Institut Eurecom

Free Deconvolution for Wireless Communications

Mérouane Debbah
http://www.eurecom.fr/~debbah
debbah@eurecom.fr

February, 2007
A starting example: estimation of the covariance matrix

We would like to infer on the eigenvalue of the covariance matrix.

Let
\[ y_i = x_i \]
be a zero mean Gaussian vector of size \( N \) with covariance \( R \).

One can write
\[ y_i = R^{\frac{1}{2}} u_i \]
where \( u_i \) is zero mean i.i.d Gaussian.

Usual covariance estimates compute:
\[ \hat{R} = \frac{1}{L} \sum_{i=1}^{L} y_i y_i^H = R^{\frac{1}{2}} U U^H R^{\frac{1}{2}}. \]

The non-zero eigenvalues of \( \hat{R} \) are the same as the eigenvalues of \( U U^H R \).

We know the eigenvalues of \( U U^H \) and \( \hat{R} \). Can we determine the eigenvalues of \( R \)?
Second example: Capacity estimation problem

We measure the $N \times N$ MIMO channel:

$$\hat{H} = H + W$$

where $W$ is an $N \times N$ zero mean Gaussian matrix.

The system capacity is:

$$C' = \log \det (I + \frac{\text{SNR}}{N} HH^H)$$

Usual measurements provide:

$$C' = \log \det (I + \frac{\text{SNR}}{N} \hat{H} \hat{H}^H)$$

How are the eigenvalues of $\hat{H} \hat{H}^H$ linked to the eigenvalues of $HH^H$ and $WW^H$?
Given classical random variables $a, b, c$ so that $c = ab$.

$a$ and $b$ are assumed independent.

How do we find the distribution of $b$ (moments $E(b^n)$) when only the distributions of $a$ and $c$ are given?

**Solution:** $E(c^n) = E(a^n)E(b^n)$, so that $E(b^n) = E(c^n)/E(a^n)$.

If $c = a + b$, form the moment generating functions

$$M_a(s) = E(e^{sa}), M_c(s) = E(e^{sc}).$$

The moment generating function of $b$ is then

$$M_b(s) = M_c(s)/M_a(s).$$

The distribution of $b$ can be recovered from $M_s(b)$. 
Questions

• What if we instead have independent $n \times n$ random matrices $A$ and $B$, and $C = A + B$ ($C = AB$)?

• How do we find the eigenvalue distribution of $B$ when the eigenvalue distributions of $A$ and $C$ are known?

• If the matrices are diagonal, the eigenvalue distributions are easily found. $A$ and $B$ have equal eigenvectors. The entries in the matrices are the eigenvalues, and the eigenvalues in the product are the product of the eigenvalues. So, the eigenvalues in $AB$ can be determined from the eigenvalues in $A$ and $B$.

Deconvolution for our purposes means to determine/predict the eigenvalue distribution of $B$ from that of $AB$ and $A$. 
Our Case: Given two random matrices $A_N$ and $B_N$ with limiting spectral measures $f_A$ and $f_B$, we would like to compute the limiting spectral measure for $A_N + B_N$ and $A_N B_N$ in terms of the moments of $f_A$ and $f_B$.

Difference with classical probability: $A_N$ and $B_N$ do not commute so we are in the realm of non-commutative algebra.

However, it turns out that there is a deterministic and treatable result if:

- The eigenspaces are in **generic** positions
- if $N \to \infty$
Where does the result come from: Free Probability Theory


The theory of free probability theory allows to compute the moments of the sum/product in the limit of large matrices as long as at least one of $A_N$ or $B_N$ has eigenvectors that essentially are uniformly distributed with Haar measure.

**Intuition:** With this condition, the eigenvectors of both matrices are "deconnected" (no specific structure in some sense) and become "free".

**For example,** $A_N$ and $\Theta_N B_N \Theta_N^H$, where $A_N$ and $B_N$ are independent and $\Theta_N$ is a $N \times N$ Haar Unitary matrix, are free.
More generally, it can be expected that the eigenvalue distribution of $f(A_N, B_N)$ depends only on the asymptotic eigenvalue distribution of $A_N$ and $B_N$ if:

- $A_N$ and $B_N$ are independent.
- One of them is unitarily invariant (the joint distribution of the entries does not change under unitary transform).

**Example of unitarily invariant matrices:** Gaussian and Wishart matrices.
How can we force random matrices to have uniformly distributed eigenvector structures?

An \( n \times n \) unitary random matrix is called standard if it is distributed uniformly on the set \( U(n) \) of \( n \times n \) unitary matrices w.r.t. Haar measure.

