Supelec

Random Matrix Theory for Wireless Communications

Mérouane Debbah
http://www.supelec.fr
merouane.debbah@supelec.fr

February, 2008

Presentation

Stieltjes Transform Method

Important Lemma

Z. D. Bai and J. W. Silverstein, " No eigenvalue outside the support of the limiting spectral distribution of large dimensional sample covariance matrices", Ann. Probab. vol. 26, no. 1, pp. 316-345, 1998.

Let $\mathbf{A}$ be a deterministic $N \times N$ complex matrix with uniformly bounded spectral radius for all N. Let $\mathbf{x} = \frac{1}{\sqrt{N}}[x_1, \ldots, x_N]^T$ where the $\{x_i\}$ are i.i.d complex random variables with zero mean, unit variance and finite eight moment. Then:

$$\mathbb{E}\left[\mid \mathbf{x}^H \mathbf{A}\mathbf{x} - \frac{1}{N}\mathrm{Trace}\mathbf{A} \mid^4 \right] \leq \frac{C}{N^2}$$

where C is constant that does not depend on N or $\mathbf{A}$.

Corollary. This result ensures that

$$\mathbf{x}^H \mathbf{A}\mathbf{x} - \frac{1}{N}\mathrm{Trace}\mathbf{A} \to 0$$

almost surely.

Definition

Uniformly Bounded Spectral Radius: There exists a real number, independent of N, which bounds the magnitudes of the eigenvalues of $\mathbf{A}$ for all N.

Why Eighth Moment Bounded?

The proof proceeds by induction on p on the following inequality:

$$\mathbb{E} \mid \mathbf{x}^H \mathbf{A} \mathbf{x} - \frac{1}{N} \mathrm{Trace}(\mathbf{A}) \mid^p \leq K_p \left(\left(\mathbb{E} \mid x_1 \mid^4 \mathrm{Trace}(\mathbf{A}\mathbf{A}^H) \right)^{\frac{p}{2}} + \mathbb{E} \mid x_1 \mid^{2p} \mathrm{Trace}(\mathbf{A}\mathbf{A}^H)^{\frac{p}{2}} \right)$$

An Even Stronger Result holds

D.N.C. Tse and O. Zeitouni, "Linear multiuser receivers in random environments", IEEE Transactions on Information Theory, 46(1): 171-188, Jan. 2000

A stronger result was shown to hold:

$$\sqrt{N}\left[\mathbf{x}^H\mathbf{A}\mathbf{x} - \frac{1}{N}\mathrm{Trace}\mathbf{A}\right] \rightarrow \mathcal{N}(0, \sigma^2)$$

Elements of proof for the Marchenko-pastur law

Write

$$\mathbf{W} = \begin{bmatrix} \mathbf{u} \\ \mathbf{U} \end{bmatrix}$$

where $\mathbf{W}$ is of size $N \times K$ and $\mathbf{u}$ is of size $1 \times K$ with centered i.i.d elements and variance $\frac{1}{N}$.

$$\left(\mathbf{W}\mathbf{W}^H - z\mathbf{I}_N\right)^{-1} = \begin{bmatrix} \mathbf{u}\mathbf{u}^H - z & \mathbf{u}\mathbf{U}^H \\ \mathbf{U}\mathbf{u}^H & \mathbf{U}\mathbf{U}^H - z\mathbf{I}_{N-1} \end{bmatrix}^{-1}$$

Element $(1,1)$ of $\left(\mathbf{W}\mathbf{W}^H - z\mathbf{I}_N\right)^{-1}$ can be rewritten (using inversion formula of block matrices):

$$\left[\left(\mathbf{W}\mathbf{W}^H - z\mathbf{I}_N\right)^{-1}\right]_{11} = \frac{1}{\mathbf{u}\mathbf{u}^H - z - \mathbf{u}\mathbf{U}^H\left(\mathbf{U}\mathbf{U}^H - z\mathbf{I}_{N-1}\right)^{-1}\mathbf{U}\mathbf{u}^H}.$$

Elements of proof for the Marchenko-pastur law

Using the identity:

$$\mathbf{I}_K - \mathbf{U}^H \left(\mathbf{U}\mathbf{U}^H - z\mathbf{I}_{N-1}\right)^{-1} \mathbf{U} = -z \left(\mathbf{U}^H\mathbf{U} - z\mathbf{I}_K\right)^{-1}$$

We obtain:

$$\left[\left(\mathbf{W}\mathbf{W}^H - z\mathbf{I}_N\right)^{-1}\right]_{11} = \frac{1}{-z - z\mathbf{u}\left(\mathbf{U}^H\mathbf{U} - z\mathbf{I}_K\right)^{-1}\mathbf{u}^H}.$$

Since $\mathbf{u}$ has i.i.d elements and is independent of $\left(\mathbf{U}^H\mathbf{U} - z\mathbf{I}_K\right)^{-1}$, we get:

$$\left[\left(\mathbf{W}\mathbf{W}^H - z\mathbf{I}_N\right)^{-1}\right]_{11} \simeq \frac{1}{-z - z\frac{1}{N}\mathrm{Trace}\left(\mathbf{U}^H\mathbf{U} - z\mathbf{I}_K\right)^{-1}}.$$

when $N \to \infty$.

