Projet ANR SESAME

Intermediate report, $t_0 + 24$ (WP2) : Eigen-inference methods in large systems

January 2010

Contents

Abstract

The problem we are concerned with enters the general scope of *inverse problems*; that is, from the observation of a random process, we want to infer general characteristics of that process. In the present telecommunication setting, the problem consists in inferring signal characteristics from noisy observations. Specifically, we study here the case where a set of K wireless devices simultaneously transmit signals in the open air; to be very general, device k is assumed to be composed of a number n_k of antennas and transmits with power P_k. The sum signal is captured by a set of N sensors, which aim at reliably inferring (i) the total number K of transmit sources, (ii) the respective powers $P_1, \ldots, P_K$ of the K signal sources.

The objective of this document is to provide recent results from the field of random matrix theory which enable (i) the derivation of solutions to such inverse problems when N, $n_1, \ldots, n_K$ are large, (ii) the derivation of computationally cheap practical algorithms. These results rely on inference methods of the (hidden) eigenvalue distribution of signal source matrices from the eigenvalue distribution of the observed received signal matrix. These methods, mathematically easy to derive when the matrix dimensions grow large, are referred to as *eigen-inference* methods.

In order to perform eigen-inference, we will successively consider moment-based approaches in Chapter 1 as well as analytic Stieltjes transform-based approaches in Chapter 2.

Chapter 1

Moment-based methods

Consider a set of n independent and identically distributed N-dimensional complex vectors $\mathbf{y}_1, \ldots, \mathbf{y}_n \in \mathbb{C}^N$, such that $\mathrm{E}[\mathbf{y}_1] = 0$ and $\mathrm{E}[\mathbf{y}_1 \mathbf{y}_1^{\mathsf{H}}] = \mathbf{T}$. It is a classical problem to try to determine the eigenvalue distribution of $\mathbf{T}$ from the sole realization of $\mathbf{y}_1, \ldots, \mathbf{y}_n$. This is in particular the case if $\mathbf{T}$ is diagonal for instance. Stacking the vectors $\mathbf{y}_1, \ldots, \mathbf{y}_n$ in the $N \times n$ matrix $\mathbf{Y}$, we can model the observation as

$$\mathbf{Y} = \mathbf{T}^{\frac{1}{2}} \mathbf{X} \tag{1.1}$$

where $\mathbf{X} \in \mathbb{C}^{N \times n}$ has independent and identically distributed (i.i.d.) entries.

The eigen-inference problem here consists in retrieving the eigenvalue distribution of $\mathbf{T}$ from the matrix $\mathbf{Y}$. It turns out that an often suboptimal but extremely efficient way to infer the eigenvalues of $\mathbf{T}$ is to consider the eigenvalues of the *Gram* matrix $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ associated to $\mathbf{Y}$, instead of $\mathbf{Y}$ itself. In this chapter, we introduce methods approaches, which enable the identification of the eigenvalue distribution of $\mathbf{T}$ through its successive moments, as a function of the successive moments of the eigenvalue distribution of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$.

1.1 Free probability

1.1.1 Basics of free probability

In [1], Voiculescu generalizes the classical mathematical field of probability theory, in which random variables are classically reals or complex scalars, to general-purpose $*$-algebras of random variables. A case of particular interest to us is the *non-commutative* $*$-algebras of asymptotically large Hermitian random matrices.

Voiculescu defines its generalized probability space as a pair $(\mathcal{A}, \phi)$ of a given $*$-algebra $\mathcal{A}$ and an expectation operator ϕ such that $\phi(1) = 1$. In the case of random Hermitian matrices, an operator ϕ of practical interest is the normalized expected trace operator, i.e.

$$\phi(\mathbf{A}) \triangleq \lim_{N \to \infty} \frac{1}{N} \mathrm{E}[\mathrm{tr}(\mathbf{A}_N)] \tag{1.2}$$

where $\{\mathbf{A}_N \in \mathbb{C}^{N \times N}\}$ is a series of matrices converging in spectrum to the hypothetical infinite-size matrix $\mathbf{A}$.

In the same way as there exists an independence notion in scalar random variables, Voiculescu defines the so-called notion of *freeness*.

Definition 1 *A family of unital $*$-subalgebras $(\mathcal{A}_i)_{i\in I}$ is called a free family if, for all n,*

$$\left\{ \begin{array}{l} A_j \in \mathcal{A}_{i_j} \\ i_1 \neq i_2, \ldots, \ i_{n-1} \neq i_n \\ \phi(A_1) = \ldots = \phi(A_n) = 0 \end{array} \right\} \Rightarrow \phi(A_1 \ldots A_n) = 0$$

We will say that the variables $A_1, \ldots, A_n$ are free if the underlying algebras they generate form a free family.

Note that the condition for freeness is tighter than the condition for independence. As such, two (asymptotically large) Hermitian random matrices $\mathbf{A}$ and $\mathbf{B}$ with independent entries are not necessarily free. The additional condition for $\mathbf{A}$ and $\mathbf{B}$ to be free is that the eigenvectors of $\mathbf{A}$ and $\mathbf{B}$ are distributed in directions that are totally disconnected.

Here are some classical examples of families of free random matrices

- $(\mathbf{A}, \mathbf{U}\mathbf{B}\mathbf{U}^{\mathsf{H}})$ for independent $\mathbf{A}$, $\mathbf{B}$, $\mathbf{U}$ with $\mathbf{U}$ a unitary (Haar distributed) matrix
- the family of Haar distributed matrices
- the family of random Gaussian matrices
- a Haar or Gaussian matrix and a deterministic matrix

Most of the matrices that do not enter this list are not free. If so, in the same way that little can be said of non-independent random variables in classical probability theory, very little can be said about random matrices that are not free.

Because of the non-commutativity of free random variables, the expectation relations of products of free random variables are different from those derived in classical probability theory. In particular, for A, B free random variables, we have that

$$\begin{aligned} \phi(AB) &= \phi(A)\phi(B) \\ \phi(ABAB) &= \phi(A^2)\phi(B)^2 + \phi(A)^2\phi(B^2) - \phi(A)^2\phi(B)^2 \\ \phi(AB^2A) &= \phi(A^2)\phi(B^2) \end{aligned}$$

It appears in particular that the second relation here does not exist in classical probability theory. In the next section, we exhibit the relation between free probability and eigenvalue distribution.

1.1.2 Free probability and eigenvalue distributions

The free probability setting applied to random matrices is particularly suited to treat direct problems (for which we want to evaluate the eigenvalue distribution of the output of a random process, given its input) as well as inverse problems (for which we want to perform just the opposite) involving large dimensional random matrices.

The key here is to observe that, for a given Hermitian matrix $\mathbf{A}$, $\phi(\mathbf{A}^k)$, referred to as the K^{th} free order moment of $\mathbf{A}$, reads

$$\phi(\mathbf{A}^k) = \lim_{N\to\infty} \frac{1}{N}\mathrm{E}[\mathrm{tr}(\mathbf{A}_N^k)] \tag{1.3}$$

$$= \lim_{N\to\infty} \int a^k dF^{\mathbf{A}_N}(a) \tag{1.4}$$

with $F^{\mathbf{A}_N}$ the distribution function of the eigenvalues of $\mathbf{A}_N$.

