SIGNAL PROCESSING APPLICATIONS OF FREE PROBABILITY THEORY

Øyvind Ryan, and Mérouane Debbah

Department of Informatics, University of Oslo, P.O Box 1080 Blindern, NO-0316 Oslo, Norway
email: oyvindry@ifi.uio.no. web: www.ifi.uio.no/∼oyvindry

SUPELEC, Gif-sur-Yvette, France
email: merouane.debbah@supelec.fr

ABSTRACT

Situations in many fields of research, such as digital communications, nuclear physics and mathematical finance, can be modelled with random matrices. When the matrices get large, free probability theory is an invaluable tool for describing the asymptotic behaviour of many systems. It will be shown how free probability can be used to aid in source detection for certain systems. Sample covariance matrices for systems with noise are the starting point in our source detection problem. Multiplicative free deconvolution is shown to be a method which can aid in expressing limit eigenvalue distributions for sample covariance matrices, and to simplify estimators for eigenvalue distributions of covariance matrices.

1. INTRODUCTION

Random matrices, and in particular limit distributions of sample covariance matrices, have proved to be a useful tool for modelling systems, for instance in digital communications [1], nuclear physics [2] and mathematical finance [3]. A typical random matrix model is the information-plus-noise model,

$$\mathbf{W}_n = \frac{1}{N}(\mathbf{R}_n + \sigma\mathbf{X}_n)(\mathbf{R}_n + \sigma\mathbf{X}_n)^H. \tag{1}$$

$\mathbf{R}_n$ and $\mathbf{X}_n$ are assumed independent random matrices of dimension $n \times N$ throughout the paper, where $\mathbf{X}_n$ contains i.i.d. standard (i.e. mean 0, variance 1) complex Gaussian entries. (1) can be thought of as the sample covariance matrices of random vectors $\mathbf{r}_n + \sigma\mathbf{x}_n$, with $\mathbf{r}_n$ a vector containing the system characteristics (direction of arrival for instance in radar applications or impulse response in channel estimation applications)a, and $\mathbf{x}_n$ additive noise, with σ a measure of the strength of the noise. Throughout the paper, n and N will be increased so that $\lim_{n\to\infty} \frac{n}{N} = c$, i.e. the number of observations is increased at the same rate as the number of parameters of the system. This is typical of many situations arising in signal processing applications where one can gather only a limited number of observations during which the characteristics of the signal do not change.

The situation motivating our problem is the following: Assume that N observations are taken by n sensors. Observed values at each sensor may be the result of an unknown number of sources with unknown origins. In addition, each sensor is under the influence of noise. The sensors thus form a random vector $\mathbf{r}_n + \sigma\mathbf{x}_n$, and the observed values form a realization of the sample covariance matrix $\mathbf{W}_n$. Based on the fact that $\mathbf{W}_n$ is known, one is interested in inferring as much as possible about the random vector $\mathbf{r}_n$, and hence on the system (1). One would like to connect the following quantities:

1. The eigenvalue distribution of $\mathbf{W}_n$,
2. The eigenvalue distribution of $\mathbf{\Gamma}_n = \frac{1}{N}\mathbf{R}_n\mathbf{R}_n^H$,
3. The eigenvalue distribution of the covariance matrix $\mathbf{\Theta}_n = E\left(\mathbf{r}_n\mathbf{r}_n^H\right)$.

In [4], Dozier and Silverstein explain how one can use 2) to estimate 1) by solving a given equation. However, no algorithm for solving it was provided. In fact, many applications are interested in going from 1) to 2) when attempting to retrieve information about the system. Unfortunately, [4] does not provide any hint on this direction. Recently, in [5], it is shown that the framework of [4] is an interpretation of the concept of *multiplicative free convolution*.

3) can be adressed by the G_2-*estimator* [6], which provides a consistent estimator for the Stieltjes transform of covariance matrices. G-estimators have already shown their usefulness in many applications [7] but still lack intuitive interpretations. In [5], it is also shown that the G_2-estimator can be derived within the framework of multiplicative free convolution. This provides a computational algorithm for finding 3).

Interestingly, multiplicative free convolution admits a convenient implementation; [8] describes two implementations of free convolution. An implementation of of one of these, called *combinatorial computation of free convolution* in [8] (an exact implementation of free convolution based solely on moments), will be used for simulations in this paper to address several problems related to signal processing. For communication systems, estimation of the rank of the signal subspace, channel correlation and noise variance will be addressed.

