$\begin{aligned}\frac{N_k}{N} f(t_k) &= \frac{1}{2\pi i} \oint_{\mathcal{C}_k} f(\omega) m_H(\omega) d\omega \\ &= \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \frac{1}{N} \sum_{r=1}^{K} N_r \frac{f(\omega)}{t_r - \omega} d\omega\end{aligned}$$

with H the limit $F^{\mathbf{T}_N} \Rightarrow H$. This provides a link between $f(t_k)$ for all continuous f and the Stieltjes transform $m_H(z)$.

Letting $f(x) = x$ and taking the limit $N \to \infty$, $N_k/N \to c_k/c$, with $c \triangleq c_1 + \ldots + c_K$ the limit of N/n, we have

$$\frac{c_k}{c} t_k = \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \omega m_H(\omega) d\omega. \tag{17}$$

We now want to express m_H as a function of m_F, the Stieltjes transform of the l.s.d. F of $\mathbf{B}_N$. For this, we have the two relations (10), i.e.,

$$m_{\underline{F}}(z) = c m_F(z) + (c-1)\frac{1}{z}$$

and (16) with $F^T = H$, i.e.,

$$\frac{c}{m_{\underline{F}}(z)} m_H\left(-\frac{1}{m_{\underline{F}}(z)}\right) = -z m_{\underline{F}}(z) + (c-1).$$

Together, those two equations give the simpler expression

$$m_H\left(-\frac{1}{m_{\underline{F}}(z)}\right) = -z m_{\underline{F}}(z) m_F(z).$$

Applying the variable change $\omega = -1/m_{\underline{F}}(z)$ in (17), we obtain

$$\begin{aligned}\frac{c_k}{c} t_k &= \frac{1}{2\pi i} \oint_{\mathcal{C}_{\underline{F},k}} z \frac{m'_{\underline{F}}(z)}{m_{\underline{F}}(z) c} + \frac{1-c}{c} \frac{m_{\underline{F}}(z)'}{m^2_{\underline{F}}(z)} dz \\ &= \frac{1}{c} \frac{1}{2\pi i} \oint_{\mathcal{C}_{\underline{F},k}} z \frac{m'_{\underline{F}}(z)}{m_{\underline{F}}(z)} dz,\end{aligned} \tag{18}$$

where $\mathcal{C}_{\underline{F},k}$ is the preimage of $\mathcal{C}_k$ by $-1/m_{\underline{F}}$. The second equality (18) comes from the fact that the second term in the previous relation is the derivative of $(c-1)/(c m_{\underline{F}}(z))$, which therefore integrates to 0 on a closed path, from classical real or complex integration rules [31]. Obviously, since $z \in \mathbb{C}^+$ is equivalent to $-1/m_{\underline{F}}(z) \in \mathbb{C}^+$ (the same being true if $\mathbb{C}^+$ is replaced by $\mathbb{C}^-$), $\mathcal{C}_{\underline{F},k}$ is clearly continuous and of non-zero imaginary part whenever $\Im[z] \neq 0$. Now, one must be careful about the exact choice of $\mathcal{C}_{\underline{F},k}$.

We make the important assumption that the index k satisfies the separability conditions of Theorem 0.15. This is, the cluster k_F associated to k in F is distinct from $(k-1)_F$ and $(k+1)_F$ (whenever they exist). Let us then pick $x_F^{(l)}$ and $x_F^{(r)}$ two real values such that

$$x^+_{(k-1)_F} < x_F^{(l)} < x^-_{k_F} < x^+_{k_F} < x_F^{(r)} < x^-_{(k+1)_F}$$

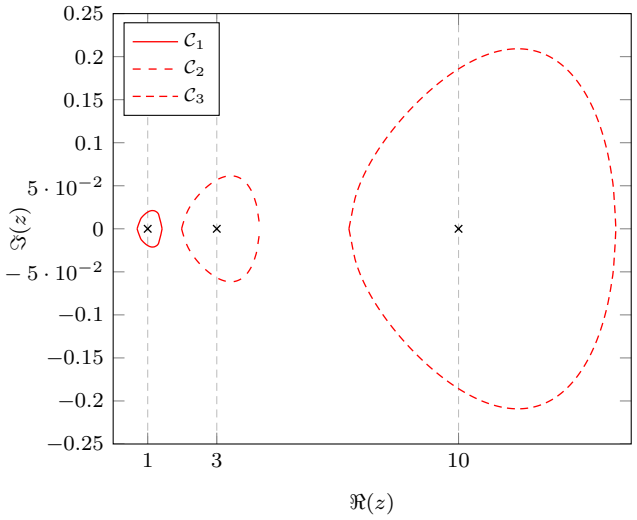

Figure 4: Integration contours $\mathcal{C}_k$, $k \in \{1, 2, 3\}$, preimage of $\mathcal{C}_{\underline{F},k}$ by $-1/m_{\underline{F}}$, for $\mathcal{C}_{\underline{F},k}$ a circular contour around cluster k_F, when $\mathbf{T}_N$ composed of three distinct entries, $t_1 = 1$, $t_2 = 3$, $t_3 = 10$, $N_1 = N_2 = N_3$, $N/n = 1/10$.

with $\{x_1^-, x_1^+, \ldots, x_{K_F}^-, x_{K_F}^+\}$ the support boundary of F, as defined in Theorem 0.14. Now remember Theorem 0.13 and Figure 3; for $x_F^{(l)}$ as defined previously, $m_{\underline{F}}(z)$ has a limit $m^{(l)} \in \mathbb{R}$ as $z \to x_F^{(l)}$, $z \in \mathbb{C}^+$, and a limit $m^{(r)} \in \mathbb{R}$ as $z \to x_F^{(r)}$, $z \in \mathbb{C}^+$, those two limits verifying

$$t_{k-1} < x^{(l)} < t_k < x^{(r)} < t_{k+1}, \tag{19}$$

with $x^{(l)} \triangleq -1/m^{(l)}$ and $x^{(r)} \triangleq -1/m^{(r)}$.

This is the most important outcome of the integration process. Let us define $\mathcal{C}_{\underline{F},k}$ to be *any* continuous contour surrounding cluster k_F such that $\mathcal{C}_{\underline{F},k}$ crosses the real axis in only two points, namely $x_F^{(l)}$ and $x_F^{(r)}$. Since $-1/m_{\underline{F}}(\mathbb{C}^+) \subset \mathbb{C}^+$ and $-1/m_{\underline{F}}(\mathbb{C}^-) \subset \mathbb{C}^-$, $\mathcal{C}_k$ does not cross the real axis whenever $\Im[z] \neq 0$ and is obviously continuously differentiable there; now $\mathcal{C}_k$ crosses the real axis in $x^{(l)}$ and $x^{(r)}$, and is in fact continuous there. Because of (19), we then have that $\mathcal{C}_k$ is (at least) continuous and piecewise continuously differentiable and encloses *only* t_k. This is what is required to ensure the validity of (18). In Figure 4, we consider the case when $\mathbf{T}_N$ is formed of three evenly weighted eigenvalues $t_1 = 1$, $t_2 = 3$ and $t_3 = 10$, and we depict the contours $\mathcal{C}_k$, preimages of $\mathcal{C}_{\underline{F},k}$, $k \in \{1, 2, 3\}$, circular contours around the clusters k_F such that they cross the real line in the positions $x_F^{(l)}$ and $x_F^{(r)}$, corresponding to the inflexion points of $x_{\underline{F}}(\underline{m})$ (and an arbitrary large value for the extreme right point).

The difficult part of the proof is completed. The rest will unfold more naturally. We start by considering the following expression,

$$\begin{aligned}\hat{t}_k &\triangleq \frac{1}{2\pi i}\frac{n}{N_k}\oint_{\mathcal{C}_{\underline{F},k}} z\frac{m'_{F^{\mathbf{B}_N}}(z)}{m_{F^{\mathbf{B}_N}}(z)}dz \\ &= \frac{1}{2\pi i}\frac{n}{N_k}\oint_{\mathcal{C}_{\underline{F},k}} z\frac{\frac{1}{n}\sum_{i=1}^n \frac{1}{(\lambda_i - z)^2}}{\frac{1}{n}\sum_{i=1}^n \frac{1}{\lambda_i - z}}dz,\end{aligned} \tag{20}$$

where we remind that $\underline{\mathbf{B}}_N \triangleq \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N \mathbf{X}_N$ and where, if $n \geq N$, we defined $\lambda_{N+1} = \ldots = \lambda_n = 0$.

The value $\hat{t}_k$ can be viewed as the empirical counterpart of t_k. Now, we know from Theorem 0.9 that $m_{F^{\mathbf{B}_N}}(z) \xrightarrow{\text{a.s.}} m_F(z)$ and $m_{F^{\underline{\mathbf{B}}_N}}(z) \to m_{\underline{F}}(z)$. It is not difficult to verify, from the fact that $m_{\underline{F}}$ is holomorphic, that the same convergence holds for the successive derivatives.

At this point, we need the two fundamental results that are Theorem 0.10 and Theorem 0.12. We know that, for all matrices $\mathbf{B}_N$ in a set of probability one, all the eigenvalues of $\mathbf{B}_N$ are contained in the support of F for all large N, and that the eigenvalues of $\mathbf{B}_N$ contained in cluster k_F are exactly $\{\lambda_i, i \in \mathcal{N}_k\}$ for these large N, with $\mathcal{N}_k = \{\sum_{j=1}^{k-1} N_j + 1, \ldots, \sum_{j=1}^{k} N_j\}$. Take such a $\mathbf{B}_N$. For all large N, $m_{\mathbf{B}_N}(z)$ is uniformly bounded over N and $z \in \mathcal{C}_{\underline{F},k}$, since $\mathcal{C}_{\underline{F},k}$ is away from the support of F. The integrand in the right-hand side of (20) is then uniformly bounded for all large N and for all $z \in \mathcal{C}_{\underline{F},k}$. By the dominated convergence theorem, Theorem 16.4 in [14], we then have that $\hat{t}_k - t_k \xrightarrow{\text{a.s.}} 0$.

It then remains to evaluate $\hat{t}_k$ explicitly. This is performed by residue calculus [31], i.e., by determining the poles in the expanded expression of $\hat{t}_k$ (when developing $m_{F^{\mathbf{B}_N}}(z)$ in its full expression). Those poles are found to be $\lambda_1, \ldots, \lambda_N$ (indeed, the integrand of (20) behaves like $O(1/(\lambda_i - z))$ for $z \simeq \lambda_i$) and $\mu_1, \ldots, \mu_N$, the N real roots of the equation in μ, $m_{F^{\underline{\mathbf{B}}_N}}(\mu) = 0$ (indeed, the denominator of the integrand cancels for $z = \mu_i$ while the numerator is non zero). Since $\mathcal{C}_{\underline{F},k}$ encloses only those values λ_i such that $i \in \mathcal{N}_k$, the other poles are discarded. Noticing now that $m_{F^{\mathbf{B}_N}}(\mu) \to \pm\infty$ as $\mu \to \lambda_i$, we deduce that $\mu_1 < \lambda_1 < \mu_2 < \ldots < \mu_N < \lambda_N$, and therefore we have that μ_i, $i \in \mathcal{N}_k$ are all in $\mathcal{C}_{\underline{F},k}$ but maybe for μ_j, $j = \min \mathcal{N}_k$. It can in fact be shown that μ_j is also in $\mathcal{C}_{\underline{F},k}$. To notice this last remaining fact, observe simply that

$$\frac{1}{2\pi i} \oint_{\mathcal{C}_k} \frac{1}{\omega} d\omega = 0.$$

since 0 is not contained in the contour $\mathcal{C}_k$. Applying the variable change $\omega = -1/m_{\underline{F}}(z)$ as previously, this gives

$$\oint_{\mathcal{C}_{\underline{F},k}} \frac{m'_{\underline{F}}(z)}{m^2_{\underline{F}}(z)} dz = 0. \tag{21}$$

From the same reasoning as above, with the dominated convergence theorem argument, we have that for sufficiently large N and almost surely,

$$\left| \oint_{\mathcal{C}_{\underline{F},k}} \frac{m'_{F^{\underline{\mathbf{B}}_N}}(z)}{m^2_{F^{\underline{\mathbf{B}}_N}}(z)} dz \right| < \frac{1}{2}. \tag{22}$$

At this point, we need to proceed to residue calculus in order to compute the integral in the left-hand side of (22). We will in fact prove that the value of this integral is an integer, hence necessarily equal to zero from the inequality (22). Notice indeed that the poles of (21) are the λ_i and the μ_i that lie inside the integration contour $\mathcal{C}_{\underline{F},k}$, all of order one with residues equal to -1 and 1, respectively. These residues are obtained using in particular L'Hospital rule, as detailed below. Therefore, (21) equals the number of such λ_i minus the number of such μ_i (remember that the integration contour is negatively oriented, so we need to reverse the signs). We however already know that this difference, for large N, equals either 0 or 1, since only the position of the leftmost μ_i is unknown yet. But since the integral is asymptotically less than $1/2$, this implies that it is identically zero, and therefore the leftmost μ_i (indexed by $\min \mathcal{N}_k$) also lies inside the integration contour.