Taking $\mathbf{A}$, $\mathbf{B}$ two free random variables, it is then possible to evaluate, for instance,

$$\phi(\mathbf{AB}) = \phi(\mathbf{A})\phi(\mathbf{B}) \tag{1.5}$$

$$= \left[\lim_{N\to\infty} \int a dF^{\mathbf{A}_N}(a)\right] \times \left[\lim_{N\to\infty} \int b dF^{\mathbf{B}_N}(b)\right] \tag{1.6}$$

$$= \left[\int a dF^{\mathbf{A}}(a)\right] \times \left[\int b dF^{\mathbf{B}}(b)\right] \tag{1.7}$$

where $F^{\mathbf{A}}$ and $F^{\mathbf{B}}$ are the limiting eigenvalue distributions of $\mathbf{A}_N$ and $\mathbf{B}_N$. We can then derive the first moment of the product of $\mathbf{A}$ and $\mathbf{B}$; but we can also write

$$\phi(\mathbf{A}) = \frac{\phi(\mathbf{AB})}{\phi(\mathbf{B})} \tag{1.8}$$

$$= \left[\int c dF^{\mathbf{AB}}(c)\right] \times \left[\int b dF^{\mathbf{B}}(b)\right]^{-1} \tag{1.9}$$

which gives the first moment of $\mathbf{A}$ in the inverse problem of recovering $\mathbf{A}$ from $\mathbf{AB}$.

We will see in what follows that, in a similar fashion, all free moments of products of $\mathbf{A}$ and $\mathbf{B}$ matrices can be derived (direct problems), as well as all free moments of $\mathbf{A}$ given $\mathbf{AB}$ and $\mathbf{B}$ (inverse problems).

For practical applications though, matrices are never of infinite size. Still, it turns out that, for matrix sizes of practical usage (matrices larger than 8×8 are often considered large enough), the above relations can be blindly applied and are good approximations of the true moments of the finite-size random matrices. However, the relations (1.3) are increasingly complex for higher moment orders. Fortunately, combinatorics approaches exist that allow for a systematic derivation of the successive moments.

1.2 Combinatorics methods

In classical probability theory, if A and B are independent, the moments of $A+B$ are functions of the moments of A and those of B. In particular, for A, B independent,

$$c_k(A+B) = c_k(A) + c_k(B) \tag{1.10}$$

with $c_k(X)$ the cumulants of X (polynomial functions of the moments m_k of X). The cumulants c_n are connected to the moments m_n through formulas invoking partitions,

$$m_n = \sum_{\pi\in\mathcal{P}(n)} \prod_{V\in\pi} c_{|V|} \tag{1.11}$$

In free probability theory, if $\mathbf{A}$, $\mathbf{B}$ are Hermitian matrices, there exists a relationship between the eigenvalue distribution moments $M_k(\mathbf{A}+\mathbf{B}) = \phi((\mathbf{A}+\mathbf{B})^k)$ of the sum of $\mathbf{A}$ and $\mathbf{B}$ and the eigenvalue distribution moments of $\mathbf{A}$ and $\mathbf{B}$. This involves the definition of free cumulants C_k, interpreted by Speicher [2] in terms of non-crossing partitions, as

$$M_n = \sum_{\pi \in \mathcal{NC}(n)} \prod_{V \in \pi} C_{|V|} \tag{1.12}$$

with $NC(n)$ the set of non-crossing partitions of $\{1, \ldots, n\}$. One element of the set of $NC(8)$ is depicted in Figure 1.1. With $M_k = M_k(\mathbf{A})$, $C_k = C_k(\mathbf{A})$, we then have relations of the type,

$$\begin{aligned} C_1 &= M_1 \\ C_2 &= M_2 - M_1^2 \\ C_3 &= M_3 - 3M_1M_2 + 2M_1^2 \end{aligned}$$

As in classical probability theory, those cumulants verify

$$C_k(\mathbf{A}+\mathbf{B}) = C_k(\mathbf{A}) + C_k(\mathbf{B}) \tag{1.13}$$

Equations (1.12) and (1.13) together allow to evaluate all moments of sums of free random matrices. The same type of derivation operates for products of such free matrices. For these, we have the cumulant-to-moment formula

$$M_n(\mathbf{AB}) = \sum_{(\pi_1,\pi_2) \in NC(n)} \prod_{\substack{V_1 \in \pi_1 \\ V_2 \in \pi_2}} C_{|V_1|}(\mathbf{A}) C_{|V_2|}(\mathbf{B}) \tag{1.14}$$

As a consequence, thanks to (1.12), (1.13) and (1.14), all multinomial forms of matrices can be treated. The eigenvalue distribution of such matrices can then be recovered. Now observe also that the moment relations are mere polynomials and that the order k free moment of (the eigenvalue distribution of) a sum/product of matrices is given as a function of all order free moments up to k of the (eigenvalue distribution of the) underlying matrices. The polynomial relations can then be commuted to obtain the free moments of the underlying matrices as a function of the free moments of their sum/product. This means that the inverse problems of interest can be easily treated through moments computation; the treatment is made even easier and computationally cheap by the fact that only polynomials have to be computed.

As an example, consider the product $\mathbf{DXX}^{\mathsf{H}}$ of a deterministic matrix $\mathbf{D} \in \mathbb{C}^{n \times n}$ and the Wishart matrix $\mathbf{XX}^{\mathsf{H}}$ where $\mathbf{X} \in \mathbb{C}^{n \times N}$ has i.i.d. complex Gaussian entries with zero mean and variance $1/N$. Denoting $D_k = \phi(\mathbf{D}^k)$ and $M_k = \phi((\mathbf{DXX}^{\mathsf{H}})^k)$, the combinatorial calculus can be automatically processed in software and, with $c = \lim n/N$, the output is

$$M_1 = D_1 \tag{1.15}$$

$$M_2 = D_2 + cD_1^2 \tag{1.16}$$

$$M_3 = D_3 + 3cD_1D_2 + c^2D_1^3 \tag{1.17}$$

$$M_4 = D_4 + 2cD_2^2 + 4cD_3D_1 + 6c^2D_2D_1 + c^3D_1^4 \tag{1.18}$$

$$\vdots = \vdots \tag{1.19}$$

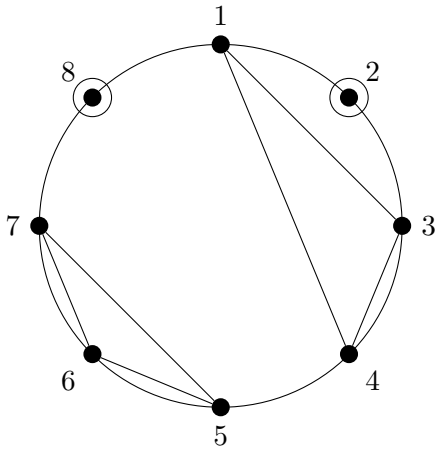

Figure 1.1: Non-crossing partition $\pi = \{\{1,3,4\},\{2\},\{5,6,7\},\{8\}\}$ of $NC(8)$.