Elements of proof for the Marchenko-pastur law

Since $\mathbf{W}^H\mathbf{W} = \mathbf{U}^H\mathbf{U} + \mathbf{u}^H\mathbf{u}$ (rank 1 perturbation), then for high values of N,

$$\frac{1}{N}\mathrm{Trace}\left(\mathbf{U}^H\mathbf{U} - z\mathbf{I}_K\right)^{-1} \simeq \frac{1}{N}\mathrm{Trace}\left(\mathbf{W}^H\mathbf{W} - z\mathbf{I}_K\right)^{-1}$$

Moreover, one can easily verify that:

$$\frac{1}{N}\mathrm{Trace}\left(\mathbf{W}^H\mathbf{W} - z\mathbf{I}_K\right)^{-1} = \frac{1}{N}\mathrm{Trace}\left(\mathbf{W}\mathbf{W}^H - z\mathbf{I}_N\right)^{-1} + \frac{N-K}{N}\frac{1}{z}$$

which yields:

$$\begin{aligned}\left[\left(\mathbf{W}\mathbf{W}^H - z\mathbf{I}_N\right)^{-1}\right]_{11} &\simeq \frac{1}{\alpha - 1 - z - z\frac{1}{N}\mathrm{Trace}\left(\mathbf{W}\mathbf{W}^H - z\mathbf{I}_N\right)^{-1}} \\ &= \frac{1}{\alpha - 1 - z - zG_{\mathbf{W}\mathbf{W}^H}(z)}\end{aligned}$$

Elements of proof for the Marchenko-pastur law

Remark.

$$G_{\mathbf{W}\mathbf{W}^H}(z) = \frac{1}{N}\mathrm{Trace}\left(\mathbf{W}\mathbf{W}^H - z\mathbf{I}_N\right)^{-1}$$

$$G_{\mathbf{W}^H\mathbf{W}}(z) = \frac{1}{K}\mathrm{Trace}\left(\mathbf{W}^H\mathbf{W} - z\mathbf{I}_K\right)^{-1}$$

Elements of proof for the Marchenko-pastur law

Note that the equality is independent of the index of the diagonal term ($\alpha = \frac{K}{N}$ when $N \to \infty$):

$$\left[\left(\mathbf{W}\mathbf{W}^H - z\mathbf{I}_N\right)^{-1}\right]_{11} = \frac{1}{\alpha - 1 - z - zG_{\mathbf{W}\mathbf{W}^H}(z)}$$

Since $G_{\mathbf{W}\mathbf{W}^H}(z) = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\mathbf{W}\mathbf{W}^H - z\mathbf{I}_N\right)^{-1}\right]_{ii}$, we have for large values of N:

$$G_{\mathbf{W}\mathbf{W}^H}(z) = \frac{1}{\alpha - 1 - z - zG_{\mathbf{W}\mathbf{W}^H}(z)}$$

which admits as a solution:

$$G_{\mathbf{W}\mathbf{W}^H}(z) = \frac{\alpha - 1}{2z} - \frac{1}{2} + \frac{\sqrt{(\alpha - 1 - z)^2 - 4z}}{2z}$$

Some references using the Stieltjes Transform Method

V.A. Marchenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices." Math. USSR Sbornik, 1(4): 457-483, 1967.

J.W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices." J. Multivariate Anal. 54(2) : 175-192, 1995

V.L. Girko, "Thirty years of the central resolvent law and three laws on the $\frac{1}{n}$ expansion for resolvent of random matrices.", Random Operators and Stochastic Equations, 11(2): 167-212, 2003.