From this point on, we can evaluate (20), which is clearly determined since we know exactly which eigenvalues of $\mathbf{B}_N$ are contained (with probability one for all large N) within the integration contour. This calls again for residue calculus, the steps of which are detailed below. Denoting

$$f(z) = z\frac{m'_{F^{\underline{\mathbf{B}}_N}}(z)}{m_{F^{\underline{\mathbf{B}}_N}}(z)},$$

we find that λ_i (inside $\mathcal{C}_{\underline{F},k}$) is a pole of order 1 with residue

$$\lim_{z\to\lambda_i}(z-\lambda_i)f(z) = -\lambda_i,$$

which is straightforwardly obtained from the fact that $f(z) \sim \frac{1}{\lambda_i - z}$ as $z \sim \lambda_i$. Also μ_i (inside $\mathcal{C}_{\underline{F},k}$) is a pole of order 1 with residue

$$\lim_{z\to\mu_i}(z-\mu_i)f(z) = \mu_i,$$

which is obtained using L'Hospital rule: upon existence of a limit, we indeed have

$$\begin{aligned}
\lim_{z\to\mu_i}(z-\mu_i)f(z) &= \lim_{z\to\mu_i}\frac{\frac{d}{dz}\left[(z-\mu_i)zm'_{F^{\underline{\mathbf{B}}_N}(z)}\right]}{\frac{d}{dz}\left[m_{F^{\underline{\mathbf{B}}_N}(z)}\right]} \\
&= \lim_{z\to\mu_i}\frac{zm'_{F^{\underline{\mathbf{B}}_N}(z)} + z(z-\mu_i)m''_{F^{\underline{\mathbf{B}}_N}(z)} + (z-\mu_i)m'_{F^{\underline{\mathbf{B}}_N}(z)}}{m'_{F^{\underline{\mathbf{B}}_N}(z)}} \\
&= \lim_{z\to\mu_i} z \\
&= \mu_i.
\end{aligned}$$

Since the integration contour is chosen to be *negatively oriented*, it must be kept in mind that the signs of the residues need be inverted in the final relation.

Noticing finally that $\mu_1, \ldots, \mu_N$ are also the eigenvalues of $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{n}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$, with $\boldsymbol{\lambda} \triangleq (\lambda_1, \ldots, \lambda_N)^{\mathsf{T}}$, from a lemma provided in [24], Lemma 1, and [32], we finally have the following statistical inference result for sample covariance matrices.

Theorem 0.16 ([25])**.** *Let $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}} \in \mathbb{C}^{N\times N}$ be defined as in Theorem 0.15, i.e., $\mathbf{T}_N$ has K distinct eigenvalues $t_1 < \ldots < t_K$ with multiplicities $N_1, \ldots, N_K$, respectively, for all r, $N_r/n \to c_r$, $0 < c_r < \infty$, and the separability conditions* (13) *are satisfied. Further denote $\lambda_1 \leq \ldots \leq \lambda_N$ the eigenvalues of $\mathbf{B}_N$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)^{\mathsf{T}}$. Let $k \in \{1, \ldots, K\}$, and define*

$$\hat{t}_k = \frac{n}{N_k}\sum_{m\in\mathcal{N}_k}(\lambda_m - \mu_m) \tag{23}$$

with $\mathcal{N}_k = \{\sum_{j=1}^{k-1} N_j + 1, \ldots, \sum_{j=1}^{k} N_j\}$ and $\mu_1 \leq \ldots \leq \mu_N$ are the ordered eigenvalues of the matrix $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{n}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$.

Then, if condition (13) *is fulfilled, we have*

$$\hat{t}_k - t_k \to 0$$

almost surely as $N, n \to \infty$, $N/n \to c$, $0 < c < \infty$.

Similarly, for the quadratic form, the following holds.

Theorem 0.17 ([25]). *Let $\mathbf{B}_N$ be defined as in Theorem 0.16, and denote $\mathbf{B}_N = \sum_{k=1}^{N} \lambda_k \mathbf{b}_k \mathbf{b}_k^{\mathsf{H}}$, $\mathbf{b}_k^{\mathsf{H}} \mathbf{b}_i = \delta_k^i$, the spectral decomposition of $\mathbf{B}_N$. Similarly, denote $\mathbf{T}_N = \sum_{k=1}^{K} t_k \mathbf{U}_k \mathbf{U}_k^{\mathsf{H}}$, $\mathbf{U}_k^{\mathsf{H}} \mathbf{U}_k = \mathbf{I}_{n_k}$, with $\mathbf{U}_k \in \mathbb{C}^{N \times N_k}$ the eigenspace associated to t_k. For given vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$, denote*

$$u(k; \mathbf{x}, \mathbf{y}) \triangleq \mathbf{x}^{\mathsf{H}} \mathbf{U}_k \mathbf{U}_k^{\mathsf{H}} \mathbf{y}.$$

Then we have

$$\hat{u}(k; \mathbf{x}, \mathbf{y}) - u(k; \mathbf{x}, \mathbf{y}) \xrightarrow{\text{a.s.}} 0$$

as $N, n \to \infty$ with ratio $c_N = N/n \to c$, where

$$\hat{u}(k; \mathbf{x}, \mathbf{y}) \triangleq \sum_{i=1}^{N} \theta_k(i) \mathbf{x}^{\mathsf{H}} \mathbf{b}_k \mathbf{b}_k^{\mathsf{H}} \mathbf{y}$$

and $\theta_k(i)$ is defined by

$$\theta_i(k) = \begin{cases} -\phi_k(i) & , \ i \notin \mathcal{N}_k \\ 1 + \psi_k(i) & , \ i \in \mathcal{N}_k, \end{cases}$$

with

$$\phi_k(i) = \sum_{r \in \mathcal{N}_k} \left(\frac{\lambda_r}{\lambda_i - \lambda_r} - \frac{\mu_r}{\lambda_i - \mu_r} \right),$$

$$\psi_k(i) = \sum_{r \notin \mathcal{N}_k} \left(\frac{\lambda_r}{\lambda_i - \lambda_r} - \frac{\mu_r}{\lambda_i - \mu_r} \right)$$

and $\mathcal{N}_k$, $\mu_1, \ldots, \mu_N$ defined as in Theorem 0.16.

The estimator proposed in Theorem 0.16 is extremely accurate and is in fact much more flexible and precise than free deconvolution approaches. A visual comparison is proposed in Figure 5 for the same scenario as in the top Figure 3, where the free deconvolution (also called moment-based) method is based on the inference techniques proposed in e.g., [33], [27]. Nonetheless, it must be stressed that the cluster separability condition, necessary to the validity of the Stieltjes transform approach, is mandatory and sometimes a rather strong assumption. Typically, the number of observations must be rather large compared to the number of sensors in order to be able to resolve close values of t_k.

0.5 Applications

In this section, we apply the random matrix methods developed above to the successive problems of multi-dimensional binary hypothesis testing and parameter estimation.

0.5.1 Binary hypothesis testing

We first consider the problem of detecting the presence of a signal source impaired by white Gaussian noise. The question is therefore to decide whether only noise is being sensed or if some coherent signal plus noise is sensed.

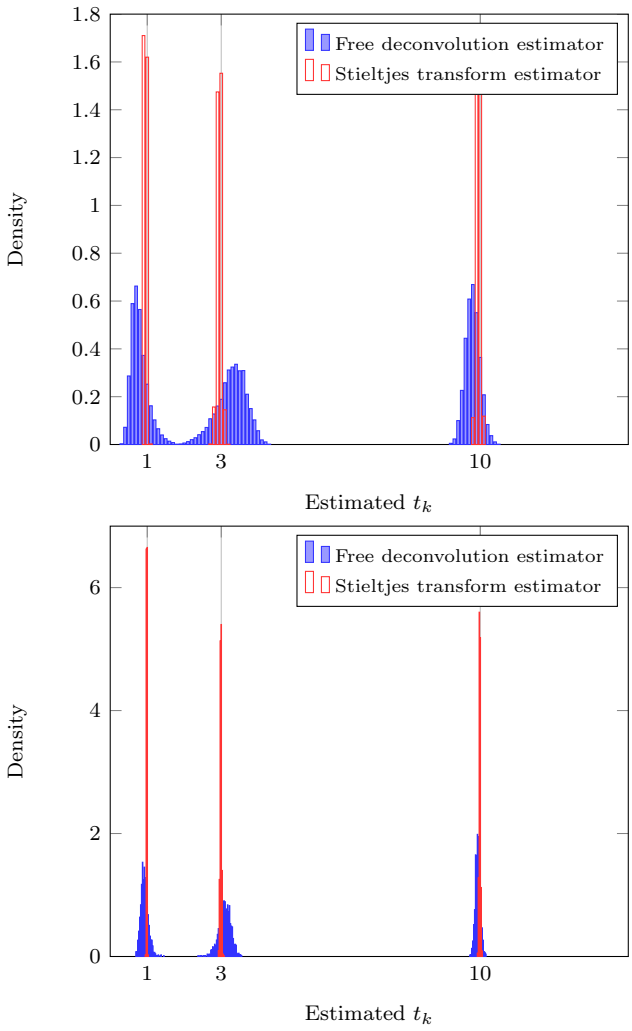

Figure 5: Estimation of t_1, t_2, t_3 in the model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$ based on first three empirical moments of $\mathbf{B}_N$ and Newton-Girard inversion, see [33], for $N_1/N = N_2/N = N_3/N = 1/3$,$N/n = 1/10$, for $100,000$ simulation runs; Top $N = 30$, $n = 90$, bottom $N = 90$, $n = 270$. Comparison is made against the Stieltjes transform estimator of Theorem 0.16.

Precisely, we consider a signal source or transmitter of dimension K and a sink or receiver composed of N sensors. The linear filter between the transmitter and the receiver is modelled by the matrix $\mathbf{H} \in \mathbb{C}^{N\times K}$, with $(i,j)^{th}$ entries h_{ij}. If at time l the transmitter emits data, those are denoted by the K-dimensional vector $\mathbf{x}^{(l)} = (x_1^{(l)}, \ldots, x_K^{(l)})^\mathsf{T} \in \mathbb{C}^K$. The additive white Gaussian noise at the receiver is modelled, at time l, by the vector $\sigma\mathbf{w}^{(l)} = \sigma(w_1^{(l)}, \ldots, w_N^{(l)})^\mathsf{T} \in \mathbb{C}^N$, where σ^2 denotes the variance of the noise vector entries. Without generality restriction, we consider in the following zero mean and unit variance of the entries of both $\mathbf{w}^{(l)}$ and $\mathbf{x}^{(l)}$, i.e., $\mathrm{E}[|w_i^{(l)}|^2] = 1$, $\mathrm{E}[|x_i^{(l)}|^2] = 1$ for all i. We then denote $\mathbf{y}^{(l)} = (y_1^{(l)}, \ldots, y_N^{(l)})^\mathsf{T}$ the N-dimensional data received at time l. Assuming the filter is static during at least M sampling periods, we finally denote $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}] \in \mathbb{C}^{N\times M}$ the matrix of the concatenated receive vectors.

Depending on whether the transmitter emits data, we consider the following hypotheses

- $\mathcal{H}_0$. Only background noise is received.
- $\mathcal{H}_1$. Data plus background noise are received.