These relations can be iteratively inverted so to obtain the solution of the inverse problem that allows one to recover the free moments of $\mathbf{D}$ from those of $\mathbf{DXX}^{\mathsf{H}}$. The procedure that consists in obtaining the moments of $\mathbf{D}$ from those of $\mathbf{DXX}^{\mathsf{H}}$ is referred to as free (product) deconvolution.

However, these methods do not allow to treat all problems of practical interest in wireless communications. In particular, the widely spread information plus noise model cannot be processed here. As a consequence, we need extra tools to process the case

$$\mathbf{YY}^{\mathsf{H}} = (\mathbf{R}+\mathbf{X})(\mathbf{R}+\mathbf{X})^{\mathsf{H}} \tag{1.20}$$

where $\mathbf{R} \in \mathbb{C}^{n\times N}$ is deterministic and $\mathbf{X} \in \mathbb{C}^{n\times N}$ is Gaussian. This is the subject of the subsequent section.

1.3 Free deconvolution of the information plus noise model

In [3], the information plus noise model is processed in the free probability scheme described above. Specifically, we have the following result

Theorem 1 *Let $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ be a random matrix with i.i.d. Gaussian entries of zero mean and unit variance, and $\mathbf{R}_N$ be a (non-necessarily random) matrix such that the eigenvalue distribution of $\mathbf{\Gamma}_N = \frac{1}{n}\mathbf{R}_N\mathbf{R}_N^{\mathsf{H}}$ converges weakly and almost surely to the compactly supported probability distribution μ_Γ, as $n, N \to \infty$ with limit ratio $N/n \to c > 0$. Then the eigenvalue distribution of*

$$\mathbf{B}_N = \frac{1}{n}\left(\mathbf{R}_N + \sigma\mathbf{X}_N\right)\left(\mathbf{R}_N + \sigma\mathbf{X}_N\right)^{\mathsf{H}}$$

converges weakly and almost surely to the compact supported measure μ_B such that

$$\mu_B = ((\mu_\Gamma \boxslash \mu_c) \boxplus \delta_{\sigma^2}) \boxtimes \mu_c \tag{1.21}$$

with μ_c the probability distribution with distribution function the Marčenko-Pastur law, and δ_{σ^2} the probability distribution of a single mass in σ^2. Equation (1.21) is the free convolution of the

information plus noise model. This can be reverted as

$$\mu_\Gamma = ((\mu_B \boxtimes \mu_c) \boxminus \delta_{\sigma^2}) \boxslash \mu_c$$

which is the free deconvolution of the information plus noise model.

In the theorem above, the operators '$\boxplus$', '$\boxminus$', '$\boxtimes$' and '$\boxslash$' stand respectively for the additive free convolution and deconvolution and the product free convolution and deconvolution operations; for instance, in terms of moments, when the eigenvalue distribution of $\mathbf{A}$ and $\mathbf{B}$ is compactly supported, $\mu_\mathbf{A} \boxplus \mu_\mathbf{B}$ is equivalent to $C_k(\mathbf{A}+\mathbf{B}) = C_k(\mathbf{A})+C_k(\mathbf{B})$ for all k. Therefore (1) can be translated in terms of moments and allows one to retrieve all moments of $\mathbf{R}\mathbf{R}^\mathsf{H}$ as a function of the moments of $\mathbf{B}_N$.

Remark 1 *Incidentally, the operations '$\cdot \boxtimes \mu_c$' and '$\cdot \boxslash \mu_c$', with the Marčenko-Pastur law (i.e. the asymptotic eigenvalue distribution of Wishart matrix $\mathbf{X}\mathbf{X}^\mathsf{H}$), correspond respectively, in moment terms, as a cumulant-to-moment and a moment-to-cumulant transform, which further simplifies the free deconvolution operation.*

1.4 Application to power estimation

Consider the uplink orthogonal CDMA scenario constituted of n uplink transmitters and a base station. Transmitter k has power P_k, spreading code $\mathbf{w}_k \in \mathbb{C}^N$ (N is the chip length) and sends symbol $s_k^{(l)}$ at time l. Denoting $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_n] \in \mathbb{C}^{N\times n}$ the orthogonal CDMA code matrix, $\sigma\mathbf{n}^{(l)} \in \mathbb{C}^N$ the AWGN noise at time l, the base station receives at time l the signal $\mathbf{y}^{(l)}$ given by

$$\mathbf{y}^{(l)} = \sum_{k=1}^{n} \mathbf{w}_k \sqrt{P_k} s_k^{(l)} + \sigma\mathbf{n}^{(l)} \tag{1.22}$$

$$= \mathbf{W}\mathbf{P}^{\frac{1}{2}}\mathbf{s}^{(l)} + \sigma\mathbf{n}^{(l)} \tag{1.23}$$

with $\mathbf{P} = \mathrm{diag}(P_1, \ldots, P_n)$ and $\mathbf{s}^{(l)} = (s_1^{(l)}, \ldots, s_n^{(l)})^\mathsf{T}$.

Concatenating the received data for $l = 1, \ldots, L$ into $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(L)}]$, we finally have the model

$$\mathbf{Y} = \mathbf{W}\mathbf{P}^{\frac{1}{2}}\mathbf{S} + \sigma\mathbf{N} \tag{1.24}$$

with $\mathbf{S} = [\mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(L)}]$ and $\mathbf{N} = [\mathbf{n}^{(1}, \ldots, \mathbf{n}^{(L)}]$.

In this model, $\mathbf{W} \in \mathbb{C}^{N\times n}$, $n \leq N$ is such that $\mathbf{W}^\mathsf{H}\mathbf{W} = \mathbf{I}_n$, $\mathbf{N} \in \mathbb{C}^{N\times L}$ has Gaussian entries with zero mean and variance 1, $\mathbf{S} \in \mathbb{C}^{n\times L}$ has Gaussian entries with zero mean and variance $1/L$. Then, as the system dimensions grow large, with constant ratio, we have, with $\mathbf{B}_N = \frac{1}{L}\mathbf{Y}\mathbf{Y}^\mathsf{H}$, from Theorem 1

$$\mu_{\mathbf{W}\mathbf{P}^{\frac{1}{2}}\mathbf{S}\mathbf{S}^\mathsf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{W}^\mathsf{H}} = ((\mu_{\mathbf{B}_N} \boxtimes \mu_c) \boxminus \delta_{\sigma^2}) \boxslash \mu_c \tag{1.25}$$

where $c = N/L$.

Now, since $\mathbf{W}^\mathsf{H}\mathbf{W} = \mathbf{I}_n$, we have straightforwardly that $\mu_{\mathbf{W}\mathbf{P}^{\frac{1}{2}}\mathbf{S}\mathbf{S}^\mathsf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{W}^\mathsf{H}} = \mu_{\mathbf{P}\mathbf{S}\mathbf{S}^\mathsf{H}}$.