Another useful result on quadratic forms

L. Cottatellucci and R. Müller, "CDMA Systems with Correlated Spatial Diversity: A generalized Resource Pooling", to appear in IEEE transactions on Information Theory, 2007.
Z.D. Bai and J. W. Silverstein, "On the Signal-to-Interference-Ratio of CDMA Systems in Wireless Communications", To appear in Annals of Applied Probability, 2007

Theorem. Let $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_K]$ be $N \times K$ matrix with zero mean i.i.d entries and variance $\frac{1}{N}$. Define for each N and k, $\beta_k = \sqrt{T_k}[\gamma_k(1), \ldots, \gamma_k(L)]^T$ such that the empirical distribution of $\beta_1, \ldots, \beta_K$ (independent of $\mathbf{W}$) weakly converges to a probability distribution P_H in $\mathbb{C}^L$. Defined $\mathbf{x}_k = \sqrt{T_k}[\gamma_k(1)\mathbf{w}_k; \ldots; \gamma_k(L)\mathbf{w}_k]$ be a column vector of size $NL \times 1$. Then, when $N, K \to \infty$

$$\mathbf{x}_k^H \left(\sum_{j \neq k} \mathbf{x}_j \mathbf{x}_j^H - z\mathbf{I}_{NL} \right)^{-1} \mathbf{x}_k \to \beta_k^H \mathbf{A} \beta_k$$

where

$$\mathbf{A}^{-1} = \left(\alpha \mathbb{E} \left(\frac{\beta \beta^H}{1 + \beta^H \mathbf{A} \beta} \right) - z\mathbf{I}_L \right)$$

Results on Random Unitary Matrices

Haar Matrices

Definition. A square matrix $\mathbf{\Theta}$ is unitary if

$$\mathbf{\Theta}\mathbf{\Theta}^H = \mathbf{\Theta}^H\mathbf{\Theta} = \mathbf{I}$$

Definition. An $N \times N$ random matrix $\mathbf{\Theta}$ is a Haar matrix if it is uniformly distributed on the set $\mathcal{U}(N)$ of $N \times N$ unitary matrices. Its density function on $\mathcal{U}(N)$ is given by:

$$2^{-N}\pi^{-\frac{1}{2}N(N+1)}\prod_{i=1}^{N}(N-i)!$$

Characterization: For each $\mathbf{U}$ deterministic unitary, $\mathbf{\Theta}$ and $\mathbf{U}\mathbf{\Theta}$ have the same distribution (left invariance characterizes the Haar measure on a compact group).
Lemma. The eigenvalues λ_i for $i = 1, \ldots, N$ of an $N \times N$ Haar matrix lie on the unit circle i.e $\lambda_i = e^{j\theta_i}$ and their joint p.d.f is

$$\frac{1}{N!}\prod_{i<l} \mid \lambda_i - \lambda_l \mid^2$$

The Birth of the Haar measure

Alfred Haar, 1885-1933

Haar Matrices

Generation. $\boldsymbol{\Theta} = \mathbf{W}(\mathbf{W}^H\mathbf{W})^{-1/2}$ where $\mathbf{W} = [w_{i,j}]_{1\leq i,j\leq N}$ has i.i.d complex Gaussian unit variance entries is Haar distributed.

To see this, notice that for each constant unitary matrix $\mathbf{U}$,

$$\mathbf{UW}(\mathbf{W}^H\mathbf{W})^{-1/2} = \mathbf{UW}[(\mathbf{UW})^H\mathbf{UW}]^{-1/2} .$$

Since the probability distribution of $\mathbf{W}$ and $\mathbf{UW}$ coincide, matrices $\mathbf{W}(\mathbf{W}^H\mathbf{W})^{-1/2}$ and $\mathbf{UW}((\mathbf{UW})^H\mathbf{UW})^{-1/2}$ have the same distribution.

The above equality thus implies that $\mathbf{UW}(\mathbf{W}^H\mathbf{W})^{-1/2}$ and $\mathbf{W}(\mathbf{W}^H\mathbf{W})^{-1/2}$ are identically distributed.

Haar Matrices

There is another way for building Haar distributed unitary matrices that will be useful to our context. Instead of multiplying $\mathbf{W}$ by the inverse of the Hermitian square root of $\mathbf{W}^H\mathbf{W}$, one can introduce the uniquely defined upper triangular matrix with positive diagonal elements $\mathbf{Q}(\mathbf{X})$ defined by

$$\mathbf{W}^H\mathbf{W} = \mathbf{Q}(\mathbf{W})^H\mathbf{Q}(\mathbf{W}) \; .$$

The unitary matrix $\mathbf{\Theta}(\mathbf{W})$ defined by

$$\mathbf{\Theta}(\mathbf{W}) = \mathbf{W}\mathbf{Q}(\mathbf{W})^{-1} \qquad (1)$$