Therefore, under condition $\mathcal{H}_0$, we have the model,

$$\mathbf{Y} = \sigma \begin{pmatrix} w_1^{(1)} & \cdots & w_1^{(M)} \\ \vdots & \ddots & \vdots \\ w_N^{(1)} & \cdots & w_N^{(M)} \end{pmatrix}$$

and under condition $\mathcal{H}_1$,

$$\mathbf{Y} = \begin{pmatrix} h_{11} & \ldots & h_{1K} & \sigma & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{N1} & \ldots & h_{NK} & 0 & \cdots & \sigma \end{pmatrix} \begin{pmatrix} x_1^{(1)} & \cdots & \cdots & x_1^{(M)} \\ \vdots & \vdots & \vdots & \vdots \\ x_K^{(1)} & \cdots & \cdots & x_K^{(M)} \\ w_1^{(1)} & \cdots & \cdots & w_1^{(M)} \\ \vdots & \vdots & \vdots & \vdots \\ w_N^{(1)} & \cdots & \cdots & w_N^{(M)} \end{pmatrix}. \tag{24}$$

Under this hypothesis, we further denote $\mathbf{\Sigma}$ the covariance matrix of $\mathbf{y}^{(1)}$,

$$\mathbf{\Sigma} = \mathrm{E}[\mathbf{y}^{(1)}\mathbf{y}^{(1)\mathsf{H}}] = \mathbf{H}\mathbf{H}^\mathsf{H} + \sigma^2\mathbf{I}_N = \mathbf{U}\mathbf{G}\mathbf{U}^\mathsf{H}$$

where $\mathbf{G} = \mathrm{diag}\left(\nu_1 + \sigma^2, \ldots, \nu_N + \sigma^2\right) \in \mathbb{R}^{N\times N}$, with $\{\nu_1, \ldots, \nu_N\}$ the eigenvalues of $\mathbf{H}\mathbf{H}^\mathsf{H}$ and $\mathbf{U} \in \mathbb{C}^{N\times N}$ a certain unitary matrix.

The receiver is entitled to decide whether data were transmitted or not. It is a common assumption to be in the scenario where σ^2 is known in advance, although it is uncommon to know the transfer matrix $\mathbf{H}$. This is true in particular of the wireless signal sensing scenario where $\mathbf{H}$ is the wireless fading channel matrix between two antenna arrays. We consider specifically the Bayesian scenario where some a priori probability distribution for $\mathbf{H}$ is known and that this probability distribution is unitarily invariant. This is in particular relevant when the filter $\mathbf{H}$ presents rotational invariance properties. For simplicity we take $\mathbf{H}$ and $\mathbf{x}^{(l)}$ to be i.i.d. Gaussian

with zero mean and $\mathrm{E}[|h_{ij}|^2] = 1/K$, although our study could go well beyond the Gaussian case.

For simplicity, we consider in the following the case $K = 1$, although a generalized result exists for $K \geq 1$ [2]. The Neyman-Pearson criterion for the receiver to establish whether data were transmitted is based on the ratio

$$C(\mathbf{Y}) = \frac{P_{\mathcal{H}_1|\mathbf{Y}}(\mathbf{Y})}{P_{\mathcal{H}_0|\mathbf{Y}}(\mathbf{Y})}, \tag{25}$$

where $P_{\mathcal{H}_i|\mathbf{Y}}(\mathbf{Y})$ is the probability of the event $\mathcal{H}_i$ conditioned on the observation $\mathbf{Y}$. For a given receive space-time matrix $\mathbf{Y}$, if $C(\mathbf{Y}) > 1$, then the odds are that an informative signal was transmitted, while if $C(\mathbf{Y}) < 1$, it is more likely that only background noise was captured. To ensure a low probability of false alarm (or false positive), i.e., the probability to declare a pure noise sample to carry an informative signal, a certain threshold ξ is generally set such that, when $C(\mathbf{Y}) > \xi$, the receiver declares data were transmitted, while when $C(\mathbf{Y}) < \xi$ the receiver declares that no data were sent. The question of what ratio ξ to be set to ensure a given maximally acceptable false alarm rate will not be treated here. We will however provide an explicit expression of (25) for the aforementioned model, and shall compare its performance to that achieved by the classical energy detector. The results provided in this section are taken from [2].

Applying Bayes' rule, (25) becomes

$$C(\mathbf{Y}) = \frac{P_{\mathcal{H}_1} \cdot P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y})}{P_{\mathcal{H}_0} \cdot P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y})}$$

with $P_{\mathcal{H}_i}$ the a priori probability for hypothesis $\mathcal{H}_i$ to hold. We suppose that no side information allows the receiver to consider that $\mathcal{H}_1$ is more or less probable than $\mathcal{H}_0$, and therefore set $P_{\mathcal{H}_1} = P_{\mathcal{H}_0} = \frac{1}{2}$, so that

$$C(\mathbf{Y}) = \frac{P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y})}{P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y})} \tag{26}$$

reduces to a maximum likelihood ratio.

Likelihood under $\mathcal{H}_0$. In this first scenario, the noise entries $w_i^{(l)}$ are Gaussian and independent. The probability density of $\mathbf{Y}$, that can be seen as a random vector with NM entries, is then an NM multivariate uncorrelated complex Gaussian with covariance matrix $\sigma^2\mathbf{I}_{NM}$,

$$P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y}) = \frac{1}{(\pi\sigma^2)^{NM}} e^{-\frac{1}{\sigma^2}\mathrm{tr}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}}. \tag{27}$$

Denoting $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)^{\mathsf{T}}$ the eigenvalues of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$, (27) only depends on $\sum_{i=1}^N \lambda_i$, as follows

$$P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y}) = \frac{1}{(\pi\sigma^2)^{NM}} e^{-\frac{1}{\sigma^2}\sum_{i=1}^N \lambda_i}.$$

Likelihood under $\mathcal{H}_1$. Under the data plus noise hypothesis $\mathcal{H}_1$, the problem is more involved. The entries of the channel matrix $\mathbf{H}$ are modeled as jointly uncorrelated Gaussian, with $\mathrm{E}[|h_{ij}|^2] = 1/K$. Therefore, since here $K = 1$, $\mathbf{H} \in \mathbb{C}^{N\times 1}$ and $\mathbf{\Sigma} = \mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2\mathbf{I}_N$ has $N-1$ eigenvalues $g_2 = \ldots = g_N$ equal to σ^2 and another distinct eigenvalue $g_1 = \nu_1 + \sigma^2 = (\sum_{i=1}^{N} |h_{i1}|^2) + \sigma^2$. The density of $g_1 - \sigma^2$ is a complex chi-square distribution of N degrees of freedom (denoted χ^2_N), which up to a scaling factor 2 is equivalent to a real χ^2_{2N} distribution. Hence, the eigenvalue distribution of $\mathbf{\Sigma}$, defined on $\mathbb{R}^{+N}$, reads

$$P_{\mathbf{G}}(\mathbf{G}) = \frac{1}{N}(g_1 - \sigma^2)_+^{N-1} \frac{e^{-(g_1-\sigma^2)}}{(N-1)!} \prod_{i=2}^{N} \delta(g_i - \sigma^2).$$

From the model $\mathcal{H}_1$, $\mathbf{Y}$ is distributed as correlated Gaussian, as follows

$$P_{\mathbf{Y}|\mathbf{\Sigma},I_1}(\mathbf{Y}, \mathbf{\Sigma}) = \frac{1}{\pi^{MN} \det(\mathbf{G})^M} e^{-\mathrm{tr}\left(\mathbf{Y}\mathbf{Y}^{\mathsf{H}}\mathbf{U}\mathbf{G}^{-1}\mathbf{U}^{\mathsf{H}}\right)},$$

where I_k denotes the prior information at the receiver "$\mathcal{H}_1$ and $K = k$".

Since $\mathbf{H}$ is unknown, we need to integrate out all possible linear filters for the transmission model under $\mathcal{H}_1$ over the probability space of $N \times K$ matrices with Gaussian i.i.d. distribution. From the unitarily invariance of Gaussian i.i.d. random matrices, this is equivalent to integrating out all possible covariance matrices $\mathbf{\Sigma}$ over the space of such nonnegative definite Hermitian matrices, as follows

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\mathbf{\Sigma}} P_{\mathbf{Y}|\mathbf{\Sigma},\mathcal{H}_1}(\mathbf{Y}, \mathbf{\Sigma}) P_{\mathbf{\Sigma}}(\mathbf{\Sigma}) d\mathbf{\Sigma}.$$

Eventually, after complete integration calculus given in the proof below, the Neyman-Pearson decision ratio (25) for the single-input multiple-output channel takes an explicit expression, given by the following theorem.

Theorem 0.18. *The Neyman-Pearson test ratio $C_{\mathbf{Y}}(\mathbf{Y})$ for the presence of data reads*

$$C_{\mathbf{Y}}(\mathbf{Y}) = \frac{1}{N} \sum_{l=1}^{N} \frac{\sigma^{2(N+M-1)} e^{\sigma^2 + \frac{\lambda_l}{\sigma^2}}}{\prod_{\substack{i=1 \\ i\neq l}}^{N} (\lambda_l - \lambda_i)} J_{N-M-1}(\sigma^2, \lambda_l), \tag{28}$$

with $\lambda_1, \ldots, \lambda_N$ the eigenvalues of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ and where

$$J_k(x, y) \triangleq \int_x^{+\infty} t^k e^{-t - \frac{y}{t}} dt.$$

The proof of Theorem 0.18 is provided below. Among the interesting features of (28), note that the Neyman-Pearson test does only depend on the eigenvalues of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$. This suggests that the eigenvectors of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ do not provide any information regarding the presence of data. The essential reason is that, both under $\mathcal{H}_0$ and $\mathcal{H}_1$, the eigenvectors of $\mathbf{Y}$ are isotropically distributed on the unit N-dimensional complex sphere due to the Gaussian assumptions made here. As such, a given realization of the eigenvectors of $\mathbf{Y}$ does indeed not carry any relevant information to the hypothesis test. The Gaussian assumption for $\mathbf{H}$ brought by the maximum

entropy principle is in fact essential here. Note however that (28) is not reduced to a function of the sum $\sum_i \lambda_i$ of the eigenvalues, as suggests the classical energy detector.

On the practical side, note that the integral $J_k(x,y)$ does not take a closed-form expression, but for $x=0$, see e.g., pp. 561 of [34]. This is rather inconvenient for practical purposes, since $J_k(x,y)$ must either be evaluated every time, or be tabulated. It is also difficult to get any insight on the performance of such a detector for different values of σ^2, N and K. We provide hereafter a proof of Theorem 0.18, in which classical multi-dimensional integration techniques are introduced. In particular, the tools introduced in Section 0.2 will be shown to be key ingredients of the derivation.

Proof. We start by noticing that $\mathbf{H}$ is Gaussian and therefore that the joint density of its entries is invariant by left and right unitary products. As a consequence, the distribution of the matrix $\mathbf{\Sigma} = \mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2\mathbf{I}$ is unitarily invariant. This allows us to write, similar to [35],

$$\begin{aligned} P_{\mathbf{Y}|I_1}(\mathbf{Y}) &= \int_{\mathbf{\Sigma}} P_{\mathbf{Y}|\mathbf{\Sigma},\mathcal{H}_1}(\mathbf{Y},\mathbf{\Sigma}) P_{\mathbf{\Sigma}}(\mathbf{\Sigma}) d\mathbf{\Sigma} \\ &= \int_{\mathcal{U}(N)\times(\mathbb{R}^+)^N} P_{\mathbf{Y}|\mathbf{\Sigma},\mathcal{H}_1}(\mathbf{Y},\mathbf{\Sigma}) P_{\mathbf{G}}(\mathbf{G}) d\mathbf{U} d\mathbf{G} \\ &= \int_{\mathcal{U}(N)\times\mathbb{R}^+} P_{\mathbf{Y}|\mathbf{\Sigma},\mathcal{H}_1}(\mathbf{Y},\mathbf{\Sigma}) P_{g_1}(g_1) d\mathbf{U} dg_1 \end{aligned}$$

with $\mathcal{U}(N)$ the space of $N\times N$ unitary matrices and $\mathbf{\Sigma} = \mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}}$.

The latter can further be equated to

$$P_{\mathbf{Y}|I_1}(\mathbf{Y}) = \int_{\mathcal{U}(N)\times\mathbb{R}^+} \frac{e^{-\mathrm{tr}\left(\mathbf{Y}\mathbf{Y}^{\mathsf{H}}\mathbf{U}\mathbf{G}^{-1}\mathbf{U}^{\mathsf{H}}\right)}}{\pi^{NM}\det(\mathbf{G})^M} (g_1-\sigma^2)_+^{N-1} \frac{e^{-(g_1-\sigma^2)}}{N!} d\mathbf{U} dg_1$$

with $(x)_+ \triangleq \max(x,0)$ here.