This paper is organized as follows. Section 2 presents the basic concepts needed on free probability, including multiplicative and additive free convolution and deconvolution. Section 3 states the results for systems of type (1). In particular, finding quantities 2) and 3) from quantity 1) will be addressed here. Section 4 will explain through examples and simulations the importance of the system (1) for digital communications. In the following, upper (lower boldface) symbols will be used for matrices (column vectors) whereas lower symbols will represent scalar values, $(.)^T$ will denote transpose operator, $(.)^\star$ conjugation and $(.)^H = \left((.)^T\right)^\star$ hermitian transpose. $\mathbf{I}$ will represent the identity matrix.

2. FRAMEWORK FOR FREE CONVOLUTION

Free probability [9] theory has grown into an entire field of research through the pioneering work of Voiculescu in the 1980's. The basic definitions of free probability are quite abstract, as the aim was to introduce an analogy to independence in classical probability that can be used for non-commutative random variables like matrices. These more general random variables are elements in what is called a *noncommutative probability space*. This can be defined by a pair (A, ϕ), where A is a unital $*$-*algebra* with unit I, and ϕ is a normalized (i.e. $\phi(I) = 1$) linear functional on A. The elements of A are called random variables. In all our examples, A will consist of $n \times n$ matrices or random matrices. For matrices, ϕ will be the normalized trace tr_n, defined by (for any $a \in A$) $tr_n(a) = \frac{1}{n}Tr(a) = \frac{1}{n}\sum_{i=1}^{n} a_{ii}$. The unit in these $*$-algebras is the $n \times n$ identity matrix $\mathbf{I_n}$. The analogy to independence is called freeness:

Definition 1 *A family of unital $*$-subalgebras $(A_i)_{i\in I}$ will be called a free family if*

$$\left\{ \begin{array}{c} a_j \in A_{i_j} \\ i_1 \neq i_2, i_2 \neq i_3, \cdots, i_{n-1} \neq i_n \\ \phi(a_1) = \phi(a_2) = \cdots = \phi(a_n) = 0 \end{array} \right\} \Rightarrow \phi(a_1 \cdots a_n) = 0. \tag{2}$$

A family of random variables a_i is called a free family if the algebras they generate form a free family.

Definition 2 *We will say that a sequence of random variables $a_{n1}, a_{n2}, ...$ in probability spaces (A_n, ϕ_n) converge in distribution if*

$$\lim_{n\to\infty} \phi_n(a_{nk_1}^{m_1} \cdots a_{nk_r}^{m_r})$$

exists for any $m_1, ..., m_r \in \mathbb{Z}$, $k_1, ..., k_r \in \{1, 2, ...\}$. If also

$$\lim_{n\to\infty} \phi_n(a_{nk_1}^{m_1} \cdots a_{nk_r}^{m_r}) = \phi(a_{k_1}^{m_1} \cdots a_{k_r}^{m_r})$$

for some noncommutative probability space (A, ϕ) and free random variables $a_1, a_2, ... \in (A, \phi)$, we will say that the $a_{n1}, a_{n2}, ...$ are asymptotically free.

Asymptotic freeness is a very useful concept for our purposes, since many types of random matrices exhibit asymptotic freeness when their sizes get large. For instance, consider random matrices $\frac{1}{\sqrt{n}}\mathbf{A}_{n1}, \frac{1}{\sqrt{n}}\mathbf{A}_{n2}, ...$, where the $\mathbf{A}_{ni}$ are $n \times n$ with all entries independent and standard Gaussian (i.e. mean 0 and variance 1). Then it is well-known [9] that the $\frac{1}{\sqrt{n}}\mathbf{A}_{ni}$ are asymptotically free.