is also Haar distributed. To see this, we first remark that for each constant unitary matrix $\mathbf{U}$, the probability distribution of $\mathbf{\Theta}(\mathbf{W})$ and of $\mathbf{\Theta}(\mathbf{UW})$ coincide. But, it is obvious that $\mathbf{Q}(\mathbf{UW}) = \mathbf{Q}(\mathbf{W})$, so that $\mathbf{U}\mathbf{\Theta}(\mathbf{W}) = \mathbf{\Theta}(\mathbf{UW})$. Therefore, the probability distribution of $\mathbf{\Theta}(\mathbf{W})$ and of $\mathbf{U}\mathbf{\Theta}(\mathbf{W})$ coincide. Remark that the columns of $\mathbf{\Theta}(\mathbf{W})$ are obtained by a Gram-Schmidt orthogonalization of the column space of $\mathbf{W}$.

Haar Matrices

Useful Properties:

$$
\begin{aligned}
\mathbb{E}[\mid \theta_{ij} \mid^2] &= \frac{1}{N} \\
\mathbb{E}[\mid \theta_{ij} \mid^4] &= \frac{2}{N(N+1)} \\
\mathbb{E}[\mid \theta_{ij} \mid^2 \mid \theta_{kj} \mid^2] &= \mathbb{E}[\mid \theta_{ij} \mid^2 \mid \theta_{il} \mid^2] = \frac{1}{N(N+1)} \\
\mathbb{E}[\mid \theta_{ij} \mid^2 \mid \theta_{kl} \mid^2] &= \frac{2}{N^2-1} \\
\mathbb{E}[\theta_{ij}\theta_{kl}\theta_{il}^*\theta_{kj}^*] &= -\frac{1}{N(N^2-1)}
\end{aligned}
$$

Asymptotic Haar matrices

D. Petz and J. Réffy, "On Asymptotics of large Haar distributed unitary matrices," Periodica Math. Hungar. 49 (2004), 103-117.

Theorem. Let $\mathbf{\Theta}_N$ be a sequence of $N \times N$ Haar unitary random matrices. Then $\mathrm{Trace}\mathbf{\Theta}_N$ converges in distribution to a standard complex normal variable as $N \to \infty$.

Theorem. Let Z be a standard complex normal distributed random variable, then for the sequence of $\mathbf{\Theta}_N$ $N \times N$ Haar unitary random matrices

$$\mathrm{Trace}\mathbf{\Theta}_N^l \to \sqrt{l}Z.$$

in distribution.

Another important lemma

Let $\mathbf{\Theta}$ be a $K < N$ columns of an $N \times N$ Haar distributed random matrix and suppose θ is a column of $\mathbf{\Theta}$. Let $\mathbf{B}_N$ be a $N \times N$ random matrix, which is a function of all columns of $\mathbf{\Theta}$ except θ and $B = \sup_N \mid\mid \mathbf{B}_N \mid\mid < \infty$, then

$$\mathbb{E} \mid \theta^H \mathbf{B}_N \theta - \frac{1}{N-K} \mathrm{Tr}(\mathbf{\Pi} \mathbf{B}_N) \mid^4 \leq \frac{C}{N^2}$$

Where $\mathbf{\Pi} = \mathbf{I}_N - \mathbf{\Theta}\mathbf{\Theta}^H + \theta\theta^H$ and C is a deterministic finite constant which depends only on B and $\alpha = \frac{K}{N}$.

Proof

$\mathbf{\Theta} = [\theta_1, \ldots, \theta_K]$ extracted from a Haar matrix.

Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ be an i.i.d. Gaussian matrix.

$$\theta_1 = \frac{\mathbf{x}_1}{||\mathbf{x}_1||}$$

$$\theta_2 = \frac{\mathbf{\Pi}_1 \mathbf{x}_2}{||\mathbf{\Pi}_1 \mathbf{x}_2||}$$

where $\mathbf{\Pi}_1 = \mathbf{I} - \frac{\mathbf{x}_1 \mathbf{x}_1^H}{||\mathbf{x}_1||^2}$

$$\theta_K = \frac{\mathbf{\Pi}_{K-1} \mathbf{x}_K}{||\mathbf{\Pi}_{K-1} \mathbf{x}_K||}$$

Note that $\mathbf{\Pi}_{k-1}$ and $\mathbf{x}_k$ are stochastically independent.

Proof

Suppose $\mathbf{B}_N$ is a function of $\theta_2, \ldots, \theta_K$.

$$\theta_1^H \mathbf{B}_N \theta_1 - \frac{1}{N-K} \mathrm{Tr}(\mathbf{\Pi}_0 \mathbf{B}_N) = f_1(\theta_1, \ldots, \theta_K)$$

with

$$\mathbf{\Pi}_0 = \mathbf{I} - \mathbf{\Theta}\mathbf{\Theta}^H + \theta_1 \theta_1^H$$

Let $\mathbf{P}$ be a permutation matrix exchanging column 1 with K.