To go further, we use the Harish-Chandra identity provided in Theorem 0.3. Denoting $\Delta(\mathbf{Z})$ the Vandermonde determinant of matrix $\mathbf{Z}\in\mathbb{C}^{N\times N}$ with eigenvalues $z_1\leq\ldots\leq z_N$

$$\Delta(\mathbf{Z}) \triangleq \prod_{i>j}(z_i - z_j), \tag{29}$$

the likelihood $P_{\mathbf{Y}|I_1}(\mathbf{Y})$ further develops as

$$\begin{aligned} &P_{\mathbf{Y}|I_1}(\mathbf{Y}) \\ &= \lim_{g_2,\ldots,g_N\to\sigma^2} \frac{e^{\sigma^2}(-1)^{\frac{N(N-1)}{2}}\prod_{j=1}^{N-1} j!}{\pi^{MN}\sigma^{2M(N-1)}N!} \int_{\sigma^2}^{+\infty} \frac{1}{g_1^M}(g_1-\sigma^2)^{N-1} e^{-g_1} \frac{\det\left(\left\{e^{-\frac{\lambda_i}{g_j}}\right\}\right)}{\Delta(\mathbf{Y}\mathbf{Y}^{\mathsf{H}})\Delta(\mathbf{G}^{-1})} dg_1 \\ &= \lim_{g_2,\ldots,g_N\to\sigma^2} \frac{e^{\sigma^2}\prod_{j=1}^{N-1} j!}{\pi^{MN}\sigma^{2M(N-1)}N!} \int_{\sigma^2}^{+\infty} \frac{1}{g_1^M}(g_1-\sigma^2)^{N-1} e^{-g_1} \det\left(\mathbf{G}^{N-1}\right) \frac{\det\left(\left\{e^{-\frac{\lambda_i}{g_j}}\right\}\right)}{\Delta(\mathbf{Y}\mathbf{Y}^{\mathsf{H}})\Delta(\mathbf{G})} dg_1 \end{aligned} \tag{30}$$

$$= \lim_{g_2,\ldots,g_N\to\sigma^2} \frac{e^{\sigma^2}\sigma^{2(N-1)(N-M-1)}\prod_{j=1}^{N-1} j!}{\pi^{MN}N!} \int_{\sigma^2}^{+\infty} {g_1}^{N-M-1}(g_1-\sigma^2)^{N-1} e^{-g_1} \frac{\det\left(\left\{e^{-\frac{\lambda_i}{g_j}}\right\}\right)}{\Delta(\mathbf{Y}\mathbf{Y}^{\mathsf{H}})\Delta(\mathbf{G})} dg_1 \tag{31}$$

in which we remind that $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$. The equality (30) comes from the fact that $\Delta(\mathbf{G}^{-1}) = (-1)^{N(N+3)/2}\frac{\Delta(\mathbf{G})}{\det(\mathbf{G})^{N-1}}$. Note the trick of replacing the known values of $g_2, \ldots, g_N$ by limits of scalars converging to these known values, which dodges the problem of improper ratios. To derive the explicit limits, we then proceed as follows.

To go further, we need the following result, Lemma 6 of [36].

Theorem 0.19. *Let $f_1, \ldots, f_N$ be a family of infinitely differentiable functions and let $x_1, \ldots, x_N \in \mathbb{R}$. Denote*

$$R(x_1, \ldots, x_N) \triangleq \frac{\det\left(\{f_i(x_j)\}_{i,j}\right)}{\prod_{i>j}(x_i - x_j)}.$$

Then, for $p \leq N$ and for $x_0 \in \mathbb{R}$,

$$\lim_{x_1,\ldots,x_p \to x_0} R(x_1, \ldots, x_N) = \frac{\det\left[f_i(x_0), f_i'(x_0), \ldots, f_i^{(p-1)}(x_0), f_i(x_{p+1}), \ldots, f_i(x_N)\right]}{\prod_{p<j<i}(x_i - x_j)\prod_{i=p+1}^{N}(x_i - x_0)^p \prod_{j=1}^{p-1} j!}.$$

Denoting $\mathbf{y} = (\gamma_1, \ldots, \gamma_{N-1}, \gamma_N) = (g_2, \ldots, g_N, g_1)$ and defining the functions,

$$\begin{aligned} f(x_i, \gamma_j) &\triangleq e^{-\frac{x_i}{\gamma_j}}, \\ f_i(\gamma_j) &\triangleq f(x_i, \gamma_j), \end{aligned}$$

from Theorem 0.19, we obtain

$$\begin{aligned} \lim_{g_2,\ldots,g_N \to \sigma^2} \frac{\det\left(\left\{e^{-\frac{\lambda_i}{g_j}}\right\}_{\substack{1\leq i\leq N \\ 1\leq j\leq N}}\right)}{\Delta(\mathbf{Y}\mathbf{Y}^{\mathsf{H}})\Delta(\mathbf{G})} &= \lim_{\substack{\gamma_1,\ldots,\gamma_{N-1}\to\sigma^2 \\ \gamma_M \to g_1}} (-1)^{N-1}\frac{\det\left(\{f_i(\lambda_j)\}_{i,j}\right)}{\Delta(\mathbf{Y}\mathbf{Y}^{\mathsf{H}})\Delta(\mathbf{G})} \\ &= (-1)^{N-1}\frac{\det\left[f_i(\sigma^2),\ f_i'(\sigma^2), \ldots,\ f^{(N-2)}(\sigma^2),\ f_i(g_1)\right]}{\prod_{i<j}(\lambda_i - \lambda_j)(g_1 - \sigma^2)^{N-1}\prod_{j=1}^{N-2} j!}. \end{aligned}$$

The change of variables led to a switch of one column and explains the $(-1)^{N-1}$ factor appearing when computing the resulting determinant. The partial derivatives of f along the second variable is

$$\begin{aligned} \left(\frac{\partial}{\partial \gamma^k} f\right)_{k\geq 1}(a, b) &= \sum_{m=1}^{k} \frac{(-1)^{k+m}}{b^{m+k}}\binom{m}{k}\frac{(k-1)!}{(m-1)!}a^m e^{-\frac{a}{b}} \\ &\triangleq \kappa_k(a, b)e^{-\frac{a}{b}}. \end{aligned}$$

Back to the full expression of $P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y})$, we then have

$$
\begin{aligned}
&P_{\mathbf{Y}|I_1}(\mathbf{Y}) \\
&= \frac{e^{\sigma^2}\sigma^{2(N-1)(N-M-1)}}{N\pi^{MN}} \int_{\sigma^2}^{+\infty} (-1)^{N-1} g_1^{N-M-1} e^{-g_1} \frac{\det\left[f_i(\sigma^2), f_i'(\sigma^2), \ldots, f^{(N-2)}(\sigma^2), f_i(g_1)\right]}{\prod_{i<j}(\lambda_i - \lambda_j)} dg_1 \\
&= \frac{e^{\sigma^2}\sigma^{2(N-1)(N-M-1)}}{N\pi^{MN}\prod_{i<j}(\lambda_i - \lambda_j)} \\
&\times \int_{\sigma^2}^{+\infty} (-1)^{N-1} g_1^{N-M-1} e^{-g_1} \det \left[\begin{array}{c|c|c} e^{-\frac{x_1}{\sigma^2}} & \left(\kappa_j(\lambda_i,\sigma^2) e^{-\frac{\lambda_j}{\sigma^2}}\right)_{\substack{1\le i\le N \\ 1\le j\le N-2}} & e^{-\frac{\lambda_1}{g_1}} \\ \vdots & & \vdots \\ e^{-\frac{x_N}{\sigma^2}} & & e^{-\frac{\lambda_N}{g_1}} \end{array} \right] dg_1.
\end{aligned}
$$

Before going further, we need the following result, often required in the calculus of marginal eigenvalue distributions for Gaussian matrices.

Lemma 0.2. *For any family* $\{a_1, \ldots, a_N\} \in \mathbb{R}^N$, $N \geq 2$, *and for any* $b \in \mathbb{R}^*$,

$$
\det \left[\begin{array}{c|c} 1 & \\ \vdots & (\kappa_j(a_i,b))_{\substack{1\le i\le N \\ 1\le j\le N-1}} \\ 1 & \end{array} \right] = \frac{1}{b^{N(N-1)}} \prod_{i<j}(a_j - a_i).
$$

This identity follows from the observation that column k of the matrix above is a polynomial of order k. Since summations of linear combinations of the columns do not affect the determinant, each polynomial can be replaced by the monomial of highest order, i.e., $b^{-2(k-1)}a_i^k$ in row i. Extracting the product $1 \cdot b^{-2} \cdots b^{-2(N-1)} = b^{-(N-1)N}$ from the determinant, what remains is the determinant of a Vandermonde matrix based on the vector $a_1, \ldots, a_N$.

By factorizing every row of the matrix by $e^{-\frac{\lambda_j}{\sigma^2}}$ and developing the determinant on the last column, one obtains

$$
\begin{aligned}
&P_{\mathbf{Y}|I_1}(\mathbf{Y}) \\
&= \frac{e^{\sigma^2}\sigma^{2(N-1)(N-M-1)}}{N\pi^{MN}\prod_{i<j}(\lambda_i - \lambda_j)} \int_{\sigma^2}^{+\infty} g_1^{N-M-1} e^{-g_1 - \frac{\sum_{i=1}^N \lambda_i}{\sigma^2}} \sum_{l=1}^{N} \frac{(-1)^{2N+l-1} e^{-\lambda_l\left(\frac{1}{g_1} - \frac{1}{\sigma^2}\right)}}{\sigma^{2(N-1)(N-2)}} \prod_{\substack{i<j \\ i\neq l \\ j\neq l}} (\lambda_i - \lambda_j) dg_1 \\
&= \frac{e^{\sigma^2 - \frac{1}{\sigma^2}\sum_{i=1}^N \lambda_i}}{N\pi^{MN}\sigma^{2(N-1)(M-1)}} \sum_{l=1}^{N} (-1)^{l-1} \int_{\sigma^2}^{+\infty} g_1^{N-M-1} e^{-g_1} \frac{e^{-\lambda_l\left(\frac{1}{g_1} - \frac{1}{\sigma^2}\right)}}{\prod_{i<l}(\lambda_i - \lambda_l)\prod_{i>l}(\lambda_l - \lambda_i)} dg_1 \\
&= \frac{e^{\sigma^2 - \frac{1}{\sigma^2}\sum_{i=1}^N \lambda_i}}{N\pi^{MN}\sigma^{2(N-1)(M-1)}} \sum_{l=1}^{N} \frac{e^{\frac{\lambda_l}{\sigma^2}}}{\prod_{\substack{i=1 \\ i\neq l}}^N (\lambda_l - \lambda_i)} \int_{\sigma^2}^{+\infty} g_1^{N-M-1} e^{-\left(g_1 + \frac{\lambda_l}{g_1}\right)} dg_1,
\end{aligned}
$$

which finally gives

$$
P_{\mathbf{Y}|I_1}(\mathbf{Y}) = \frac{e^{\sigma^2 - \frac{1}{\sigma^2}\sum_{i=1}^N \lambda_i}}{N\pi^{MN}\sigma^{2(N-1)(M-1)}} \sum_{l=1}^{N} \frac{e^{\frac{\lambda_l}{\sigma^2}}}{\prod_{\substack{i=1 \\ i\neq l}}^N (\lambda_l - \lambda_i)} J_{N-M-1}(\sigma^2, \lambda_l),
$$

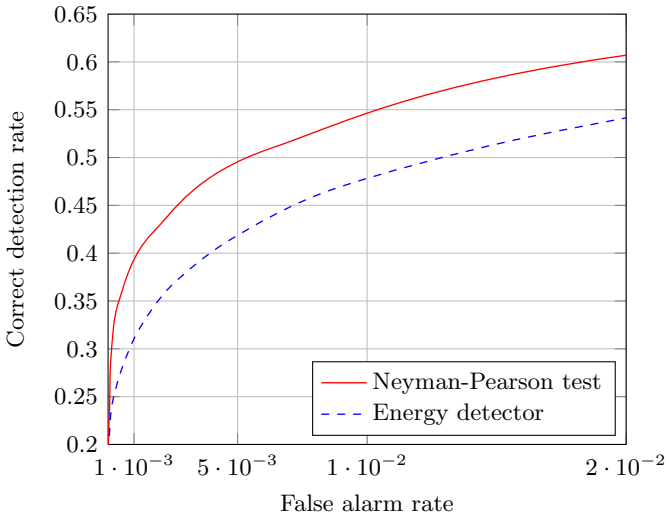

Figure 6: ROC curve for single-source detection, $K = 1$, $N = 4$, $M = 8$, SNR $= -3$ dB, FAR range of practical interest.

where

$$J_k(x, y) = \int_x^{+\infty} t^k e^{-t-\frac{y}{t}} dt = 2y^{\frac{k+1}{2}} K_{-k-1}(2\sqrt{y}) - \int_0^x t^k e^{-t-\frac{y}{t}} dt$$

and K_n denotes the modified Bessel function of the second kind. □

The scenario where $K \geq 1$ unfolds similarly. The final theorem can be found in [2]. The receiver operating characteristic (ROC) curve of the Neyman-Pearson test against that of the energy detector is provided in Figure 6 for $N = 4$, $M = 8$ and $\sigma^2 = 3$ dBm. We observed a significant performance gain in terms of detection rate incurred by the Neyman-Pearson test compared to the classical energy detector.