Applying a further product deconvolution, we have

$$\mu_{\mathbf{P}} = \mu_{\mathbf{PSS}^{\mathsf{H}}} \boxslash \mu_{\mathbf{SS}^{\mathsf{H}}} \tag{1.26}$$

Hence finally

$$\mu_{\mathbf{P}} = ((\mu_{\mathbf{B}_N} \boxtimes \mu_c) \boxminus \delta_{\sigma^2}) \boxslash \mu_c \boxslash \mu_{c'} \tag{1.27}$$

with $c' = n/L$.

This equality can then be translated in terms of moments, evaluated in software, given the equations above. If now the n users are grouped in a finite number K of groups with same transmit power, each group composed of n/K users, then the diagonal matrix $\mathbf{P}$ is composed of exactly K distinct values. From the first K *estimated* moments $\hat{p}_k$ of $\mu_{\mathbf{P}}$, it is then possible to solve the K equation

$$\frac{1}{K}\left(\hat{P}_1 + \hat{P}_2 + \ldots + \hat{P}_K\right) = \hat{p}_1 \tag{1.28}$$

$$\frac{1}{K}\left(\hat{P}_1^2 + \hat{P}_2^2 + \ldots + \hat{P}_K^2\right) = \hat{p}_2 \tag{1.29}$$

$$\ldots = \ldots \tag{1.30}$$

$$\frac{1}{K}\left(\hat{P}_1^K + \hat{P}_2^K + \ldots + \hat{P}_K^K\right) = \hat{p}_K \tag{1.31}$$

the unique solution $\hat{P}_1, \ldots, \hat{P}_K$ of which is obtained from the Newton-Girard formulas [5]. This allows one to get an estimate $\hat{P}_1, \ldots, \hat{P}_K$ of the K distinct transmit powers.

Note however that we need to use exactly K estimated moments here to obtain an estimate of $P_1, \ldots, P_K$. Relying on classical maximum likelihood estimators or minimum mean square error estimators, it is possible, but at a highly more expensive computational price, to obtain an estimate of $P_1, \ldots, P_K$ based on a larger range of moments. For instance, $\hat{P}_1, \ldots, \hat{P}_K$ can be given by

$$(\hat{P}_1, \ldots, \hat{P}_K) = \arg \min_{(\tilde{P}_1, \ldots, \tilde{P}_K)} (\tilde{b}_1 - b_1, \ldots, \tilde{b}_M - b_M)^{\mathsf{T}} (\tilde{b}_1 - b_1, \ldots, \tilde{b}_M - b_M) \tag{1.32}$$

with $\tilde{b}_k$ the estimated moment of $\mathbf{B}_N$ when $\mathbf{P} = \mathrm{diag}(\tilde{P}_1, \ldots, \tilde{P}_K)$.

1.5 Conclusion

Thanks to the free probability theory, it is possible to obtain computationally cheap expressions that link the free moments of sums and products of free random matrices. The extension of sums and products to the information plus noise model is shown to perform in a similar fashion, based again on free convolution operations on the moments of the underlying matrix eigenvalue distributions. This generates a general framework for determining the eigenvalue distribution of a large range of random matrices expressed as polynomial functions of Gaussian Gaussian and deterministic matrices.

However, the moment approach has some limitations. First, the theorems derived here are used in practice in the finite N dimension, for which equalities are turned into approximations. Typically, higher order moments are very badly approximated compared to smaller moments,

but there is no trivial way to error distribution (which is by the way linked to the parameters we wish to estimate); for instance, in determining the moments of $\mathbf{P}$ from $\mathbf{B}_N$ in the section above, be there a central limit theorem for the estimated moments, (1.33) above must be replaced by

$$(\hat{P}_1, \dots, \hat{P}_K) = \arg \min_{(\tilde{P}_1, \dots, \tilde{P}_K)} (\tilde{b}_1 - b_1, \dots, \tilde{b}_M - b_M)^{\mathsf{T}} \mathbf{C}^{-1} (\tilde{b}_1 - b_1, \dots, \tilde{b}_M - b_M) \tag{1.33}$$

where $\mathbf{C}$ is the covariance matrix of the estimated moments. This matrix is however dependent on $\tilde{P}_1, \dots, \tilde{P}_K$, so it varies constantly along the search for the optimal $\hat{P}_1, \dots, \hat{P}_K$. This however increases immensely the complexity of the estimator.

Also, note importantly that the method derived above is only valid for a very limited class of matrices, i.e. Gaussian matrices, unitary matrices and deterministic matrices. Any other matrix with additional structure requires a whole new setup. Finally, moment methods only work when the moments actually determine the eigenvalue distribution. If, for instance, the support of the eigenvalue distribution under study is not compact, then it is very likely that Carleman's condition is not met (Appendix B in [6]) and then the moments do not determine uniquely the underlying distribution. In this case, even if the above methods are extended to more structured matrices, they remain useless. In the next section, we develop an alternative approach through the Stieltjes transform of the eigenvalue distribution.

Chapter 2

Analytic approaches

In this section, we discuss a very different approach to the inverse problem involving large random matrices. This approach will be shown to be more general than the moment-based approach in the sense that (i) it covers a larger range of random matrices, including general purpose matrices with i.i.d. entries (i.e. not necessarily Gaussian), (ii) it does really only on independence of random matrices, therefore releasing freeness assumptions and (iii) it does not require strong assumptions on the moments, such as the Carleman condition discussed in Chapter 1.

However, we shall see that the mathematics involved in this quite novel theory are more involved, to the point that model generalizations are not straightforward. Also, we shall see that a strong condition is requirement on the parameters to be estimated so that the analytic approach is feasible; this condition will be referred to as the cluster separability condition.

2.1 Sample covariance matrices with few distinct eigenvalues

We consider here the sample covariance matrix model, i.e. we study the data matrix $\mathbf{Y} \in \mathbb{C}^{n \times N}$ that is the concatenation of independent columns $\mathbf{y}_1, \ldots, \mathbf{y}_N \in \mathbb{C}^n$ with covariance matrix $\mathbf{T} = \mathrm{E}[\mathbf{y}_1 \mathbf{y}_1^{\mathsf{H}}] \in \mathbb{C}^{n \times n}$. We assume that $\mathbf{T}_n$ has a finite number of distinct eigenvalues, which we wish to recover. Prior to the proper estimation of the eigenvalues of $\mathbf{T}_n$, we need to study the asymptotic spectrum of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$. For simplicity in the following derivations, we shall consider the matrix $\mathbf{B}_n = \frac{1}{n}\mathbf{Y}^{\mathsf{H}}\mathbf{Y}$ instead of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$, which has the same eigenvalue distribution but for a mass in zero.