Define $\tilde{\mathbf{\Theta}} = \mathbf{\Theta}\mathbf{P}$.

It is clear that $f_K(\tilde{\theta}_1, \ldots, \tilde{\theta}_K) = f_1(\theta_1, \ldots, \theta_K)$

Since $\tilde{\mathbf{\Theta}}$ and $\mathbf{\Theta}$ have the same distribution, random variables $f_1(\theta_1, \ldots, \theta_K)$ and $f_K(\theta_1, \ldots, \theta_K)$ are identically distributed which implies:

$$\mathbb{E}\| f_1(\theta_1, \ldots, \theta_K) \|^4 = \mathbb{E}\| f_K(\theta_1, \ldots, \theta_K) \|^4$$

In the following, we focus therefore on column θ_K of $\mathbf{\Theta}$.

Proof

$\mathbb{E}\| f_1(\theta_1, \ldots, \theta_K)\|^4$ does not depend on the particular way $\mathbf{\Theta}$ is generated provided it is Haar distributed. We can therefore focus on our construction of Haar distributed matrices.

$$
\begin{aligned}
e_N &= \theta_K^H \mathbf{B}_N \theta_K - \frac{1}{N-K}\mathrm{Tr}(\mathbf{\Pi}_{K-1}\mathbf{B}_N) \\
&= \frac{\mathbf{x}_K^H \mathbf{\Pi}_{K-1}^H}{\|\mathbf{x}_K^H \mathbf{\Pi}_{K-1}^H\|}\mathbf{B}_N \frac{\mathbf{\Pi}_{K-1}\mathbf{x}_K}{\|\mathbf{\Pi}_{K-1}\mathbf{x}_K\|} - \frac{1}{N-K}\mathrm{Tr}(\mathbf{\Pi}_{K-1}\mathbf{B}_N) \\
&= e_{1,N} + e_{2,N}
\end{aligned}
$$

$$
e_{1,N} = \frac{\mathbf{x}_K^H \mathbf{\Pi}_{K-1}^H \mathbf{B}_N \mathbf{\Pi}_{K-1}\mathbf{x}_K}{\|\mathbf{x}_K \mathbf{\Pi}_{K-1}\|^2} - \frac{\mathbf{x}_K^H \mathbf{\Pi}_{K-1}^H \mathbf{B}_N \mathbf{\Pi}_{K-1}\mathbf{x}_K}{N-K}
$$

$$
e_{2,N} = \frac{\mathbf{x}_K^H \mathbf{\Pi}_{K-1}^H \mathbf{B}_N \mathbf{\Pi}_{K-1}\mathbf{x}_K}{N-K} - \frac{1}{N-K}\mathrm{Tr}(\mathbf{\Pi}_{K-1}\mathbf{B}_N).
$$

Proof

$$e_{2,N} = \frac{\mathbf{x}_K^H \mathbf{\Pi}_{K-1}^H \mathbf{B}_N \mathbf{\Pi}_{K-1} \mathbf{x}_K}{N-K} - \frac{1}{N-K} \mathrm{Tr}(\mathbf{\Pi}_{\mathbf{K-1}} \mathbf{B}_N)$$

$$\begin{aligned}
\frac{\mathbf{x}_K^H \mathbf{\Pi}_{K-1}^H \mathbf{B}_N \mathbf{\Pi}_{K-1} \mathbf{x}_K}{N-K} &\simeq \frac{1}{N-K} \mathrm{Trace}(\mathbf{\Pi}_{K-1} \mathbf{B}_N \mathbf{\Pi}_{K-1}) \\
&= \frac{1}{N-K} \mathrm{Trace}(\mathbf{\Pi}_{K-1}^2 \mathbf{B}_N) \\
&= \frac{1}{N-K} \mathrm{Trace}(\mathbf{\Pi}_{K-1} \mathbf{B}_N)
\end{aligned}$$

Proof

$$e_{1,N} = \frac{\mathbf{x}_K^H \mathbf{\Pi}_{K-1}^H \mathbf{B}_N \mathbf{\Pi}_{K-1} \mathbf{x}_K}{\|\mathbf{x}_K \mathbf{\Pi}_{K-1}\|^2} - \frac{\mathbf{x}_K^H \mathbf{\Pi}_{K-1}^H \mathbf{B}_N \mathbf{\Pi}_{K-1} \mathbf{x}_K}{N-K}$$

We rewrite the form:

$$e_{1,N} = \frac{\mathbf{x}_K^H \mathbf{\Pi}_{K-1}^H \mathbf{B}_N \mathbf{\Pi}_{K-1} \mathbf{x}_K}{N-K} \left(\frac{N-K}{\|\mathbf{x}_K \mathbf{\Pi}_{K-1}\|^2} - 1 \right)$$

Let us first show that $\frac{\mathbf{x}_K^H \mathbf{\Pi}_{K-1}^H \mathbf{B}_N \mathbf{\Pi}_{K-1} \mathbf{x}_K}{N-K}$ is bounded.