This completes this section on hypothesis testing. In the following section, we go beyond the hypothesis test and move to the question of parameter inference in a slightly more complex data plus noise model than above.

0.5.2 Parameter estimation

We consider a similar scenario as in the previous section, where now the K sources use different transmit powers $P_1, \ldots, P_K$, which the receiver is entitled to infer from successive observations.

Consider K data sources which are transmitting data simultaneously. Transmitter $k \in \{1, \ldots, K\}$ has power P_k and has space dimension n_k, e.g. is composed of n_k antennas. We denote $n \triangleq \sum_{k=1}^{K} n_k$ the total number of transmit dimensions. Consider also a sink or receiver with space dimension N, $N > n$. Denote $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$ the multi-dimensional filter matrix between transmitter k and the receiver. We assume that the entries of $\sqrt{N}\mathbf{H}_k$ are independent and identically distributed with zero mean, unit variance and finite fourth order moment. At time instant m, transmitter k emits the signal $\mathbf{x}_k^{(m)} \in \mathbb{C}^{n_k}$, with entries assumed to be independent, independent along m, k, identically distributed along m, and have all zero mean, unit

variance and finite fourth order moment (the $\mathbf{x}_k^{(m)}$ need not be identically distributed along k). Assume further that at time instant m the receive signal is impaired by additive white Gaussian noise with entries of zero mean and variance σ^2, denoted $\sigma\mathbf{w}^{(m)} \in \mathbb{C}^N$. At time m, the receiver therefore senses the signal $\mathbf{y}^{(m)} \in \mathbb{C}^N$ defined as

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \sqrt{P_k}\mathbf{H}_k\mathbf{x}_k^{(m)} + \sigma\mathbf{w}^{(m)}.$$

Assuming the filter coefficients are constant over at least M consecutive sampling periods, by concatenating M successive signal realizations into $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}] \in \mathbb{C}^{N\times M}$, we have

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{P_k}\mathbf{H}_k\mathbf{X}_k + \sigma\mathbf{W},$$

where $\mathbf{X}_k = [\mathbf{x}_k^{(1)}, \ldots, \mathbf{x}_k^{(M)}] \in \mathbb{C}^{n_k\times M}$, for every k, and $\mathbf{W} = [\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(M)}] \in \mathbb{C}^{N\times M}$. This can be further rewritten as

$$\mathbf{Y} = \mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W}, \tag{32}$$

where $\mathbf{P} \in \mathbb{R}^{n\times n}$ is diagonal with first n_1 entries P_1, subsequent n_2 entries P_2, etc. and last n_K entries P_K, $\mathbf{H} = [\mathbf{H}_1, \ldots, \mathbf{H}_K] \in \mathbb{C}^{N\times n}$ and $\mathbf{X} = [\mathbf{X}_1^{\mathsf{T}}, \ldots, \mathbf{X}_K^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{C}^{n\times M}$. By convention, we assume $P_1 \leq \ldots \leq P_K$.

Our objective is to infer the values of the powers $P_1, \ldots, P_K$ from the realization of a single random matrix $\mathbf{Y}$. This is successively performed from different approaches in the following sections. We first consider the conventional approach that assumes n small, N much larger than n, and M much larger than N. This will lead to a simple although largely biased estimation algorithm. This algorithm will be improved using Stieltjes transform approaches in the same spirit as in Section 0.4.

Conventional approach

The first approach assumes numerous sensors in order to have much diversity in the observation vectors, as well as an even larger number of observations so to create an averaging effect on the incoming random data. In this situation, let us rewrite (32) under the form

$$\mathbf{Y} = \begin{pmatrix}\mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I}_N\end{pmatrix}\begin{pmatrix}\mathbf{X}\\ \mathbf{W}\end{pmatrix}. \tag{33}$$

We shall denote $\lambda_1 \leq \ldots \leq \lambda_N$ the ordered eigenvalues of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ (the non-zero eigenvalues of which are almost surely different).

Appending $\mathbf{Y} \in \mathbb{C}^{N\times M}$ into the larger matrix $\underline{\mathbf{Y}} \in \mathbb{C}^{(N+n)\times M}$

$$\underline{\mathbf{Y}} = \begin{pmatrix}\mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I}_N\\ 0 & 0\end{pmatrix}\begin{pmatrix}\mathbf{X}\\ \mathbf{W}\end{pmatrix},$$

we recognize that, conditioned on $\mathbf{H}$, $\frac{1}{M}\underline{\mathbf{Y}}\underline{\mathbf{Y}}^{\mathsf{H}}$ is a *sample covariance matrix*, for which the *population covariance matrix* is

$$\mathbf{T} \triangleq \begin{pmatrix}\mathbf{H}\mathbf{P}\mathbf{H}^{\mathsf{H}} + \sigma^2\mathbf{I}_N & 0\\ 0 & 0\end{pmatrix}$$

and the random matrix

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{W} \end{pmatrix}$$

has independent (non-necessarily identically distributed) entries with zero mean and unit variance. The population covariance matrix $\mathbf{T}$, whose upper left entries also form a matrix unitarily equivalent to a sample covariance matrix, clearly has an almost sure limit spectral distribution as N grows large for fixed or slowly growing n. Extending Theorem 0.9 and Theorem 0.11 to $c = 0$ and applying them twice (once for the population covariance matrix $\mathbf{T}$ and once for $\frac{1}{M}\underline{\mathbf{YY}}^{\mathsf{H}}$), we finally have that, as $M, N, n \to \infty$ with $M/N \to \infty$ and $N/n \to \infty$, the distribution of the largest n eigenvalues of $\frac{1}{M}\mathbf{YY}^{\mathsf{H}}$ is asymptotically almost surely composed of a mass $\sigma^2 + P_1$ of weight $\lim n_1/n$, a mass $\sigma^2 + P_2$ of weight $\lim n_2/n$, etc. and a mass $\sigma^2 + P_K$ of weight $\lim n_K/n$. As for the distribution of the smallest $N - n$ eigenvalues of $\frac{1}{M}\mathbf{YY}^{\mathsf{H}}$, it converges to a single mass in σ^2.

If σ^2 is a priori known, a rather trivial estimator of P_k is then given by

$$\frac{1}{n_k} \sum_{i \in \mathcal{N}_k} (\lambda_i - \sigma^2),$$

where $\mathcal{N}_k = \{\sum_{j=1}^{k-1} n_j + 1, \ldots, \sum_{j=1}^{k} n_j\}$ and we recall that $\lambda_1 \leq \ldots \leq \lambda_N$ are the ordered eigenvalues of $\frac{1}{M}\mathbf{YY}^{\mathsf{H}}$.

This means in practice that P_K is asymptotically well approximated by the averaged value of the n_K largest eigenvalues of $\frac{1}{M}\mathbf{YY}^{\mathsf{H}}$, P_{K-1} is well approximated by the averaged value of the n_{K-1} eigenvalues before that, etc. This also assumes that σ^2 is perfectly known at the receiver. If it were not, observe that the averaged value of the $N - n$ smallest eigenvalues of $\frac{1}{M}\mathbf{YY}^{\mathsf{H}}$ is a consistent estimate for σ^2. This therefore leads to the second estimator $\hat{P}_k^{\infty}$ for P_k, that will constitute our reference estimator,

$$\hat{P}_k^{\infty} = \frac{1}{n_k} \sum_{i \in \mathcal{N}_k} \left(\lambda_i - \hat{\sigma}^2\right),$$

where

$$\hat{\sigma}^2 = \frac{1}{N-n} \sum_{i=1}^{N-n} \lambda_i.$$

Incidentally, although not derived on purpose, the refined (n, N, M)-consistent estimator of Section 0.5.2 will appear not to depend on a prior knowledge of σ^2. Note that the estimation of P_k only relies on n_k contiguous eigenvalues of $\frac{1}{M}\mathbf{YY}^{\mathsf{H}}$, which suggests that the other eigenvalues are asymptotically uncorrelated from these. It will turn out that the improved (n, N, M)-consistent estimator does take into account all eigenvalues for each k, in a certain manner.

The Stieltjes transform method

The Stieltjes transform approach relies heavily on the techniques from Mestre, established in [25] and introduced in Section 0.4. The main strategy is the following.

- We first need to study the asymptotic spectrum of $\mathbf{B}_N$, as all system dimensions (N, n, M) grow large (remember that K is fixed). For this, we will proceed to

  - determine the almost sure l.s.d. of $\mathbf{B}_N$. Practically, this will allow us to connect the asymptotic spectrum of $\mathbf{B}_N$ to the spectrum of $\mathbf{P}$,
  - study the exact separation of the eigenvalues of $\mathbf{B}_N$ in clusters of eigenvalues. This is necessary first to determine whether the coming step of complex integration is possible and second to determine a well-chosen integration contour for the estimation of every P_k.

- Then, we will write P_k under the form of a complex integral of a functional of the spectrum of $\mathbf{P}$ over this well-chosen contour. Since the spectrum of $\mathbf{P}$ can be linked to that of $\mathbf{B}_N$ (at least asymptotically) through the previous step, a change of variable will allow us to rewrite P_k under the form of an integral of some functional of the l.s.d. of $\mathbf{B}_N$. This point is the key step in our derivation, where P_k is now connected to the observation matrix $\mathbf{Y}$ (although only in the asymptotic sense).

- Finally, the estimate $\hat{P}_k$ of P_k will be computed from the previous step by replacing the l.s.d. of $\mathbf{B}_N$ by its e.s.d., i.e., by the truly observed eigenvalues of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$, in the expression relating P_k to the l.s.d. of $\mathbf{B}_N$.

We provide here only the study of the limiting spectrum of $\mathbf{B}_N$ and of the final estimator, taking for granted the result of the contour integration study, which is more technical and is properly detailed in [24].

Limiting spectrum of $\mathbf{B}_N$ In this section, we prove the following result

Theorem 0.20. *Let $\mathbf{B}_N = \frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$, with $\mathbf{Y}$ defined as in (32). Then, for M, N, n growing large with limit ratios $M/N \to c$, $N/n_k \to c_k$, $0 < c, c_1, \ldots, c_K < \infty$, the empirical spectral distribution $F^{\mathbf{B}_N}$ of $\mathbf{B}_N$ converges almost surely to the distribution function F, whose Stieltjes transform $m_F(z)$ satisfies, for $z \in \mathbb{C}^+$,*

$$m_F(z) = cm_{\underline{F}}(z) + (c-1)\frac{1}{z}, \tag{34}$$

where $m_{\underline{F}}(z)$ is the unique solution with positive imaginary part of the implicit equation in $m_{\underline{F}}$,

$$\frac{1}{m_{\underline{F}}} = -\sigma^2 + \frac{1}{f} - \sum_{k=1}^{K} \frac{1}{c_k} \frac{P_k}{1 + P_k f} \tag{35}$$

in which we denoted f the value

$$f = (1-c)m_{\underline{F}} - czm_{\underline{F}}^2.$$

Proof. First remember that the matrix $\mathbf{Y}$ in (32) can be extended into the larger sample covariance matrix $\underline{\mathbf{Y}} \in \mathbb{C}^{(N+n)\times M}$

$$\underline{\mathbf{Y}} = \begin{pmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I}_N \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{W} \end{pmatrix}.$$

From Theorem 0.9, since $\mathbf{H}$ has independent entries with finite fourth order moment, we have that the e.s.d. of $\mathbf{H}\mathbf{P}\mathbf{H}^{\mathsf{H}}$ converges weakly and almost surely to a limit distribution G as

$N, n_1, \ldots, n_K \to \infty$ with $N/n_k \to c_k > 0$. For $z \in \mathbb{C}^+$, the Stieltjes transform $m_G(z)$ of G is the unique solution with positive imaginary part of the equation in m_G,

$$z = -\frac{1}{m_G} + \sum_{k=1}^{K} \frac{1}{c_k} \frac{P_k}{1 + P_k m_G}. \tag{36}$$

The almost sure convergence of the e.s.d. of $\mathbf{HPH}^{\mathsf{H}}$ ensures the almost sure convergence of the e.s.d. of the matrix $\begin{pmatrix} \mathbf{HPH}^{\mathsf{H}} + \sigma^2 \mathbf{I}_N & 0 \\ 0 & 0 \end{pmatrix}$. Since $m_G(z)$ evaluated at $z \in \mathbb{C}^+$ is the Stieltjes transform of the l.s.d. of $\mathbf{HPH}^{\mathsf{H}} + \sigma^2 \mathbf{I}_N$ evaluated at $z + \sigma^2$, adding n zero eigenvalues, we finally have that the e.s.d. of $\begin{pmatrix} \mathbf{HPH}^{\mathsf{H}} + \sigma^2 \mathbf{I}_N & 0 \\ 0 & 0 \end{pmatrix}$ tends almost surely to a distribution H whose Stieltjes transform $m_H(z)$ satisfies

$$m_H(z) = \frac{c_0}{1 + c_0} m_G(z - \sigma^2) - \frac{1}{1 + c_0} \frac{1}{z}, \tag{37}$$

for $z \in \mathbb{C}^+$, where we denoted c_0 the limit of the ratio N/n, i.e., $c_0 = (c_1^{-1} + \ldots + c_K^{-1})^{-1}$.