When sequences of moments uniquely identify probability measures (as for compactly supported probability measures), the distributions of $a_1 + a_2$ and $a_1 a_2$ give us (when a_1 and a_2 are free) two new probability measures, which depend only on the probability measures associated with the moments of a_1, a_2. Therefore we can define two operations on the set of probability measures: *Additive free convolution* $\mu_1 \boxplus \mu_2$ for the sum of free random variables, and *multiplicative free convolution* $\mu_1 \boxtimes \mu_2$ for the product of free random variables. These operations can be used to predict the spectrum of sums or products of asymptotically free random matrices. For instance, if a_{n1} has an eigenvalue distribution which approaches μ_1 and a_{n2} has an eigenvalue distribution which approaches μ_2, one has that the eigenvalue distribution of $a_{n1} + a_{n2}$ approaches $\mu_1 \boxplus \mu_2$.

We will also find it useful to introduce the concepts of *additive and multiplicative free deconvolution*: Given probability measures μ and μ_2. When there is a unique probability measure μ_1 such that $\mu = \mu_1 \boxplus \mu_2$ ($\mu = \mu_1 \boxtimes \mu_2$), we will write $\mu_1 = \mu \boxminus \mu_2$ ($\mu_1 = \mu \boxslash \mu_2$ respectively). We say that μ_1 is the additive (respectively multiplicative) free deconvolution of μ with μ_2.

One important measure is the Marčenko Pastur law μ_c [10], characterized by the density

$$f^{\mu_c}(x) = (1 - \frac{1}{c})^+ \delta(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi c x}, \qquad (3)$$

where $(z)^+ = \max(0, z)$, $a = (1 - \sqrt{c})^2$ and $b = (1 + \sqrt{c})^2$. It is known that μ_c describes asymptotic eigenvalue distributions of *Wishart matrices*. These have the form $\frac{1}{N}\mathbf{R}\mathbf{R}^H$, where $\mathbf{R}$ is an $n \times N$ random matrix with independent standard Gaussian entries. μ_c appears as limits of such when $\frac{n}{N} \to c$ when $n \to \infty$.

An important tool for our purposes is *the Stieltjes transform* [10]. For a probability measure μ, this is the analytic function on $\mathbb{C}^+ = \{z \in C : Imz > 0\}$ defined by

$$m_\mu(z) = \int \frac{1}{\lambda - z} dF^\mu(\lambda), \qquad (4)$$

where F^μ is the cumulative distribution function of μ.

3. INFORMATION PLUS NOISE MODEL

In this section we will indicate how the quantities 2) and 3) can be found through free convolution. By the *empirical eigenvalue distribution* of an $n \times n$ random matrix $\mathbf{X}$ we will mean the random atomic measure $\frac{1}{n}(\delta(\lambda_1(\mathbf{X})) + \cdots + \delta(\lambda_n(\mathbf{X})))$, where $\lambda_1(\mathbf{X}), ..., \lambda_n(\mathbf{X})$ are the (random) eigenvalues of $\mathbf{X}$. In [5], the following was shown:

Theorem 1 *Assume that the empirical eigenvalue distribution of $\mathbf{\Gamma}_n = \frac{1}{N}\mathbf{R}_n\mathbf{R}_n^H$ converges in distribution almost surely to a compactly supported probability measure μ_Γ. Then we have that the empirical eigenvalue distribution of $\mathbf{W}_n$ also converges in distribution almost surely to a compactly supported probability measure μ_W uniquely identified by*

$$\mu_W \boxslash \mu_c = (\mu_\Gamma \boxslash \mu_c) \boxplus \mu_{\sigma^2 I}. \qquad (5)$$

Theorem 1 addresses the relationship between 1) and 2), through deconvolution of (5) to $\mu_\Gamma = ((\mu_W \boxslash \mu_c) \boxminus \mu_{\sigma^2 I}) \boxtimes \mu_c$ and $\mu_W = ((\mu_\Gamma \boxslash \mu_c) \boxplus \mu_{\sigma^2 I}) \boxtimes \mu_c$.

To estimate the covariance matrices 3), general statistical analysis of observations, also called *G-analysis* [7], will be used. This is a mathematical theory for studying complex systems, where the number of parameters of the considered mathematical model can increase together with the growth of the number of observations of the system. The mathematical models which in some sense approach the system are called *G-estimators*. We use N for the number of observations of the system, and n for the number of parameters of the mathematical model. The condition used in G-analysis expressing the growth of the number of observations vs. the number of parameters in the mathematical model, is called the *G-condition*. In this paper this is $\lim_{n\to\infty} \frac{n}{N} = c$.