Since the $\mathbf{y}_i$'s are drawn independently, $\mathbf{Y}$ can be modelled as

$$\mathbf{Y} = \mathbf{T}_n^{\frac{1}{2}} \mathbf{X}_n^{\mathsf{H}} \tag{2.1}$$

where $\mathbf{X}_n^{\mathsf{H}} \in \mathbb{C}^{n \times N}$ has independent columns with identically distributed entries.

We first introduce the Stieltjes transform of a distribution F as follows,

Definition 2 *Let F be a distribution function. For $z \in \mathbb{C}^+$, the Stieltjes transform $m(z)$ of F is defined as*

$$m(z) = \int \frac{1}{t - z} dF(t) \tag{2.2}$$

For all $a < b \in \mathbb{R}$, we then have the inverse Stieltjes transform

$$F([a,b]) = \frac{1}{\pi} \lim_{y \to 0^+} \int_a^b \Im[m(x+iy)] \tag{2.3}$$

Therefore, through the inversion formula (2.3), the Stieltjes transform m_F uniquely determines the distribution F; in other words, it is equivalent to study the probability distribution F of some random process or to study its Stieltjes transform m_F.

Back to our previous model, from [6] and [7], we now have the following result

Theorem 2 *Let $\mathbf{X}_n \in \mathbb{C}^{N \times n}$ have complex independent entries with zero mean and unit variance and $\mathbf{T}_n \in \mathbb{R}^{N \times N}$ be Hermitian with eigenvalue distribution converging almost surely to H, as $N \to \infty$. Let $\mathbf{B}_n = \frac{1}{n}\mathbf{X}_n\mathbf{T}_n\mathbf{X}_n^{\mathsf{H}}$. Then, under some additional mild assumptions on $\mathbf{X}_n$, as $n, N \to \infty$, $N/n \to c_0 > 0$, almost surely, the eigenvalue distribution of $\mathbf{B}_n$ converges weakly to F with Stieltjes transform $m_F(z)$, $z \in \mathbb{C}^+$, being the unique solution with positive imaginary part of*

$$z = -\frac{1}{m_F(z)} + \frac{1}{c_0} \int \frac{t}{1 + t m_F(z)} dH(t) \tag{2.4}$$

Theorem 2 therefore gives access to the almost sure limiting distribution F of the eigenvalues of $\mathbf{B}_n$ through an expression of the Stieltjes transform m_F. In the particular case where the population covariance matrix $\mathbf{T}_n$ has a finite number K of distinct eigenvalues, we then have that, for $z \in \mathbb{C}^+$, $m_F(z)$ is the unique solution with positive imaginary part of

$$z = -\frac{1}{m_F(z)} + \sum_{k=1}^{K} \frac{1}{c_k} \frac{T_k}{1 + T_k m_F(z)} \tag{2.5}$$

with $T_1, \ldots, T_K$ the distinct eigenvalues of $\mathbf{T}_n$ and $c_k = \lim N/n_k$, n_k the multiplicity of T_k.

The asymptotic spectrum of F in the latter case has a peculiar behaviour, namely, for c_k large enough, the spectrum of F is divided into up to K clusters. This is depicted in Figures 2.1 and 2.2 in the case when $T_1 = 1$, $T_2 = 3$ and T_3 is alternatively 10 and 5, when $c_0 = \lim N/n = 10$. Note that, when subsequent T_i's are too close, subsequent clusters merge into a single cluster. The inference scheme proposed hereafter requires asymptotic separation of these clusters. We therefore need to identify a rule that implies this feature. In [8], we derive necessary and sufficient conditions for asymptotic spectrum separability; the conditions are merely slight modifications from those proposed in [9] and come as follow

Proposition 1 *The cluster corresponding to T_k is associated to T_k and T_k only if and only if the following condition is fulfilled*

$$\max \left\{ \sum_{i=r}^{K} \frac{1}{c_r} \frac{(T_r m_k)^2}{(1 + T_r m_k)^2}, \sum_{r=1}^{K} \frac{1}{c_r} \frac{(T_r m_{k+1})^2}{(1 + T_r m_{k+1})^2} \right\} < 1 \tag{2.6}$$

where $m_1 < \ldots < m_K$ are the K reals solutions to the equation in m,

$$\sum_{r=1}^{K} \frac{1}{c_r} \frac{(T_r m)^3}{(1 + T_r m)^3} = 1 \tag{2.7}$$

and we take the convention $m_{K+1} = \infty$.

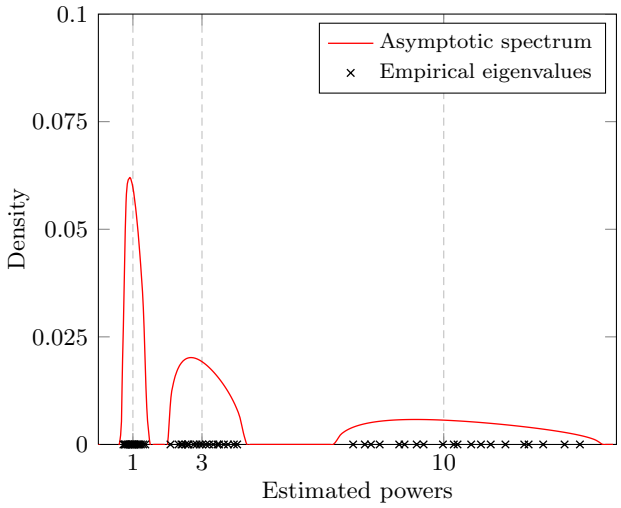

Figure 2.1: Empirical and asymptotic eigenvalue distribution of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ when $\mathbf{P}$ has three distinct entries $T_1 = 1$, $P_2 = 3$, $P_3 = 10$, $n_1 = n_2 = n_3$, $c_0 = 10$. Empirical test: $n = 60$.

If the condition is fulfilled, we can then give a consistent estimate of T_k, i.e. we can find $\hat{T}_k$, such that

$$\hat{T}_k - T_k \xrightarrow{\text{a.s.}} 0 \tag{2.8}$$

as $N, n_1, \ldots, n_K \to \infty$ and $0 < \lim N/n_k = c_k < \infty$. Before performing this estimation, let us explore further the condition given by Proposition 1. It is clear that a scaling of all T_i's by a constant value does not impact the equations and therefore the condition (2.6) only depends on the ratios $T_2/T_1, \ldots, T_K/T_1$ and on the values $c_1, \ldots, c_K$. Assume for instance that $c_1 = \ldots = c_K$, and let us observe the cases $K = 2$ and $K = 3$. For $K = 2$, i.e. the matrix $\mathbf{T}$ is formed only of 2 distrinct eigenvalues, then both eigenvalues can be inferred under the condition that c_0 is above the plot in Figure 2.3. In the case $K = 3$, then, for $c_0 = 10$, we have the regions of decidability for the 3 eigenvalues given by Figure 2.4. Note in particular that the case $T_1 = 1, T_2 = 3, T_3 = 10$ is clearly inside the detectability region, while the triplet $T_1 = 1, T_2 = 3, T_3 = 5$ is not, which confirms the tendency observed in Figure 2.1 and Figure 2.2.