As a consequence, the sample covariance matrix $\frac{1}{M} \underline{\mathbf{Y}} \underline{\mathbf{Y}}^{\mathsf{H}}$ has a population covariance matrix which is not deterministic but whose e.s.d. has an almost sure limit H for increasing dimensions. Since $\mathbf{X}$ and $\mathbf{W}$ have entries with finite fourth order moment, we can again apply Theorem 0.9 and we have that the e.s.d. of $\underline{\mathbf{B}}_N \triangleq \frac{1}{M} \underline{\mathbf{Y}}^{\mathsf{H}} \underline{\mathbf{Y}}$ converges almost surely to the limit $\underline{F}$ whose Stieltjes transform $m_{\underline{F}}(z)$ is the unique solution in $\mathbb{C}^+$ of the equation in $m_{\underline{F}}$

$$\begin{aligned} z &= -\frac{1}{m_{\underline{F}}} + \frac{1}{c} \left(1 + \frac{1}{c_0}\right) \int \frac{t}{1 + t m_{\underline{F}}} dH(t) \\ &= -\frac{1}{m_{\underline{F}}} + \frac{1 + \frac{1}{c_0}}{c m_{\underline{F}}} \left[1 - \frac{1}{m_{\underline{F}}} m_H \left(-\frac{1}{m_{\underline{F}}}\right)\right] \end{aligned} \tag{38}$$

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

For $z \in \mathbb{C}^+$, $m_{\underline{F}}(z) \in \mathbb{C}^+$. Therefore $-1/m_{\underline{F}}(z) \in \mathbb{C}^+$ and one can evaluate (37) at $-1/m_{\underline{F}}(z)$. Combining (37) and (38), we then have

$$z = -\frac{1}{c} \frac{1}{m_{\underline{F}}(z)^2} m_G \left(-\frac{1}{m_{\underline{F}}(z)} - \sigma^2\right) + \left(\frac{1}{c} - 1\right) \frac{1}{m_{\underline{F}}(z)}, \tag{39}$$

where, according to (36), $m_G(-1/m_{\underline{F}}(z) - \sigma^2)$ satisfies

$$\frac{1}{m_{\underline{F}}(z)} = -\sigma^2 + \frac{1}{m_G(-\frac{1}{m_{\underline{F}}(z)} - \sigma^2)} - \sum_{k=1}^{K} \frac{1}{c_k} \frac{P_k}{1 + P_k m_G(-\frac{1}{m_{\underline{F}}(z)} - \sigma^2)}. \tag{40}$$

Together with (39), this is exactly (35), with $f(z) = m_G(-\frac{1}{m_{\underline{F}}(z)} - \sigma^2) = (1 - c) m_{\underline{F}}(z) - c z m_{\underline{F}}(z)^2$.

Since the eigenvalues of the matrices $\mathbf{B}_N$ and $\underline{\mathbf{B}}_N$ only differ by $M - N$ zeros, we also have that the Stieltjes transform $m_F(z)$ of the l.s.d. of $\mathbf{B}_N$ satisfies

$$m_F(z) = c m_{\underline{F}}(z) + (c - 1) \frac{1}{z}. \tag{41}$$

This completes the proof of Theorem 0.20. □

For further usage, notice here that (41) provides a simplified expression for $m_G(-1/m_{\underline{F}}(z) - \sigma^2)$. Indeed we have,

$$m_G(-1/m_{\underline{F}}(z) - \sigma^2) = -z m_F(z) m_{\underline{F}}(z). \tag{42}$$

Therefore, the support of the (almost sure) l.s.d. F of $\mathbf{B}_N$ can be evaluated as follows: for any $z \in \mathbb{C}^+$, $m_F(z)$ is given by (34), in which $m_{\underline{F}}(z)$ is solution of (35); the inverse Stieltjes transform formula (4) allows then to evaluate F from $m_F(z)$, for values of z spanning over the set $\{z = x + iy, x > 0\}$ and y small.

Multi-source power inference In the following, we finally prove the main result of this section, which provides the G-estimator $\hat{P}_1, \ldots, \hat{P}_K$ of the transmit powers $P_1, \ldots, P_K$.

Theorem 0.21. *Let $\mathbf{B}_N \in \mathbb{C}^{N \times N}$ be defined as $\mathbf{B}_N = \frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ with $\mathbf{Y}$ defined as in (32), and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$, $\lambda_1 \leq \ldots \leq \lambda_N$, be the vector of the ordered eigenvalues of $\mathbf{B}_N$. Further assume that the limiting ratios c_0, $c_1, \ldots, c_K$, c and $\mathbf{P}$ are such that the cluster mapped to P_k in $\mathbf{B}_N$ does not map another P_i, $i \neq k$. Then, as N, n, M grow large, we have*

$$\hat{P}_k - P_k \xrightarrow{\text{a.s.}} 0,$$

where the estimate $\hat{P}_k$ is given by

- *if $M \neq N$,*

$$\hat{P}_k = \frac{NM}{n_k(M-N)} \sum_{i \in \mathcal{N}_k} (\eta_i - \mu_i),$$

- *if $M = N$,*

$$\hat{P}_k = \frac{N}{n_k(N-n)} \sum_{i \in \mathcal{N}_k} \left(\sum_{j=1}^{N} \frac{\eta_i}{(\lambda_j - \eta_i)^2} \right)^{-1},$$

in which $\mathcal{N}_k = \{\sum_{i=1}^{k-1} n_i + 1, \ldots, \sum_{i=1}^{k} n_i\}$, $\eta_1 \leq \ldots \leq \eta_N)$ are the ordered eigenvalues of the matrix $\operatorname{diag}(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$ and $\mu_1 \leq \ldots \leq \mu_N$ are the ordered eigenvalues of the matrix $\operatorname{diag}(\boldsymbol{\lambda}) - \frac{1}{M}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$.

Remark 0.2. *We immediately notice that, if $N < n$, the powers $P_1, \ldots, P_l$, with l the largest integer such that $N - \sum_{i=l}^{K} n_i < 0$, cannot be estimated since clusters may be empty. The case $N \leq n$ turns out to be of no practical interest as clusters always merge and no consistent estimate of either P_i can be described.*

Proof. The approach pursued to prove Theorem 0.21 relies strongly on the original idea of [23], which was detailed for the case of sample covariance matrices in Section 0.4. From Cauchy's integration formula,

$$\begin{aligned} P_k &= c_k \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \frac{1}{c_k} \frac{\omega}{P_k - \omega} d\omega \\ &= c_k \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \sum_{r=1}^{K} \frac{1}{c_r} \frac{\omega}{P_r - \omega} d\omega \end{aligned} \tag{43}$$

for any negatively oriented contour $\mathcal{C}_k \subset \mathbb{C}$, such that P_k is contained in the surface described by the contour, while for every $i \neq k$, P_i is outside this surface. The strategy is very similar to that used for the sample covariance matrix case in Section 0.4. It comes as follows: we first propose a convenient integration contour $\mathcal{C}_k$ which is parametrized by a functional of the Stieltjes transform $m_F(z)$ of the l.s.d. of $\mathbf{B}_N$. We proceed to a variable change in (43) to express P_k as a function of $m_F(z)$. We then evaluate the complex integral resulting from replacing the limiting $m_F(z)$ in (43) by its empirical counterpart $\hat{m}_F(z) = \frac{1}{N}\mathrm{tr}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$. This new integral, whose value we name $\hat{P}_k$, is shown to be almost surely equal to P_k in the large N limit. It then suffices to evaluate $\hat{P}_k$, which is just a matter of residue calculus.

Similar to Section 0.4, it turns out that the clusters generated in the spectrum of $\mathbf{B}_N$ can be mapped to one or many power values P_k. In what follows, we assume that the clusters are disjoint so that no holomorphicity problem arises. We can prove the following. There exist $x_{k_F}^{(l)}$ and $x_{k_F}^{(r)}$ outside the support of F, on either side of cluster k_F (i.e., the cluster in F that is uniquely mapped to P_k) such that $m_{\underline{F}}(z)$ has limits $m_{\underline{F},k_G}^{(l)} \triangleq m_{\underline{F}}^{\circ}(x_{k_F}^{(l)})$ and $m_{\underline{F},k_G}^{(r)} \triangleq m_{\underline{F}}^{\circ}(x_{k_F}^{(r)})$, as $z \to x_{k_F}^{(l)}$ and $z \to x_{k_F}^{(r)}$, respectively, with $m_{\underline{F}}^{\circ}$ the analytic extension of $m_{\underline{F}}$ in the points $x_{k_F}^{(l)} \in \mathbb{R}$ and $x_{k_F}^{(r)} \in \mathbb{R}$. These limits $m_{\underline{F},k_G}^{(l)}$ and $m_{\underline{F},k_G}^{(r)}$ are on either side of cluster k_G (i.e. the cluster in G mapped uniquely to P_k) in the support of $-1/H$, and therefore $-1/m_{\underline{F},k_G}^{(l)} - \sigma^2$ and $-1/m_{\underline{F},k_G}^{(l)} - \sigma^2$ are on either side of cluster k_G in the support of G.

Consider any continuously differentiable complex path $\Gamma_{F,k}$ with endpoints $x_{k_F}^{(l)}$ and $x_{k_F}^{(r)}$, and interior points of positive imaginary part. We define the contour $\mathcal{C}_{F,k}$ as the union of $\Gamma_{F,k}$ oriented from $x_{k_F}^{(l)}$ to $x_{k_F}^{(r)}$ and its complex conjugate $\Gamma_{F,k}^*$ oriented backwards from $x_{k_F}^{(r)}$ to $x_{k_F}^{(l)}$. The contour $\mathcal{C}_{F,k}$ is clearly continuous and piecewise continuously differentiable. Also, the support of cluster k_F in $\underline{F}$ is completely inside $\mathcal{C}_{F,k}$, while the supports of the neighboring clusters are away from $\mathcal{C}_{F,k}$. The support of cluster k_G in H is then inside $-1/m_{\underline{F}}(\mathcal{C}_{F,k})$,[6] and therefore the support of cluster k_G in G is inside $\mathcal{C}_{G,k} \triangleq -1/m_{\underline{F}}(\mathcal{C}_{F,k}) - \sigma^2$. Since $m_{\underline{F}}$ is continuously differentiable on $\mathbb{C} \setminus \mathbb{R}$ (it is in fact holomorphic there [22]) and has limits in $x_{k_F}^{(l)}$ and $x_{k_F}^{(r)}$, $\mathcal{C}_{G,k}$ is also continuous and piecewise continuously differentiable. Going one last step in this process, we finally have that P_k is inside the contour $\mathcal{C}_k \triangleq -1/m_G(\mathcal{C}_{G,k})$, while P_i, for all $i \neq k$, is outside $\mathcal{C}_k$. Since m_G is also holomorphic on $\mathbb{C} \setminus \mathbb{R}$ and has limits in $-1/m_{\underline{F}}^{\circ}(x_{k_F}^{(l)}) - \sigma^2$ and $-1/m_{\underline{F}}^{\circ}(x_{k_F}^{(r)}) - \sigma^2$, $\mathcal{C}_k$ is a continuous and piecewise continuously differentiable complex path, which is sufficient to perform complex integration [31].