- Standard unitary matrices have uniformly distributed eigenvector structures.
- If the \( U_n \) are standard unitaries, and \( A_n \) is independent from \( U_n \), the random matrix \( U_n A_n U_n^* \) is called a random rotation.
- Even though \( A_n \) does not have a uniformly distributed eigenvector structure, \( U_n A_n U_n^* \) always has.
- Limit distributions of standard unitary random matrices are called Haar unitaries.
Can convolution be performed for other random matrices?

**Result.** Assume that $A_n$ and $B_n$ are random matrices which have a limit eigenvalue distribution (in some sense) as $n \to \infty$, and assume that either $A_n$ or $B_n$ have uniformly distributed eigenvectors with respect to Haar measure. Then $A_n B_n$ ($A_n + B_n$) also have a limit eigenvalue distribution, only depending on the limit eigenvalue distribution of $A_n$ and $B_n$.

If the conditions in the result are fulfilled, and the limit eigenvalue distribution of $A_n$ and $B_n$ are called $\mu_A$ and $\mu_B$ respectively, we will denote by

$$\mu_A \boxplus \mu_B$$

for the limit eigenvalue distributions of $A_n + B_n$, and

$$\mu_A \boxtimes \mu_B$$

for the limit eigenvalue distributions of $A_n B_n$. 
Let $A_n$ and $B_n$ be Gaussian random matrices with i.i.d entries, mean 0, variance $\frac{1}{n}$.

- $A_n$ and $B_n$ have a uniformly distributed eigenvector structure, hence $A_n B_n$, $A_n + B_n$ also have limiting eigenvalue distributions.
- The limiting eigenvalue distribution of $A_n$ is called the *full-circle law*.
- If $A_n$ is $N \times K$ (with $\lim_{n \to \infty} \frac{K}{N} = C$), the limiting eigenvalue distribution of $A_n A_n^*$ is called the *Marčenko Pastur law* with parameter $c$, also called $\mu_c$. 
Motivation for free probability

One can show that for the Gaussian random matrices we considered, the limits

$$\mu \left( A^{i_1} B^{j_1} \cdots A^{i_l} B^{j_l} \right) = \lim_{n \to \infty} tr_n \left( A^{i_1}_n B^{j_1}_n \cdots A^{i_l}_n B^{j_l}_n \right)$$

exist. If we linearly extend the linear functional $\mu$ to all polynomials in $A$ and $B$, the following can be shown:

**Result.** If $P_i, Q_i$ are polynomials in $A$ and $B$ respectively, with $1 \leq i \leq l$, and $\mu(P_i(A)) = 0, \mu(Q_i(B)) = 0$ for all $i$, then

$$\mu \left( P_1(A) Q_1(B) \cdots P_l(A) Q_l(B) \right) = 0.$$
Motivation for free probability

One can show that for two independent standard unitary random matrices $U_n, V_n,$

$$
\mu \left( U^{i_1}V^{j_1} \cdots U^{i_l}V^{j_l} \right) = \lim_{n \to \infty} tr_n \left( U^{i_1}_n V^{j_1}_n \cdots U^{i_l}_n V^{j_l}_n \right) = 0
$$

whenever $i_k, j_k \neq 0$. So, all higher moments and mixed moments are 0. These calculations motivate the following definition of freeness.
Definition of freeness

- Free probability was developed as a probability theory for random variables which do not commute, like matrices.
- Many similarities with classical probability.
- Random variables are elements in what we will call a noncommutative probability space: A pair \((A, \phi)\), where \(A\) is a unital \(*\)-algebra with unit \(I\), and \(\phi\) is a normalized (i.e. \(\phi(I) = 1\)) linear functional on \(A\).
- For matrices, \(\phi\) will be the normalized trace \(tr_n\), defined by

\[
tr_n(a) = \frac{1}{n} \sum_{i=1}^{n} a_{ii}.
\]

For random matrices, \(\phi = \tau_n\) is defined by

\[
\tau_n(a) = \frac{1}{n} \sum_{i=1}^{n} E(a_{ii}) = E(tr_n(a)).
\]

The unit in these \(*\)-algebras is the \(n \times n\) identity matrix \(I_n\).
Definition. A family of unital $\ast$-subalgebras $(A_i)_{i \in I}$ is called a free family if

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

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

- Enables one to calculate the mixed moments of free random variables.
- Enables us to define the operations $\mu_A \boxplus \mu_B$ and $\mu_A \boxtimes \mu_B$ formally, by associating the moments of a random variable $A$ with a probability measure $\mu_A$. $\mu_A \boxplus \mu_B$ is defined as the probability measure $\mu_{A+B}$ when $A$ and $B$ are random variables which are free. Similarly for $\mu_A \boxtimes \mu_B$.
- Given $\mu$ and $\mu_2$, when there exists a unique $\mu_1$ such that $\mu = \mu_1 \boxtimes \mu_2$, we will write $\mu_1 = \mu \oslash \mu_2$.
- The Gaussian/standard unitary random matrices considered are said to be asymptotically free.
- Freeness captures the limit eigenvalue distributions for combinations of matrices with uniformly distributed eigenvector structure.
Free relation