We restrict our analysis to systems where a number of i.i.d. random vector observations are taken. If a random vector has length n, we will use the notation $\mathbf{\Theta}_n$ to denote the covariance. Girko calls an estimator for the Stieltjes transform of covariance matrices a G_2-estimator. In chapter 2.1 of [6] he introduces a candidate $G_{2,n}(z)$ for a G_2-estimator through a complex equation of functions, and shows that it in some cases approaches the true Stieltjes transform of the involved covariance matrices. We will not state this equation here, but instead use the following result from [5]:

Theorem 2 *the following holds for real $z < 0$:*

$$G_{2,n}(z) = m_{\mu_{\Gamma_n} \boxslash \mu_c}(z) \qquad (6)$$

Theorem 2 shows that multiplicative free convolution can be used to estimate the covariance of systems, and also explains the importance of the Marčenko pastur law in terms of free (de)convolution. The implementation used by simulations in this paper works only for the case of (de)convolution with the Marčenko pastur law, as the implementation is simplest in this case. How to implement (de)convolution for the more general case is unknown to the authors, and may be much harder. Note that the G_2-estimator appears in (5) in theorem 1.

In this paper, the difference between a probability measure, μ, and an estimate of it, ν, will be measured in terms of the *Mean Square Error of the moments* (MSE). If the moments of $\int x^k d\mu(x)$, $\int x^k d\nu(x)$ are denoted by μ_k, ν_k, respectively, the MSE is defined by

$$\sum_{k\le n} |\mu_k - \nu_k|^2 \qquad (7)$$

for some number n.

4. APPLICATIONS TO SIGNAL PROCESSING

In this section, we provide several applications of free deconvolution and show how the framework can be used.

4.1 Estimation of power and the number of users

In communication applications, one needs to determine the number of users in a cell in a CDMA type network as well the power with which they are received (linked to the path loss). Denoting by n the spreading length, the received vector at the base station in an uplink CDMA system is given by:

$$\mathbf{y}_i = \mathbf{W}\mathbf{P}^{\frac{1}{2}}\mathbf{s}_i + \mathbf{b}_i \qquad (8)$$

where $\mathbf{y}_i$, $\mathbf{W}$, $\mathbf{P}$, $\mathbf{s}_i$ and $\mathbf{b}_i$ are respectively the $n \times 1$ received vector, the $n \times N$ spreading matrix with i.i.d zero mean, $\frac{1}{n}$ variance Gaussian entries, the $N \times N$ diagonal power matrix, the $N \times 1$ i.i.d gaussian unit variance modulation signals and the $n \times 1$ additive white zero mean Gaussian noise.

Usual methods determine the power of the users by finding the eigenvalues of covariance matrix of $\mathbf{y}_i$ when the signatures (matrix $\mathbf{W}$) and the noise variance are known.

$$\boldsymbol{\Theta} = \mathbb{E}\left(\mathbf{y}_i \mathbf{y}_i^H\right) = \mathbf{W}\mathbf{P}\mathbf{W}^H + \sigma^2 \mathbf{I} \tag{9}$$

However, in practice, one has only access to an estimate of the covariance matrix and does not know the signatures of the users. One can solely assume the noise variance known. In fact, usual methods compute the sample covariance matrix (based on L samples) given by:

$$\hat{\boldsymbol{\Theta}} = \frac{1}{L}\sum_{i=1}^{L} \mathbf{y}_i \mathbf{y}_i^H \tag{10}$$

and determine the number of users (and not the powers) in the cell by the non-zero eigenvalues (or up to an ad-hoc threshold for the noise variance) of $\hat{\boldsymbol{\Theta}} - \sigma^2 \mathbf{I}$. This method, referred here as classical method, is quite inadequate when L is in the same range as n. Moreover, it does not provide a method for the estimation of the power of the users.