Recall now that P_k was defined as

$$P_k = c_k \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \sum_{r=1}^{K} \frac{1}{c_r} \frac{\omega}{P_r - \omega} d\omega.$$

[6] we slightly abuse notations here and should instead say that the support of cluster k_G in H is inside the contour described by the image by $-1/m_{\underline{F}}$ of the restriction to $\mathbb{C}^+$ and $\mathbb{C}^-$ of $\mathcal{C}_{F,k}$, continuously extended to $\mathbb{R}$ in the points $-1/m_{\underline{F},k_G}^{(l)}$ and $-1/m_{\underline{F},k_G}^{(r)}$.

With the variable change $\omega = -1/m_G(t)$, this becomes

$$\begin{aligned} P_k &= \frac{c_k}{2\pi i} \oint_{\mathcal{C}_{G,k}} \sum_{r=1}^{K} \frac{1}{c_r} \frac{-1}{1+P_r m_G(t)} \frac{m_G'(t)}{m_G(t)^2} dt \\ &= \frac{c_k}{2\pi i} \oint_{\mathcal{C}_{G,k}} \left(m_G(t) \left[-\frac{1}{m_G(t)} + \sum_{r=1}^{K} \frac{1}{c_r} \frac{P_r}{1+P_r m_G(t)} \right] + \frac{c_0 - 1}{c_0} \right) \frac{m_G'(t)}{m_G(t)^2} dt. \end{aligned}$$

From Equation (36), this simplifies into

$$P_k = \frac{c_k}{c_0} \frac{1}{2\pi i} \oint_{\mathcal{C}_{G,k}} (c_0 t m_G(t) + c_0 - 1) \frac{m_G'(t)}{m_G(t)^2} dt. \tag{44}$$

Using (39) and proceeding to the further change of variable $t = -1/m_{\underline{F}}(z) - \sigma^2$, (44) becomes

$$\begin{aligned} P_k &= \frac{c_k}{2\pi i} \oint_{\mathcal{C}_{F,k}} \left(\frac{1}{m_{\underline{F}}(z)} + \sigma^2 \right) z m_{\underline{F}}(z) m_F(z) \frac{-m_{\underline{F}}(z) m_F(z) - z m_{\underline{F}}'(z) m_F(z) - z m_{\underline{F}}(z) m_F'(z)}{z^2 m_{\underline{F}}(z)^2 m_F(z)^2} dz \\ &= \frac{c_k}{2\pi i} \oint_{\mathcal{C}_{F,k}} \left(1 + \sigma^2 m_{\underline{F}}(z)\right) \left[-\frac{1}{z m_{\underline{F}}(z)} - \frac{m_{\underline{F}}'(z)}{m_{\underline{F}}(z)^2} - \frac{m_F'(z)}{m_F(z) m_{\underline{F}}(z)} \right] dz. \end{aligned} \tag{45}$$

This whole process of variable changes allows us to describe P_k as a function of $m_F(z)$, the Stieltjes transform of the almost sure limiting spectral distribution of $\mathbf{B}_N$, as $N \to \infty$. It then remains to exhibit a relation between P_k and the empirical spectral distribution of $\mathbf{B}_N$ for finite N. This is to what the subsequent section is dedicated to.

Let us now define $\hat{m}_F(z)$ and $\hat{m}_{\underline{F}}(z)$ as the Stieltjes transforms of the empirical eigenvalue distributions of $\mathbf{B}_N$ and $\underline{\mathbf{B}}_N$, respectively, i.e.,

$$\hat{m}_F(z) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\lambda_i - z} \tag{46}$$

and

$$\hat{m}_{\underline{F}}(z) = \frac{N}{M} \hat{m}_F(z) - \frac{M-N}{M} \frac{1}{z}.$$

Instead of going further with (45), define $\hat{P}_k$, the "empirical counterpart" of P_k, as

$$\hat{P}_k = \frac{n}{n_k} \frac{1}{2\pi i} \oint_{\mathcal{C}_{F,k}} \frac{N}{n} \left(1 + \sigma^2 \hat{m}_{\underline{F}}(z)\right) \left[-\frac{1}{z \hat{m}_{\underline{F}}(z)} - \frac{\hat{m}_{\underline{F}}'(z)}{\hat{m}_{\underline{F}}(z)^2} - \frac{\hat{m}_F'(z)}{\hat{m}_F(z) \hat{m}_{\underline{F}}(z)} \right] dz. \tag{47}$$

The integrand can then be expanded into nine terms, for which residue calculus can easily be performed. Denote first $\eta_1, \ldots, \eta_N$ the N real roots of $\hat{m}_F(z) = 0$ and $\mu_1, \ldots, \mu_N$ the N real roots of $\hat{m}_{\underline{F}}(z) = 0$. We identify three sets of possible poles for the nine aforementioned terms: (i) the set $\{\lambda_1, \ldots, \lambda_N\} \cap [x_{k_F}^{(l)}, x_{k_F}^{(r)}]$, (ii) the set $\{\eta_1, \ldots, \eta_N\} \cap [x_{k_F}^{(l)}, x_{k_F}^{(r)}]$ and (iii) the set

$\{\mu_1,\dots,\mu_N\}\cap[x_{k_F}^{(l)},x_{k_F}^{(r)}]$. For $M\neq N$, the full calculus leads to

$$\hat{P}_k=\frac{NM}{n_k(M-N)}\left[\sum_{\substack{1\le i\le N\\ x_{k_F}^{(l)}\le\eta_i\le x_{k_F}^{(r)}}}\eta_i-\sum_{\substack{1\le i\le N\\ x_{k_F}^{(l)}\le\mu_i\le x_{k_F}^{(r)}}}\mu_i\right]$$
$$+\frac{N}{n_k}\left[\sum_{\substack{1\le i\le N\\ x_{k_F}^{(l)}\le\eta_i\le x_{k_F}^{(r)}}}\sigma^2-\sum_{\substack{1\le i\le N\\ x_{k_F}^{(l)}\le\lambda_i\le x_{k_F}^{(r)}}}\sigma^2\right]+\frac{N}{n_k}\left[\sum_{\substack{1\le i\le N\\ x_{k_F}^{(l)}\le\mu_i\le x_{k_F}^{(r)}}}\sigma^2-\sum_{\substack{1\le i\le N\\ x_{k_F}^{(l)}\le\lambda_i\le x_{k_F}^{(r)}}}\sigma^2\right]. \tag{48}$$

Now, we know from Theorem 0.20 that $\hat{m}_F(z)\xrightarrow{\text{a.s.}}m_F(z)$ and $\hat{m}_{\underline{F}}(z)\xrightarrow{\text{a.s.}}m_{\underline{F}}(z)$ as $N\to\infty$. Observing that the integrand in (47) is uniformly bounded on the compact $\mathcal{C}_{F,k}$, the dominated convergence theorem, Theorem 16.4 of [14], ensures $\hat{P}_k\xrightarrow{\text{a.s.}}P_k$.

To go further, we now need to determine which of $\lambda_1,\dots,\lambda_N$, $\eta_1,\dots,\eta_N$ and $\mu_1,\dots,\mu_N$ lie inside $\mathcal{C}_{F,k}$. It can be proved, by extending Theorem 0.10 and Theorem 0.12 to the current model, that there will be no eigenvalue of $\mathbf{B}_N$ (or $\underline{\mathbf{B}}_N$) outside the support of F, and the number of eigenvalues inside cluster k_F is exactly n_k. Since $\mathcal{C}_{F,k}$ encloses cluster k_F and is away from the other clusters, $\{\lambda_1,\dots,\lambda_N\}\cap[x_{k_F}^{(l)},x_{k_F}^{(r)}]=\{\lambda_i,i\in\mathcal{N}_k\}$ almost surely, for all N large. Also, for any $i\in\{1,\dots,N\}$, it is easy to see from (46) that $\hat{m}_F(z)\to\infty$ when $z\uparrow\lambda_i$ and $\hat{m}_F(z)\to-\infty$ when $z\downarrow\lambda_i$. Therefore $\hat{m}_F(z)=0$ has at least one solution in each interval $(\lambda_{i-1},\lambda_i)$, with $\lambda_0=0$, hence $\mu_1<\lambda_1<\mu_2<\ldots<\mu_N<\lambda_N$. This implies that, if k_0 is the index such that $\mathcal{C}_{F,k}$ contains exactly $\lambda_{k_0},\dots,\lambda_{k_0+(n_k-1)}$, then $\mathcal{C}_{F,k}$ also contains $\{\mu_{k_0+1},\dots,\mu_{k_0+(n_k-1)}\}$. The same result holds for $\eta_{k_0+1},\dots,\eta_{k_0+(n_k-1)}$. When the indexes exist, due to cluster separability, η_{k_0-1} and μ_{k_0-1} belong, for N large, to cluster k_F-1. We are then left with determining whether μ_{k_0} and η_{k_0} are asymptotically found inside $\mathcal{C}_{F,k}$.

For this, we use the same approach as in [23] by noticing that, since 0 is not included in $\mathcal{C}_k$, one has

$$\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{1}{\omega}d\omega=0.$$

Performing the same changes of variables as previously, we have

$$\oint_{\mathcal{C}_{F,k}}\frac{-m_{\underline{F}}(z)m_F(z)-zm'_{\underline{F}}(z)m_F(z)-zm_{\underline{F}}(z)m'_F(z)}{z^2m_{\underline{F}}(z)^2m_F(z)^2}dz=0. \tag{49}$$

For N large, the dominated convergence theorem ensures again that the left-hand side of the (49) is close to

$$\oint_{\mathcal{C}_{F,k}}\frac{-\hat{m}_{\underline{F}}(z)\hat{m}_F(z)-z\hat{m}'_{\underline{F}}(z)\hat{m}_F(z)-z\hat{m}_{\underline{F}}(z)\hat{m}'_F(z)}{z^2\hat{m}_{\underline{F}}(z)^2\hat{m}_F(z)^2}dz. \tag{50}$$

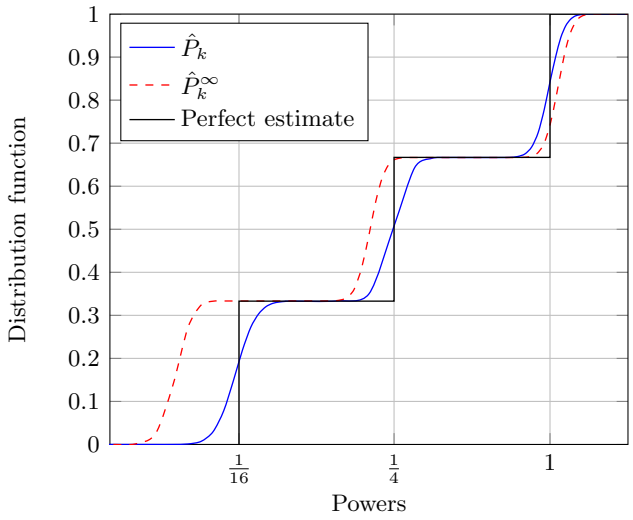

Figure 7: Distribution function of the estimators $\hat{P}_k^{\infty}$ and $\hat{P}_k$ for $k \in \{1,2,3\}$, $P_1 = 1/16$, $P_2 = 1/4$, $P_3 = 1$, $n_1 = n_2 = n_3 = 4$ antennas per user, $N = 24$ sensors, $M = 128$ samples and SNR $= 20$ dB. Optimum estimator shown in dashed lines.

Residue calculus of (50) then leads to

$$\left[\sum_{\substack{1\leq i\leq N \\ \lambda_i \in [x_{k_F}^{(l)}, x_{k_F}^{(r)}]}} 2 - \sum_{\substack{1\leq i\leq N \\ \eta_i \in [x_{k_F}^{(l)}, x_{k_F}^{(r)}]}} 1 - \sum_{\substack{1\leq i\leq N \\ \mu_i \in [x_{k_F}^{(l)}, x_{k_F}^{(r)}]}} 1 \right] \xrightarrow{\text{a.s.}} 0. \tag{51}$$

Since the cardinalities of $\{i, \eta_i \in [x_{k_F}^{(l)}, x_{k_F}^{(r)}]\}$ and $\{i, \mu_i \in [x_{k_F}^{(l)}, x_{k_F}^{(r)}]\}$ are at most n_k, (51) is satisfied only if both cardinalities equal n_k in the limit. As a consequence, $\mu_{k_0} \in [x_{k_F}^{(l)}, x_{k_F}^{(r)}]$ and $\eta_{k_0} \in [x_{k_F}^{(l)}, x_{k_F}^{(r)}]$. For N large, $N \neq M$, this allows us to simplify (48) into

$$\hat{P}_k = \frac{NM}{n_k(M-N)} \sum_{\substack{1\leq i\leq N \\ \lambda_i \in \mathcal{N}_k}} (\eta_i - \mu_i) \tag{52}$$

with probability one. The same reasoning holds for $M = N$. This is our final relation. It now remains to show that the η_i and the μ_i are the eigenvalues of $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$ and $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{M}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$ respectively. But this is merely a consequence of Lemma 1 of [24].