In the following, we assume that the condition of cluster separability given by proposition 1 is always fulfilled for the eigenvalue T_k. We shall then describe the method to give a consistent estimate of T_k, as $N, n_1, \ldots, n_K$ grow large.

2.2 Consistent estimation of the population eigenvalues

The main reason for the necessity of the separability condition originates from complex analysis. If $\mathcal{C}_k$ is some contour in the complex plane such that $\mathcal{C}_k$ revolves around T_k and T_k only, i.e. T_i, $i \neq k$, is not in the surface described by the contour, then we have the following classical result from complex analysis [10],

$$T_k = c_k \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \frac{1}{c_k} \frac{\omega}{T_k - \omega} d\omega \tag{2.9}$$

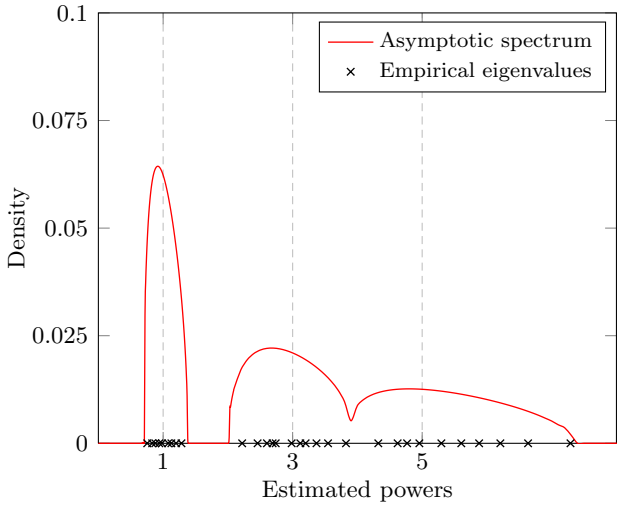

Figure 2.2: Empirical and asymptotic eigenvalue distribution of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ when $\mathbf{P}$ has three distinct entries $T_1 = 1$, $P_2 = 3$, $P_3 = 5$, $n_1 = n_2 = n_3$, $c_0 = 10$. Empirical test: $n = 60$.

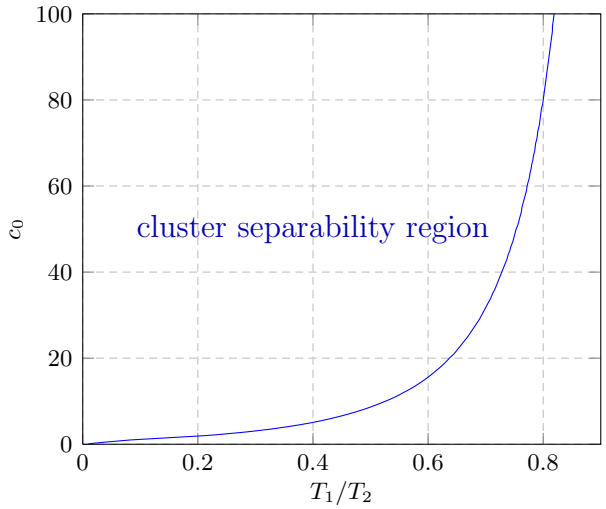

Figure 2.3: Limiting ratio c_0 to ensure separability of (T_1, T_2), $T_1 \leq T_2$, $K = 2$, $c_1 = c_2$.

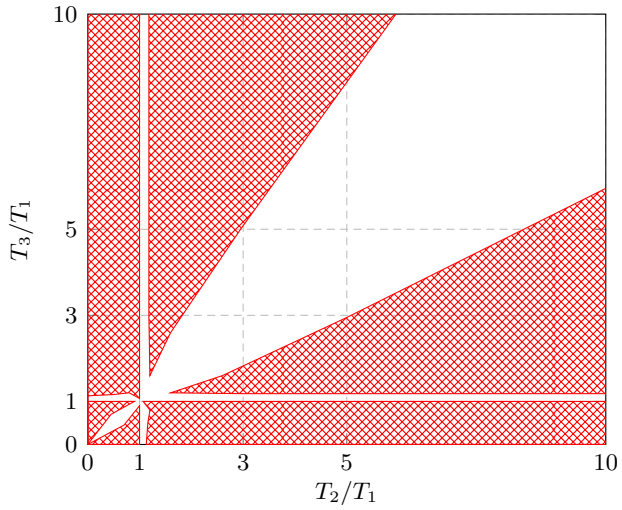

Figure 2.4: Subset of (T_1, T_2, T_3) that fulfills condition (2.6) $K = 3$, $c_1 = c_2 = c_3$, for $c_0 = 10$.

Since the other T_i's, $i \neq k$, are not inside the contour, this can be further expressed as

$$T_k = c_k \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \sum_{r=1}^{K} \frac{1}{c_r} \frac{\omega}{T_r - \omega} d\omega \tag{2.10}$$

in which we recognize an expression close to the Stieltjes transform of the eigenvalue distribution of $\mathbf{T}_n$. It now suffices to identify a satisfying integration contour $\mathcal{C}_k$. This integration contour is in fact taken to be $-1/m_F(\Gamma_k)$ for Γ_k revolving around the cluster corresponding to P_k in the limiting spectrum F of $\mathbf{B}_n$. It turns out that, for any such choice of Γ_k, $\mathcal{C}_k = -1/m_F(\Gamma_k)$ revolves around P_k and P_k only, which is exactly what we want. Now, from [7], as $N, n \to \infty$,

$$m_F(z) - \hat{m}(z) \xrightarrow{\text{a.s.}} 0 \tag{2.11}$$

where $\hat{m}(z)$ is the empirical Stieltjes transform of $\mathbf{B}_n$, i.e.

$$\hat{m}(z) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\lambda_i - z} \tag{2.12}$$

with $\lambda_1, \ldots, \lambda_N$ the eigenvalues of $\mathbf{B}_n$. We also have that the number of eigenvalues in the cluster corresponding to P_k in the spectrum of $\mathbf{B}_n$ is exactly equal to n_k as N grows large. This is referred to as the exact separation property [11].

Verifying that $m(z)$ is bounded over Γ_k, we have that

$$T_k - \hat{T}_k \xrightarrow{\text{a.s.}} 0 \tag{2.13}$$

where

$$\hat{T}_k = c_k \frac{1}{2\pi i} \oint_{\hat{\mathcal{C}}_k} \sum_{r=1}^{K} \frac{1}{c_r} \frac{\omega}{T_r - \omega} d\omega \tag{2.14}$$

where $\hat{\mathcal{C}}_k$ is now the image of $-1/\hat{m}(\Gamma_k)$. Using the change of variable $\omega = -1/\hat{m}(z)$, the calculus of (2.14) is then a mere residue calculus. From the exact separation property, the values inside the integration contour can be exactly determined.