The free deconvolution framework introduced in this paper is well suited for this case and enables to determine the power of the users without knowing their specific code structure. Indeed, the sample covariance matrix is related to the true covariance matrix $\boldsymbol{\Theta} = \mathbb{E}\left(\mathbf{y}_i \mathbf{y}_i^H\right)$ by

$$\hat{\boldsymbol{\Theta}} = \boldsymbol{\Theta}^{\frac{1}{2}} \mathbf{X}\mathbf{X}^H \boldsymbol{\Theta}^{\frac{1}{2}} \tag{11}$$

with $\mathbf{X}$ is a $n \times L$ i.i.d Gaussian zero mean matrix. Combining (11), (9), with the fact that $\mathbf{W}^H\mathbf{W}$, $\frac{1}{L}\mathbf{X}\mathbf{X}^H$ are Wishart matrices with distributions approaching $\mu_{\frac{N}{n}}$, $\mu_{\frac{n}{L}}$ respectively, and using that

$$\mu_{\mathbf{W}\mathbf{P}\mathbf{W}^H} = \frac{N}{n}\mu_{\mathbf{W}^H\mathbf{W}\mathbf{P}} + \left(1 - \frac{N}{n}\right)\delta_0,$$

we get due to asymptotic freeness the approximation

$$\left(\left(\frac{N}{n}(\mu_{\frac{N}{n}} \boxtimes \mu_{\mathbf{P}}) + \left(1 - \frac{N}{n}\right)\delta_0\right) \boxplus \mu_{\sigma^2 I}\right) \boxtimes \mu_{\frac{n}{L}} = \mu_{\hat{\boldsymbol{\Theta}}} \tag{12}$$

If one knows the noise variance, one can use this approximation in simulations in two ways:

1. Estimate the power distribution $\mu_{\mathbf{P}}$ of the users (and de facto the number of users) by isolating $\mu_{\mathbf{P}}$ on one side in (12). This can be done by performing additive and multiplicative (de)convolution on both sides: For instance, we need to perform multiplicative free deconvolution with $\mu_{\frac{n}{L}}$, and additive free deconvolution with $\mu_{\sigma^2 I}$.
2. Estimate the numbers of users N through a best-match procedure: Try values of N with $1 \leq N \leq n$. Choose the N which gives a best match between the left and right side in (12) in terms of mean square error of the moments.

To solve (12), the combinatorial computation of free convolution as described in [8] was used. In the following, a spreading length of $n = 256$ and noise variance $\sigma^2 = 0.1$ have been used.

4.1.1 Estimation of power

We use a 36×36 ($N = 36$) diagonal matrix as our power matrix $\mathbf{P}$, and use three sets of values, at 0.5, 1 and 1.5 with equal probability, so that

$$\mu_{\mathbf{P}} = \frac{1}{3}\delta_{0.5} + \frac{1}{3}\delta_1 + \frac{1}{3}\delta_{1.5}. \tag{13}$$

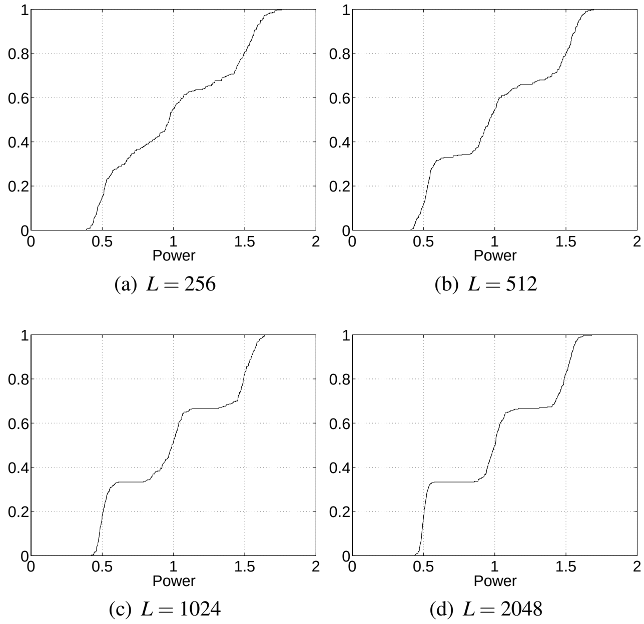


Figure 1: CDF of powers estimated from multiplicative free deconvolution from sample covariance matrices with different number of observations.

There are no existing methods for estimating such a $\mu_{\mathbf{P}}$ from the sample covariance matrices: To our knowledge, existing methods estimate the power with non-zero eigenvalues of the sample covariance matrix up to σ^2. In our case, the powers are all above σ^2.