This concludes the proof of Theorem 0.21. □

We now turn to the proper evaluation of the Stieltjes transform power inference method, for $K = 3$ sources, $P_1 = 1/16$, $P_2 = 1/4$, $N = 24$ sensors, $M = 128$ samples and $n_1 = n_2 = n_3 = 4$. In Figure 7, we compare the distribution functions of the three estimated powers under the classical and Stieltjes transform method. We observe a significant gain in terms of bias reduction for the Stieltjes transform method here.

0.6 Conclusion

Random matrix theory for signal processing is a fast growing field of research whose interest is mainly motivated by the increase of the dimensionality and complexity of today's systems. While the first years of random matrix theory were mainly focusing Gaussian or invariant matrix distributions, the last ten years of research were mainly targeting large dimensional matrices with independent entries. This provided interesting results in particular on the limiting spectrum of sample covariance matrices, but not only, which led to new results on inverse problems for large dimensional systems. These results are often surprisingly simple and efficient as they perform well against exact maximum likelihood solutions, even for systems of not too large dimensions. Much more is however needed from a mathematical viewpoint relative in particular to second order statistics, see e.g. [37], [38], in order to evaluate theoretically the performance of these methods as well as a generalization to more intricate random matrix structures, such as Vandermonde matrices for array processing, see e.g. [39], or unitary random matrices, see e.g. [40]. A more exhaustive account of random matrix methods as well as more details on the methods presented here can be found in [15, 11, 29].

0.7 Exercises

- **Exercise 1.** Based on Theorem 0.9, prove the Marčenko-Pastur law, Theorem 0.5.

  Hint: Observe that the fixed-point equation in m_{F^B} reduces now to a second order polynomial from which $m_{F^B}(z)$ takes an explicit form. The inverse Stieltjes transform formula 4 gives the expression of F^B.

- **Exercise 2.** Let $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ be a random matrix with i.i.d. Gaussian entries of zero mean and variance $1/n$. For $\mathbf{R}_N \in \mathbb{C}^{N\times N}$ and $\mathbf{T}_N \in \mathbb{C}^{n\times n}$ deterministic and of uniformly bounded spectral norm such that $F^{\mathbf{R}_N} \Rightarrow F^R$ and $F^{\mathbf{T}_N} \Rightarrow F^T$, as $N, n \to \infty$, determine an expression of the Stieltjes transform of the limiting eigenvalue distribution of $\mathbf{B}_N = \mathbf{R}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{T}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{R}_N^{\frac{1}{2}}$ as $N/n \to c$.

  Hint: Follow the proof of Theorem 0.9 by looking for a deterministic equivalent of $\frac{1}{N}\mathrm{tr}\mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$ for some deterministic $\mathbf{A}$, taken to be successively $\mathbf{R}_N$ and $\mathbf{I}_N$. A good choice of the matrix $\mathbf{D}_N$ is $\mathbf{D}_N = a_N\mathbf{R}_N$.

- **Exercise 3.** Based on the definition of the Shannon-transform and on the G-estimator for the Stieltjes transform, determine a G-estimator for

$$\mathcal{V}_{\mathbf{T}_N}(x) = \frac{1}{N}\log\det\left(x\mathbf{T}_N + \mathbf{I}_N\right)$$

  based on the observations

$$\mathbf{y}_k = \mathbf{T}_N^{\frac{1}{2}}\mathbf{x}_k$$

  with $\mathbf{x}_k \in \mathbb{C}^N$ with i.i.d. entries of zero mean and variance $1/N$, independent across k, for $k \in \{1, \ldots, n\}$.

  Hint: Write the expression of $\mathcal{V}_{\mathbf{T}_N}(x)$ as a function of the Stieltjes transform of $\mathbf{T}_N$ and operate a variable change in the resulting integral using Theorem 0.9.

- **Exercise 4.** From the result of Theorem 0.11, propose an hypothesis test for the presence of a signal transmitted by a signal source and observed by a large array of sensors, assuming that the additive noise variance is either perfectly known or not.

  Hint: Observe that the ratio of the extreme eigenvalues in both $\mathcal{H}_0$ and $\mathcal{H}_1$ hypotheses is asymptotically independent of the noise variance.

- **Exercise 5.** For $\mathbf{W} \in \mathbb{C}^{N\times n}$, $n < N$, the n columns of a random unitarily invariant unitary matrix, $\mathbf{w}$ a column vector of $\mathbf{W}$, prove that, if $\mathbf{B}_N$ is a random matrix with bounded spectral norm, function of all columns of $\mathbf{W}$ but $\mathbf{w}$, then, as $N, n \to \infty$ with $n/N \to c < 1$,

$$\mathbf{w}^{\mathsf{H}}\mathbf{B}_N\mathbf{w} - \frac{1}{N-n}\mathrm{tr}(\mathbf{I}_N - \mathbf{W}\mathbf{W}^{\mathsf{H}})\mathbf{B}_N \xrightarrow{\text{a.s.}} 0.$$

  Hint: write $\mathbf{w}$ as the normalized projection of a Gaussian vector $\mathbf{x}$ on the subspace orthogonal to the space spanned by the columns of $\mathbf{W}$ but $\mathbf{w}$, i.e. $\mathbf{w} = \mathbf{\Pi}\mathbf{x}$, with $\mathbf{\Pi} = \mathbf{I}_N - \mathbf{W}\mathbf{W}^{\mathsf{H}} + \mathbf{w}\mathbf{w}^{\mathsf{H}}$.

Bibliography

[1] J. Wishart, "The generalized product moment distribution in samples from a normal multivariate population," *Biometrika*, vol. 20, no. 1-2, pp. 32–52, Dec. 1928.

[2] R. Couillet and M. Debbah, "A Bayesian framework for collaborative multi-source signal detection," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5186–5195, Oct. 2010.

[3] P. Bianchi, J. Najim, M. Maida, and M. Debbah, "Performance of Some Eigen-based Hypothesis Tests for Collaborative Sensing," *IEEE Trans. Inf. Theory*, 2010, to appear.

[4] R. A. Fisher, "The sampling distribution of some statistics obtained from non-linear equations," *The Annals of Eugenics*, vol. 9, pp. 238–249, 1939.

[5] M. A. Girshick, "On the sampling theory of roots of determinantal equations," *The Annals of Math. Statistics*, vol. 10, pp. 203–204, 1939.

[6] P. L. Hsu, "On the distribution of roots of certain determinantal equations," *The Annals of Eugenics*, vol. 9, pp. 250–258, 1939.

[7] S. Roy, "p-statistics or some generalizations in the analysis of variance appropriate to multivariate problems," *Sankhya: The Indian Journal of Statistics*, vol. 4, pp. 381–396, 1939.

[8] Harish-Chandra, "Differential operators on a semi-simple Lie algebra," *American Journal of Mathematics*, vol. 79, pp. 87–120, 1957.

[9] C. Itzykson and J. B. Zuber, *Quantum field theory*. Dover Publications, 2006.

[10] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Feb. 1999.

[11] Z. Bai and J. W. Silverstein, "Spectral Analysis of Large Dimensional Random Matrices," *Springer Series in Statistics*, 2009.

[12] P. Billingsley, *Convergence of probability measures*. Hoboken, NJ: John Wiley & Sons, Inc., 1968.

[13] V. A. Marčenko and L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices," *Math USSR-Sbornik*, vol. 1, no. 4, pp. 457–483, Apr. 1967.

[14] P. Billingsley, *Probability and Measure*, 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc., 1995.

[15] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, 2004.

[16] Z. D. Bai and J. W. Silverstein, "No Eigenvalues Outside the Support of the Limiting Spectral Distribution of Large Dimensional Sample Covariance Matrices," *Annals of Probability*, vol. 26, no. 1, pp. 316–345, Jan. 1998.

[17] D. N. C. Tse and O. Zeitouni, "Linear multiuser receivers in random environments," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 171–188, 2000.

[18] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, 1995.

[19] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large System Analysis of Linear Precoding in MISO Broadcast Channels with Limited Feedback," *IEEE Trans. Inf. Theory*, 2010, submitted for publication. [Online]. Available: http://arxiv.org/abs/0906.3682

[20] J. Baik and J. W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," *Journal of Multivariate Analysis*, vol. 97, no. 6, pp. 1382–1408, 2006.

[21] Z. D. Bai and J. W. Silverstein, "Exact Separation of Eigenvalues of Large Dimensional Sample Covariance Matrices," *The Annals of Probability*, vol. 27, no. 3, pp. 1536–1555, 1999.

[22] J. W. Silverstein and S. Choi, "Analysis of the limiting spectral distribution of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 295–309, 1995.

[23] X. Mestre, "On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5353–5368, Nov. 2008.

[24] R. Couillet, J. W. Silverstein, and M. Debbah, "Eigen-Inference for Energy Estimation of Multiple Sources," *IEEE Trans. Inf. Theory*, 2011, to appear. [Online]. Available: http://arxiv.org/abs/1001.3934

[25] X. Mestre, "Improved estimation of eigenvalues of covariance matrices and their associated subspaces using their sample estimates," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5113–5129, Nov. 2008.

[26] F. Hiai and D. Petz, *The semicircle law, free random variables and entropy - Mathematical Surveys and Monographs No. 77.* Providence, RI, USA: American Mathematical Society, 2006.

[27] O. . Ryan and M. Debbah, "Free deconvolution for signal processing applications," in *Proc. IEEE International Symposium on Information Theory (ISIT'07)*, Nice, France, Jun. 2007, pp. 1846–1850.

[28] N. R. Rao and A. Edelman, "The polynomial method for random matrices," *Foundations of Computational Mathematics*, vol. 8, no. 6, pp. 649–702, Dec. 2008.

[29] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*, 1st ed. New York, NY, USA: Cambridge University Press, 2011, to appear.

[30] P. Vallet, P. Loubaton, and X. Mestre, "Improved subspace DoA estimation methods with large arrays: The deterministic signals case," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'09)*, 2009, pp. 2137–2140.

[31] W. Rudin, *Real and complex analysis*, 3rd ed. McGraw-Hill Series in Higher Mathematics, May 1986.

[32] D. Gregoratti and X. Mestre, "Random DS/CDMA for the amplify and forward relay channel," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 1017–1027, 2009.

[33] R. Couillet and M. Debbah, "Free deconvolution for OFDM multicell SNR detection," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'08)*, Cannes, France, 2008.

[34] I. S. Gradshteyn and I. M. Ryzhik, "Table of Integrals, Series and Products," *Academic Press, 6th edition*, 2000.

[35] M. Guillaud, M. Debbah, and A. L. Moustakas, "Maximum Entropy MIMO Wireless Channel Models," *IEEE Trans. Inf. Theory*, 2006, submitted for publication. [Online]. Available: http://arxiv.org/abs/cs.IT/0612101

[36] S. H. Simon, A. L. Moustakas, and L. Marinelli, "Capacity and character expansions: Moment generating function and other exact results for MIMO correlated channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5336–5351, 2006.

[37] Z. D. Bai and J. W. Silverstein, "CLT of linear spectral statistics of large dimensional sample covariance matrices," *Annals of Probability*, vol. 32, no. 1A, pp. 553–605, 2004.

[38] J. Yao, R. Couillet, J. Najim, E. Mouline, and M. Debbah, "CLT for eigen-inference methods in cognitive radios," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'11)*, Prague, Czech Republic, 2011, submitted for publication.

[39] Ø. Ryan and M. Debbah, "Asymptotic Behavior of Random Vandermonde Matrices With Entries on the Unit Circle," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3115–3148, Jul. 2009.

[40] R. Couillet, J. Hoydis, and M. Debbah, "Deterministic equivalents for the analysis of unitary precoded systems," *IEEE Trans. Inf. Theory*, 2011, submitted for publication.