Under separability condition, we then have our main result, originating from the work of Mestre in [9], as follows,

Theorem 3 *Let $\mathbf{B}_n \in \mathbb{C}^{N\times N}$ be defined as in Theorem 2, and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$, $\lambda_1 < \ldots < \lambda_N$, be the vector of the ordered eigenvalues of $\mathbf{B}_n$. Further assume that the limiting ratios $c_1, \ldots, c_K$ and $\mathbf{T}_n$ are such that the asymptotic spectrum separability condition is fulfilled for the eigenvalue T_k. Then, as N, n, grow large, we have*

$$\hat{P}_k - P_k \xrightarrow{\text{a.s.}} 0 \tag{2.15}$$

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

$$\hat{P}_k = \frac{N}{n_k} \sum_{i\in\mathcal{N}_k} (\lambda_i - \mu_i) \tag{2.16}$$

in which $\mathcal{N}_k = \{N-\sum_{i=k}^{K} n_i+1, \ldots, N-\sum_{i=k+1}^{K} n_i\}$ is the set of indexes matching the cluster corresponding to P_k and $(\mu_1, \ldots, \mu_N)$ are the ordered eigenvalues of the matrix $\operatorname{diag}(\boldsymbol{\lambda}) - \frac{1}{M}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$.

In fact, $(\mu_1, \ldots, \mu_N)$ are the zeros of the Stieltjes transform of the eigenvalue distribution of $\mathbf{B}_n$. This result is more interesting than the moment approach for the following reasons:

- an estimate of the eigenvalues themselves is given, instead of results on the moments of the total eigenvalue distribution.
- it is possible to isolate the estimation of individual eigenvalues, while the moment approach requires a joint estimation of all eigenvalues necessarily.
- the approach relies on the whole distribution of the eigenvalues of $\mathbf{T}$ and $\mathbf{B}_n$, or equivalently on their Stieltjes transform. As a consequence, more information than that of the first moments is used. Moment-based approach would be equivalent only if a large number of moments are computed. However, we already discussed the high computational complexity of such method.
- the practical method, that can be derived from Theorem 3 is computationally cheap.

However, while the moment-based methods are not constrained by spectrum separability, it is not possible to estimate eigenvalues that are too close to one another with the Stieltjes transform approach; see again Figures 2.3 and 2.4. If the constraint is not satisfied, the blind application of Theorem 3 on the empirical eigenvalues found in a cluster returns an estimate of the averaged value of the T_i's associated to this cluster. For practical purposes, as discussed below, this might be good enough.

The purpose of the next section is to apply the eigen-inference method derived in this section to the cognitive detection of the power of transmit sources.

2.3 Application to power inference in orthogonal CDMA networks

Let us consider the CDMA network of Chapter 1, with the model

$$\mathbf{Y} = \mathbf{W}\mathbf{P}^{\frac{1}{2}}\mathbf{S} + \sigma\mathbf{N} \tag{2.17}$$

But now, instead of taking $\mathbf{W} \in \mathbb{C}^{N\times n}$, $N \geq n$, we expand $\mathbf{W}$ into a unitary $N \times N$ matrix, and we therefore adapt $\mathbf{P}$ to an $N \times N$ matrix with additional zero diagonal entries and $\mathbf{S}$ into an $N \times L$ matrix with i.i.d. entries of zero mean and variance $1/L$. The model is then strictly equivalent to that of Chapter 1.

We recall now that $\mathbf{P}$ is composed of a finite number K of distinct eigenvalues, plus now additional null eigenvalues. This translates the fact that the different users are grouped into cellular clusters. We wish, as previously, to estimate the distinct transmit powers $P_1, \ldots, P_K$.

Note that the model of $\mathbf{Y}^{\mathsf{H}}\mathbf{Y}$ can now be rewritten in the form

$$\mathbf{Y}^{\mathsf{H}}\mathbf{Y} = \begin{bmatrix}\mathbf{X} & \mathbf{N}\end{bmatrix} \begin{bmatrix}\mathbf{P}^{\frac{1}{2}}\mathbf{W} \\ \sigma\mathbf{I}_N\end{bmatrix} \begin{bmatrix}\mathbf{W}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I}_N\end{bmatrix} \begin{bmatrix}\mathbf{X} \\ \mathbf{N}\end{bmatrix} \tag{2.18}$$

where the eigenvalue distribution of the sample covariance matrix $\begin{bmatrix}\mathbf{P}^{\frac{1}{2}}\mathbf{W}^{\mathsf{H}} \\ \sigma\mathbf{I}_N\end{bmatrix} \begin{bmatrix}\mathbf{W}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I}_N\end{bmatrix}$ is the same as that of the matrix $\mathbf{W}\mathbf{P}\mathbf{W}^{\mathsf{H}} + \sigma^2\mathbf{I}_N$ up to a mass in zero. From the fact that $\mathbf{W}$ is unitary, we then have that the eigenvalues of $\mathbf{W}\mathbf{P}\mathbf{W}^{\mathsf{H}} + \sigma^2\mathbf{I}_N$ are also those of $\mathbf{P} + \sigma^2\mathbf{I}_N$, which are then those of $\mathbf{P}$ translated by σ^2. It is then possible, from the above scheme to estimate the values of P_1, P_2, $\ldots$, P_K, assuming that two subsequent values of $P_i + \sigma^2$ and σ^2 satisfy the condition given in 1.

Note here that the relative distance

$$\frac{(P_i + \sigma^2) - \sigma^2}{P_i + \sigma^2} \tag{2.19}$$

between the eigenvalue $P_i + \sigma^2$ and the eigenvalue σ^2 becomes increasingly small as σ^2 grows large. Intuitively, this translates into the fact that the cluster corresponding to the eigenvalue σ^2 will ultimately melt into the cluster associated to $P_i + \sigma^2$ as σ^2 grows large (for fixed c_0). The order of magnitude of σ^2 such that no cluster overlapping happens is P_i. As a consequence, the Stieltjes transform method is only capable of estimating the power P_i provided that σ^2 is of order of magnitude smaller than P_i; of course, a more precise application of the separability condition is demanded to identify the exact threshold where P_i can be estimated reliably. Note also that, not only does the cluster associated to σ^2 is prone to overlap the cluster of $P_i + \sigma^2$ but also the subsequent clusters of $P_{i-1} + \sigma^2$ and $P_{i+1} + \sigma^2$ have relative distance with respect to $P_i + \sigma^2$ going small as σ^2 grows large. Consequently, depending on c_0, even smaller σ^2 can be large enough to break the separability requirement.

2.4 Conclusion

The analytical Stieltjes transform-based method is then more suited than the moment-based approach when it comes to infer the individual eigenvalues of the population matrix in a sample covariance matrix model. Moreover, the final results derived here lead to computationally cheap algorithms, which are already optimal, while moment-based approaches are only optimal when an infinite number of moments is evaluated. However, the constraint on the cluster separability is a strong limitation of the method. This limitation is somewhat controlled by the fact that we introduced here a necessary and sufficient condition to ensure spectrum separability. In fine, this condition allows the experimenter to define the ratio c_0 necessary to ensure a given detection

sensitivity. Nonetheless, for practical purposes, it might often turn out that the experimenter does not have the luxury to increase the ratio c_0 up to the desired level. In this case, alternative solutions must be found, which we discuss in the next chapter.