Proof

As $e_{2,N}$ converges to 0 almost everywhere,

$$\frac{\mathbf{x}_K^H \mathbf{\Pi}_{K-1}^H \mathbf{B}_N \mathbf{\Pi}_{K-1} \mathbf{x}_K}{N-K} < 2\frac{\mathrm{Trace}(\mathbf{\Pi}_{K-1}\mathbf{B}_N)}{N-K} \leq 2\|\mathbf{\Pi}_{K-1}\mathbf{B}_N\|$$

for N large enough (for any matrix $\mathbf{X}$, $\mathrm{Trace}(\mathbf{X}) \leq \|\mathbf{X}\|\mathrm{Rank}(\mathbf{X})$ and the rank of $\mathbf{\Pi}_{K-1}\mathbf{B}_N$ does not exceed $N-K$ **in our case**).

As $\sup_{N\in\mathbb{N}}\|\mathbf{\Pi}_{K-1}^H \mathbf{B}_N\| < \infty$, $\frac{\mathbf{x}_K^H \mathbf{\Pi}_{K-1}^H \mathbf{B}_N \mathbf{\Pi}_{K-1} \mathbf{x}_K}{N-K}$ is bounded almost everywhere.

Proof

$$e_{1,N} = \frac{\mathbf{x}_K^H \mathbf{\Pi}_{K-1}^H \mathbf{B}_N \mathbf{\Pi}_{K-1} x_K}{N-K} \left(\frac{N-K}{\|\mathbf{\Pi}_{K-1}\mathbf{x}_K\|^2} - 1 \right)$$

Note that $\mathbf{\Pi}_{K-1}$ is independent of $\mathbf{x}_K$.

$\|\mathbf{\Pi}_{K-1}\mathbf{x}_K\|^2$ is χ^2 distributed with $2(N-K)$ degrees of freedom. Its probability density is the function: $\frac{t^{N-K-1}}{(N-k-1)!} e^{-t}$. A direct computation show that:

$$\mathbb{E} \mid \frac{N-K}{\|\mathbf{x}_K \mathbf{\Pi}_{K-1}\|^2} - 1 \mid^4 = O((N-K)^{-2})$$

which coincides with $O((N)^{-2})$ if $\frac{K}{N} \rightarrow \alpha$.

Results on Random Unitary Matrices

F. Hiai and D. Petz, "The semicircle law, free random variables and entropy," American Mathematical Society, Providence, 2000

D. Voiculescu, "Limit laws for random matrices and free products", Invent. Math, 104 (1991), 201-220.

D. Shlyakhtenko, "R-transform of certain joint distributions in Free Probability Theory," D. V. Voiculescu (ed.), Fields Inst. Commun. 12, Amer. Math. Soc., 1997, pp. 253-256.

P. Diacoms and S.N. Evans, "Linear functionals of eigenvalues of random matrices," Transactions AMS 353, 2615-2633.

K. Wreand, "Eigenvalues distributions of random unitary matrices", Probability theory and related fields, 123, 202-224, 2002.

P. Biane, "Free probability for probabilists", http://www.dma.ens.fr/ biane, 2000.

Sum of Random Matrices

Let $\mathbf{X}$ and $\mathbf{T}$ be $N \times N$ Hermitian matrices with an empirical eigenvalue distribution converging respectively almost surely towards a compact supported probability distribution $P_{\mathbf{X}}$ and $P_{\mathbf{T}}$. Let $\mathbf{\Theta}$ be an $N \times N$ haar distributed matrix independent from $\mathbf{X}$ and $\mathbf{T}$. Then, almost surely, the empirical eigenvalue distribution of

$$\mathbf{H} = \mathbf{\Theta}\mathbf{T}\mathbf{\Theta}^H + \mathbf{X}$$

converges almost surely as $K, N \rightarrow \infty$ but $\frac{K}{N}$ fixed to a unique non-random distribution such as:

$$R_{\mathbf{H}}(z) = R_{\mathbf{\Theta}\mathbf{T}\mathbf{\Theta}^H}(z) + R_{\mathbf{X}}(z)$$

where

$$R_{\mathbf{X}}(z) = K_{\mathbf{X}}(-z) - \frac{1}{z}$$

with

$$G_{\mathbf{X}}(K_{\mathbf{X}}(z)) = K_{\mathbf{X}}(G_{\mathbf{X}}(z)) = z$$

One might see in the R-transform the non-commutative analog of the logarithm of the Fourier transform of a probability density function.