In figure 1, the CDF of $\mu_{\mathbf{P}}$ was estimated by solving (12), using the combinatorial computation of free convolution from [8] with three moments. The resulting moments were used to compute a characteristic polynomial, from which estimates of the eigenvalues were obtained, and the CDF was computed by averaging these eigenvalues for 100 runs for each number of observations. When L increases, we get a CDF closer to that of (13).

4.1.2 Estimation of the number of users

We use a 36×36 ($N = 36$) diagonal matrix as our power matrix $\mathbf{P}$ with $\mu_{\mathbf{P}} = \delta_1$. In this case, a common method that try to find just the rank exists. This method tries the number of eigenvalues greater than some threshold above σ^2. We will set the threshold at $1.5\sigma^2$. There are no general known rules for where the threshold should be set. Choosing a wrong threshold can lead to a need for a very high number of observations for the method to be precise.

We will compare this classical method with a free convolution method for estimating the rank, following the procedure sketched in 2). The method is tested with varying number of observations, from $L = 1$ to $L = 4000$, and the number N which gives the best match with the moments of the SCM in (12) is chosen. Only the four first moments are considered. In figure 2, it is seen that when L increases, we get a prediction of N which is closer to the actual value 36. The classical method starts to predict values close to the right one only for a number of observations close to 4000. The method using free probability predicts values close to the right one for a less greater number of realizations.

4.2 Estimation of Channel correlation

In channel modelling, the modeler would like to infer on the correlation between the different degrees of the channel. These typical cases are represented by a received signal (assuming that a unit training sequence has been sent) which is given by

$$\mathbf{y}_i = \mathbf{w}_i + \mathbf{b}_i \tag{14}$$

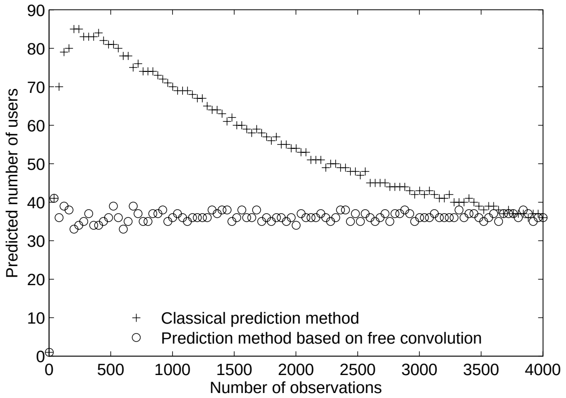


Figure 2: Estimation of the number of users with a classical method, and free convolution $L = 1024$ observations have been used.

where $\mathbf{y}_i$, $\mathbf{w}_i$ and $\mathbf{b}_i$ are respectively the $n \times 1$ received vector, the $n \times 1$ zero mean Gaussian impulse response and $n \times 1$ additive white zero mean Gaussian noise with variance σ. The cases of interest can be:

- Ultra-wide band applications [11, 12] where one measures in the frequency domain the wide-band nature of the frequency signature $\mathbf{w}_i$
- Multiple antenna applications [1] with one transmit and n receiving antennas where $\mathbf{w}_i$ is the spatial channel signature at time instant i.

Usual methods compute the sample covariance matrix given by $\hat{\mathbf{\Theta}} = \frac{1}{L}\sum_{i=1}^{L} \mathbf{y}_i \mathbf{y}_i^H$. The sample covariance matrix is related to the true covariance matrix of $\mathbf{w}_i$ by:

$$\hat{\mathbf{\Theta}} = \mathbf{\Theta}^{\frac{1}{2}} \mathbf{X}\mathbf{X}^H \mathbf{\Theta}^{\frac{1}{2}} \quad (15)$$

with $\mathbf{\Theta} = \mathbf{R} + \sigma^2 \mathbf{I}$ and $\mathbf{X}$ is an $N \times n$ i.i.d Gaussian zero mean matrix. Hence, if one knows the noise variance (measured without any signal sent), one can determine the eigenvalue distribution of the true covariance matrix following:

$$\mu_{\mathbf{R}} = (\mu_{\hat{\mathbf{\Theta}}} \boxslash \mu_{\frac{n}{L}}) \boxminus \mu_{\sigma^2 I}. \quad (16)$$

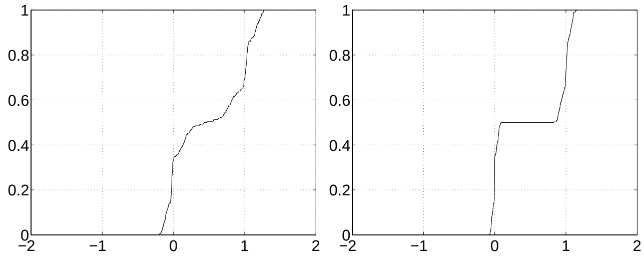
$\mu_{\mathbf{R}}$ can thus be estimated with our free convolution framework.