Chapter 3

Conclusion and further studies

3.1 Conclusion

It appears through the first steps in this study that both separate fields of random matrix theory, free probability and the Stieltjes transform approaches, are capable of coping with inverse problems that implicate large dimensional matrix models. It turns out that the now well-documented moment approaches can still be exploited and is open to further contributions, as exemplified by the recent extension to the information plus noise model [3], [4]. A large range of models of sums products and information plus noise-type of matrices can be easily treated using combinatorics methods. Specifically, specific network parameters can be recovered from the estimated moments of the asymptotic distribution of these parameters.

It has been shown however that the moment approaches have severe limitations, particularly when it comes to application and cheap eigen-inference algorithms. On the contrary, the Stieltjes transform approach, which relies though on more complex non-systematic mathematical derivations, is able to produce very cheap consistent estimators of the individual parameters under study (instead the moments of the joint distribution of all parameters). The main limitation of this approach, though, is connected to asymptotic properties of the underlying parameter distribution. It is however possible to identify these difficulties and anticipate the robustness of the estimators; specifically, it is possible to anticipate the number of time samples, or the number of sensors required to perform an accurate parameter estimation.

3.2 Contributions

At time $t_0 + 24$ of this project, the following contributions were made,

- In [3], the mathematical framework for free deconvolution of the information-plus-noise model is developed. Theorem 1 is proven accurately here.
- In [4], applications of [3] to wireless communications are developed. In particular, the above CDMA user power inference above is thoroughly detailed in this contribution, with first results, using the Newton-Girard inversion approach.

- In [12], the free deconvolution framework is applied to the capacity of multi-antenna (MIMO) communications with imperfect channel state information.
- In [13], the free deconvolution method is applied to the problem of blind detection of users in a MIMO network.
- In [8], the aforementioned criterion for spectrum separability is discussed and applied to the problem of multiple source detection.
- In [14], the theoretical fundamentals of source detection under limited knowledge at the sensor are discussed and applied to the case of source detection in a finite-antenna MIMO environment. This work is slightly orthogonal to the previous contributions as it regards eigen-inference from a finite matrix size viewpoint in order to develop optimal (instead of sub-optimal) Neyman-Pearson detection tests.

3.3 Further studies

The above studies can be further extended along the following lines

- the free deconvolution framework is currently being extended to more structured matrices. Specifically, it seems possible to widen the scope of wireless communication models, where free probability can be applied, to more involved matrix models, such as Vandermonde matrices which are used to model the phase differences of plane waves arriving at a linear antenna array. Generally speaking, it is believed that some core symmetry property of matrix models are responsible for the validity of free probability methods. The extension of the work on Gaussian and unitary matrices to these matrices will be investigated.
- it appears that, although the asymptotic inference methods provide appreciable results even for rather small matrix dimensions, it is often the case that the application of asymptotic methods to matrices of size 4×4 shows a non-negligible performance decay. The extension of the free deconvolution methods to finite-size random matrices is also being currently investigated, with first promising results. In that case, the precision of the above moment-based eigen-inference methods will be dramatically increased for systems of smaller size.
- the system models involved in the Stieltjes-transform approach can be extended to more involved models. Consider for instance

$$\mathbf{Y} = \sum_{k=1}^{K} \mathbf{H}_k \sqrt{P_k} \mathbf{X}_k + \sigma \mathbf{N} \tag{3.1}$$

  which models the simultaneous data transmission of K multi-antenna (with $n_1, \ldots, n_K$ antennas) transmitters with powers $P_1, \ldots, P_K$ through the respective MIMO channels $\mathbf{H}_1, \ldots, \mathbf{H}_K$, and additive white Gaussian noise σ^2. This model can be written into the denser form

$$\mathbf{Y} = \mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X} + \sigma \mathbf{N} \tag{3.2}$$

  by concatenating $\mathbf{H} = [\mathbf{H}_1 \ldots \mathbf{H}_K]$, $\mathbf{X}^{\mathsf{T}} = [\mathbf{X}_1^{\mathsf{H}}, \ldots, \mathbf{X}_N^{\mathsf{H}}]$ and with $\mathbf{P}$ the diagonal matrix with n_1 first entries P_1, n_2 subsequent entries P_2 etc. This model is particularly suited to

the cognitive estimation of the transmit powers by opportunistic users of the users of a primary (licensed) network.

- in the particular case of the Stieltjes transform approach when the separability condition is not met, it is believed that much more than a mere estimation of the averaged parameter values can be done. In particular, the properties of the asymptotic spectrum can be further studied to extract features that allow to perform the complex integration calculus.

- finite matrix-size considerations for the Stieltjes transform approach are also considered to be valuable for matrices of small size. However, computing the trailing terms in finite dimensions is a very involved problem, unless the random matrices under study are Gaussian matrices. The subject of the extension of the Stieltjes transform approach will then follow this requirement.

Bibliography

[1] D. V. Voiculescu, K. J. Dykema, A. Nica, "Free random variables," American Mathematical Society, 1992.

[2] R. Speicher, "Combinatorial theory of the free product with amalgamation and operator-valued free probability theory," Mem. A.M.S., vol. 627, 1998.

[3] Ø. Ryan, M. Debbah, "Multiplicative free convolution and information-plus-noise type matrices," submitted to Annals of Probability Theory, arxiv preprint math/0702342, 2007.

[4] Ø. Ryan, M. Debbah, "Free Deconvolution for Signal Processing Applications," IEEE International Symposium on Information Theory, pp. 1846-1850, 2007.

[5] R. Seroul, D. O'Shea, "Programming for Mathematicians," Springer, 2000.

[6] Z. Bai, J. W. Silverstein, "Spectral Analysis of Large Dimensional Random Matrices," Second Edition, Springer, 2010.

[7] J. W. Silverstein, 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.

[8] R. Couillet, J. W. Silverstein, M. Debbah, "Eigen-Inference for Multi-User Power Detection," submitted to IEEE Trans. on Information Theory.

[9] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," IEEE trans. on Information Theory, vol. 54, no. 11, pp. 5113-5129, 2008.

[10] W. Rudin, *Real and complex analysis*, McGraw-Hill, 2006.

[11] J. Silverstein, Z. Bai, "Exact separation of eigenvalues of large dimensional sample covariance matrices" Annals of Probability, vol. 27, no. 3, pp. 1536-1555, 1999.

[12] Ø. Ryan, M. Debbah, "Channel capacity estimation using free probability theory," IEEE Transactions on Signal Processing, vol. 56, no. 11, pp. 5654-5667, 2008.

[13] R. Couillet, M. Debbah, "Free deconvolution for OFDM multicell SNR detection", PIMRC 2008, Cannes, France.

[14] R. Couillet, M. Debbah, "R. Couillet, M. Debbah, "Bayesian Inference for Multiple Antenna Cognitive Receivers", WCNC 2009, Budapest, Hungary.