Application Example

Let $\mathbf{X}$ and $\mathbf{T}$ be two orthogonal projects on the subspace $\mathbb{C}^N$ of dimension $\frac{N}{2}$ i.e the matrices have $\frac{N}{2}$ eigenvalues equal to 1 and the rest is equal to 0. As a consequence,

$$G_{\mathbf{X}}(z) = \frac{1}{2}\frac{2z-1}{z(z-1)}$$

$$K_{\mathbf{X}}(z) = \frac{z-1-\sqrt{z^2+1}}{2z}$$

and

$$R_{\mathbf{H}}(z) = \frac{z-1+\sqrt{z^2+1}}{z}$$

which yields

$$G_{\mathbf{H}}(z) = -\frac{1}{\sqrt{z(z-2)}}$$

Arc-sinus law:

$$P_{\mathbf{H}}(x) = \frac{1}{\pi\sqrt{x(x-2)}}1_{[0,2]}(x)$$

Where does the result come from: Free Probability Theory

Classical Probability Theory: If we are given probability densities $f_{\mathbf{X}}$ and $f_{\mathbf{Y}}$ for random variables $\mathbf{X}$ and $\mathbf{Y}$ respectively, and if we know that $\mathbf{X}$ and $\mathbf{Y}$ are independent, we can compute the moments of $\mathbf{X}+\mathbf{Y}$ and $\mathbf{XY}$ for example from the moments of $\mathbf{X}$ and $\mathbf{Y}$.

Where does the result come from: Free Probability Theory

Our Case: Given two random matrices $\mathbf{A}_N$ and $\mathbf{B}_N$ with limiting spectral measures $f_{\mathbf{A}}$ and $f_{\mathbf{B}}$, we would like to compute the limiting spectral measure for $\mathbf{A}_N$+$\mathbf{B}_N$ and $\mathbf{A}_N\mathbf{B}_N$ in terms of the moments of $f_{\mathbf{A}}$ and $f_{\mathbf{B}}$.

Difference with classical probability: $\mathbf{A}_N$ and $\mathbf{B}_N$ 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

D. Voiculescu, Addition of certain noncommuting random variables, J. Funct Anal., 66 (1986), pp. 323-346

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 $\mathbf{A}_N$ or $\mathbf{B}_N$ 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, $\mathbf{A}_N$ and $\mathbf{\Theta}_N \mathbf{B}_N \mathbf{\Theta}_N^H$, where $\mathbf{A}_N$ and $\mathbf{B}_N$ are independent and $\mathbf{\Theta}_N$ is a $N \times N$ Haar Unitary matrix, are free.

Where does the result come from: Free Probability Theory

More generally, it can be expected that the eigenvalue distribution of $f(\mathbf{A}_N, \mathbf{B}_N)$ depends only on the asymptotic eigenvalue distribution of $\mathbf{A}_N$ and $\mathbf{B}_N$ if:

- $\mathbf{A}_N$ and $\mathbf{B}_N$ 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 $\mathcal{U}(n)$ 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.

Product of Random Matrices

Let $\mathbf{X}$ and $\mathbf{T}$ be $N \times N$ Hermitian matrices with an empirical eigenvalue distribution converging respectively almost surely towards a compact supported probability distribution $P_{\mathbf{X}}$ and $P_{\mathbf{T}}$. Let $\mathbf{\Theta}$ be an $N \times N$ haar distributed matrix independent from $\mathbf{X}$ and $\mathbf{T}$. Then, almost surely, the empirical eigenvalue distribution of $\mathbf{H} = \mathbf{\Theta T \Theta}^H \mathbf{X}$ converges almost surely as $K, N \to \infty$ but $\frac{K}{N}$ fixed to a unique non-random distribution such as:

$$S_{\mathbf{H}}(z) = S_{\mathbf{\Theta T \Theta}^H}(z) S_{\mathbf{X}}(z)$$

where

$$S_{\mathbf{X}}(z) = \Xi_{\mathbf{X}}(z) \frac{1+z}{z}$$

with

$$\Xi_{\mathbf{X}}(\Phi_{\mathbf{X}}(z)) = \Phi_{\mathbf{X}}(\Xi_{\mathbf{X}}(z)) = z$$

$$\Phi_{\mathbf{X}}(z) = \int \frac{zt}{1-zt} P_{\mathbf{X}}(t)\,dt = -1 - \frac{1}{z} G_{\mathbf{X}}(\frac{1}{z})$$

The S-transform can be seen as the non-commutative analog of the Mellin transform in classical probability theory.