Example: Let $a$ and $b$ be in free relation with respect to $\phi$. One can understand freeness as a rule for computing the expectation of products of $a$ and $b$. Show that:

$$\phi(ab) = \phi(a)\phi(b)$$

$$\phi(abab) = \phi(a^2)\phi(b)^2 + \phi(a)^2\phi(b^2) - \phi(a)^2\phi(b)^2$$

$$\phi(ab^2a) = \phi(a^2)\phi(b^2)$$
Recursive implementations based on non-crossing partitions have been developed:

Ø. Ryan and M. Debbah, "Free deconvolution for signal processing applications", submitted IT transaction on Information Theory

What about the capacity estimation problem: Information plus Noise Type Model

\[ W_n = \frac{1}{N}(R_n + \sigma X_n)(R_n + \sigma X_n)^H. \]

- \( R_n \) and \( X_n \) are assumed independent random matrices of dimension \( n \times N \).
- \( X_n \) contains i.i.d. standard (i.e. mean 0, variance 1) complex Gaussian entries.

The can be thought of as the sample covariance matrices of random vectors \( r_n + \sigma x_n \). \( r_n \) can be interpreted as a vector containing the system characteristics (direction of arrival for instance in radar applications or impulse response in channel estimation applications). \( x_n \) represents additive noise, with \( \sigma \) a measure of the strength of the noise.
Important Result


**Result.** Assume that $\Gamma_n = \frac{1}{N} R_n R_n^H$ converge in distribution almost surely to a compactly supported probability measure $\mu_\Gamma$. Then we have that $W_n$ also converge in distribution almost surely to a compactly supported probability measure $\mu_W$ uniquely identified by

$$\mu_W \otimes \mu_c = (\mu_\Gamma \otimes \mu_c) \boxplus \mu_{\sigma^2 I}.$$
Estimation of power and the number of users

\[ y_i = WP^{\frac{1}{2}}s_i + b_i \]

where \( y_i \), \( W \), \( P \), \( s_i \) and \( b_i \) are respectively the \( n \times 1 \) received vector, the \( n \times N \) spreading matrix with i.i.d zero mean, \( \frac{1}{n} \) variance 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 $y_i$ when the signatures (matrix $W$) and the noise variance are known.

$$\Theta = \mathbb{E} \left( y_i y_i^H \right) = WPW^H + \sigma^2 I$$  \tag{2}$$

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:

$$\hat{\Theta} = \frac{1}{L} \sum_{i=1}^{L} y_i y_i^H$$  \tag{3}$$

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:

$$\hat{\Theta} - \sigma^2 I$$
Estimation of power and the number of users

The sample covariance matrix is related to the true covariance matrix \( \Theta = \mathbb{E} (y_i y_i^H) \) by:

\[
\hat{\Theta} = \Theta^{1/2} X X^H \Theta^{1/2}
\]

with

\[
\Theta = WPW^H + \sigma^2 I
\]

and \( X \) is a \( n \times L \) i.i.d Gaussian zero mean matrix. Combining (4), (4), with the fact that \( W^H W, \frac{1}{L} X X^H \) are Wishart matrices with distributions approaching \( \mu_n N, \mu_n L \) respectively, and using that

\[
\mu_{WPW^H} = \frac{N}{n} \mu_{W^H WP} + \left(1 - \frac{N}{n}\right) \delta_0,
\]

we get due to asymptotic freeness the equation

\[
\left( \left( \frac{N}{n} (\mu_n \otimes \mu_P) + \left(1 - \frac{N}{n}\right) \delta_0 \right) \boxplus \mu \sigma^2 I \right) \otimes \mu_n L = \hat{\mu}_R
\]
Simulations

In the following simulations, a spreading length of $n = 256$ and noise variance $\sigma^2 = 0.1$ have been used.

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

$$\mu_P = \frac{1}{3} \delta_{0.5} + \frac{1}{3} \delta_1 + \frac{1}{3} \delta_{1.5}.$$
Simulations

Figure 1: Distribution of the powers estimated from multiplicative free deconvolution from sample covariance matrices with $L = 2048$. 
Figure 2: Distribution of the powers estimated from multiplicative free deconvolution from sample covariance matrices with $L = 2048$. 
We use a $36 \times 36 (N = 36)$ diagonal matrix as our power matrix $P$ with $\mu_P = \delta_1$. We would like to determine the number of users.

![Figure 3: Estimation of the number of users with a classical method, and free convolution $L = 1024$ observations have been used.](image)
THANK YOU!