We use a rank K covariance matrix of the form $\mathbf{R} = diag[1, 1, .., 1, 0, .., 0]$, and variance $\sigma^2 = 0.1$, so that $\sigma \sim 0.3162$. For simulation purposes, L vectors $\mathbf{w}_i$ with covariance $\mathbf{R}$ have been generated with $n = 256$ and $K = 128$. We would like to observe the p.d.f.

$$\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1 \quad (17)$$

in our simulations. In figure 3, (16) has been solved, using $L = 128$ and $L = 512$ observations, respectively. The same strategy as in section 4.1 was used, i.e. the CDF was produced by averaging eigenvalues from 100 runs. 4 moments were computed. Both cases suggest a p.d.f. close to that of (17). It is seen that the number of observations need not be higher than the dimensions of the systems in order for free deconvolution to work.

It may also be that the true covariance matrix is known, and that we would like to estimate the noise variance through a limited number of observations. In figure 4, $L = 128$ and $L = 512$ observations have been taken. In accordance with (16), we compute $(\mu_{\mathbf{R}} \boxplus \mu_{\eta^2 I}) \boxtimes \mu_{\frac{n}{L}}$ for a set of noise variance candidates η^2, and an MSE of the four first moments of this with the moments of the observed sample covariance matrix is computed. Values of η in $(\sigma - 0.1, \sigma + 0.1) \sim (0.2162, 0.4162)$ have been tested, with a



(a) $L = 128$ (b) $L = 512$

Figure 3: CDF of eigenvalues estimated from multiplicative free deconvolution from sample covariance matrices with different number of observations.

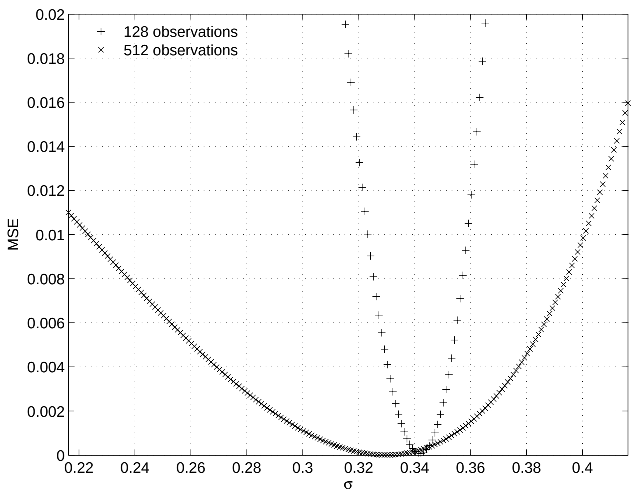


Figure 4: Estimation of the noise variance. $L = 128$ and $L = 512$ observations have been used.

spacing of 0.001. It is seen that the MMSE occurs close to the value $\sigma = \sqrt{0.1} = 0.3162$, even if the number of observations is smaller than the rank. The MMSE occurs closer to σ for $L = 512$ than for $L = 128$, so the estimate of σ improves slightly with L. It is also seen that the MSE curve for $L = 512$ lies lower than the MSE curve for $L = 128$. An explanation for this lies in the free convolution with $\mu_{\frac{n}{L}}$: As $L \to \infty$, this has the effect of concentrating all energy at 1.

5. FURTHER WORK

In this work, we have only touched upon a fraction of the potential of free deconvolution in the field of signal processing. The framework is well adapted for any problem where one needs to infer on one of the mixing matrices. Interestingly, although the results are valid in the asymptotic case, the work presented in this paper shows that it is well suited for sizes of interest for signal processing applications. The examples draw upon some basic wireless communications problems but can be extended to other cases. In particular, classical blind methods [13] which assume an infinite number of observations or noisyless problems can be revisited in light of the results of this paper.

5.1 Other types of sample matrices

One topic of interest is the use of free deconvolution with other types of matrices than the sample covariance matrix. In fact, based on a given set of observations, one can construct higher sample moment matrices than the sample covariance matrix (third product matrix for example). These matrices contain useful information that could be used in the problem. The difficult issue here is to prove freeness of the convolved measures. The free deconvolution framework could also be applied to tensor problems [14] and this has not been considered yet to our knowledge.

5.2 Colored Noise

In this work, the noise considered was supposed to be temporally and spatially white with standard Gaussian entries. This yields the Marčenko pastur law as the operand measure. However, the analysis can be extended, with the assumption that freeness is proved, to other types of noises: the case for example of an additive noise with a given correlation. In this case, the operand measure is not the Marčenko pastur law but depends on the limiting distribution of the sample noise covariance matrix.

5.3 Parametrized distribution

In the previous example (signal impaired with noise), the Marčenko Pastur law μ_c was one of the operand measures, while the other was either estimated or considered to be a discrete measure, i.e. with density $f^{\mu}(x) = \sum_{i=1}^{n} p_i \delta_{\lambda_i}(x)$. It turns out that one can find also the parameterized distribution (best fit by adjusting the parameter) that deconvolves up to certain minimum mean square error. For example, one could approximate the measure of interest with two diracs (instead of the set of n diracs) and find the best set of diracs that minimizes the mean square error. One can also approximate the measure with the Marčenko pastur law for which the parameter c needs to be optimized. In both cases, the interesting point is that the expressions can be derived explicitly.

6. CONCLUSION

In this paper, we have shown that free probability provides a neat framework for estimation problems when the number of observations is of the same order as the dimensions of the problem. In particular, we have introduced a free deconvolution framework which is very appealing from a mathematical point of view and provides an intuitive understanding of some G-estimators. Moreover, an implementation of free convolution was used in classical signal processing applications.

Acknowledgment

This project has been partially sponsored by the project IFANY (INRIA).

REFERENCES

[1] E. Telatar, “Capacity of multi-antenna gaussian channels,” *Eur. Trans. Telecomm. ETT*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[2] T. Guhr, A. Müller-Groeling, and H. A. Weidenmüller, “Random matrix theories in quantum physics: Common concepts,” *Physica Rep.*, pp. 190–, 299 1998.

[3] J.-P. Bouchaud and M. Potters, *Theory of Financial Risks-From Statistical Physics to Risk Management*. Cambridge: Cambridge University Press, 2000.

[4] B. Dozier and J. W. Silverstein, “On the empirical distribution of eigenvalues of large dimensional information-plus-noise type matrices,” *J. Multivariate Anal.*, vol. 98, no. 4, pp. 678–694, 2007.

[5] Ø. Ryan and M. Debbah, “Multiplicative free convolution and information-plus-noise type matrices,” 2007, arxiv.org/math.PR/0702342.

[6] V. L. Girko, “Ten years of general statistical analysis,” http://general-statistical-analysis.girko.freewebspace.com/chapter14.pdf.

[7] X. Mestre, “Designing good estimators for low sample sizes: random matrix theory in array processing applications,” in *12th European Signal Processing Conference, (EUSIPCO’2004)*, Sept. 2004.

[8] Ø. Ryan and M. Debbah, “Free deconvolution for signal processing applications,” *Submitted to IEEE Trans. Inform. Theory*, 2007, arxiv.org/cs.IT/0701025.

[9] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*. American Mathematical Society, 2000.

[10] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. www.nowpublishers.com, 2004.

[11] I. E. Telatar and D. N. C. Tse, “Capacity and mutual information of wideband multipath fading channels,” *IEEE Trans. Inform. Theory*, pp. 1384 –1400, Jul. 2000.

[12] D. Porrat, D. N. C. Tse, and S. Nacu, “Channel uncertainty in ultra wide band communication systems,” *IEEE Trans. Inform. Theory*, submitted 2005.

[13] D. Donoho, “On minimum entropy deconvolution,” *In Applied Time-series analysis II, p-565-609, Academic Press*, 1981.

[14] J. B. Kruskal, “Three-way arrays: rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics,” *Linear Algebra Applicat., vol. 18, pp. 95-138*, 1977.