Free Probability Theory and Random Matrices

Convolution for random matrices

- In general, we cannot find the eigenvalues of sums/products of independent random matrices from the eigenvalues of individual matrices.
- The exception is when the matrices have the same eigenvectors, as for diagonal matrices.
- If this is not the case, it is hard to combine the eigenvectors of A and B to find the eigenvectors of AB.

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 μ_A and μ_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 $A_n B_n$.

Examples

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 *Marčhenko Pastur law* with parameter c, also called μ_c.

Motivation for free probability

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

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

exist. If we linearly extend the linear functional μ 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,

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

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, ϕ), where A is a unital $*$-*algebra* with unit I, and ϕ is a normalized (i.e. $\phi(I) = 1$) linear functional on A.
- For matrices, ϕ 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 of freeness

Definition. A family of unital $*$-subalgebras $(A_i)_{i \in I}$ is 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.

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 μ_A. $\mu_A \boxplus \mu_B$ is defined as the probability measure μ_{A+B} when A and B are random variables which are free. Similarly for $\mu_A \boxtimes \mu_B$.
- Given μ and μ_2, when there exists a unique μ_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 ϕ. One can understand freeness as a rule for computing the expectation of products of a and b. Show that:

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

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

$$\phi(ab^2a) = \phi(a^2)\phi(b^2)$$

Examples

Examples of asymptotically free random matrices are

1. $(A_n, U_n B_n U_n^*)$, where U_n is a standard unitary random matrix independent from A_n, B_n,
2. Families of standard unitary random matrices,
3. Families of Gaussian random matrices,
4. Deterministic matrices which have a limit distribution, combined with the previous mentioned families.

Similarities between classical and free probability

1. Freeness is the analogy of independence.
2. The R-transform: Transform on probability distributions which satisfies
$$R_{\mu_a \boxplus \mu_b}(z) = R_{\mu_a}(z) + R_{\mu_b}(z).$$
Analogy to the logarithm of the Fourier transform for classical random variables.
3. The S-transform: Transform on probability distributions which satisfies
$$S_{\mu_a \boxtimes \mu_b}(z) = S_{\mu_a}(z) S_{\mu_b}(z).$$
4. Free central limit theorem: If $a_1, a_2, \ldots$ are free, and $\phi(a_i) = 0$, $\phi(a_i^2) = 1$, $\sup_i |\phi(a_i^k)| < \infty$ for all k. Then the sequence $(a_1 + a_2 + \cdots + a_n)/\sqrt{n}$ converges in distribution to a probability measure with density $\frac{1}{2\pi}\sqrt{4 - x^2}$. This is called the semicircular law, and is the analogy to the normal law of classical probability.

Similarities between classical and free probability

1. Poisson distributions in classical probability have their analogy in the free Poisson distributions. These are given by the Marčenko Pastur laws μ_c with parameter c. Their density is
$$f^{\mu_c}(x) = (1 - \frac{1}{c})^+ \delta(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi c x}, \tag{3}$$
where $(z)^+ = \max(0, z)$, $a = (1 - \sqrt{c})^+$ and $a = (1 + \sqrt{c})^+$.
2. Both classical and free convolution can be expressed in terms of constructs from combinatorics.

Free Deconvolution for signal processing applications

Introduction

- 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.

Definition

Given probability measures μ and μ_2. When there is a unique probability measure μ_1 such that

$$\mu = \mu_1 \boxplus \mu_2,\ \mu = \mu_1 \boxtimes \mu_2 \text{ respectively,}$$

we will write

$$\mu_1 = \mu \ominus \mu_2,\ \mu_1 = \mu \oslash \mu_2 \text{ respectively.}$$

We say that μ_1 is the additive free deconvolution (respectively multiplicative free deconvolution) of μ with μ_2.

Information plus noise model

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

- $\mathbf{R}_n$ and $\mathbf{X}_n$ are assumed independent random matrices of dimension $n \times N$.
- $\mathbf{X}_n$ 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 $\mathbf{r}_n + \sigma\mathbf{x}_n$. $\mathbf{r}_n$ 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). $\mathbf{x}_n$ represents additive noise, with σ a measure of the strength of the noise.

Important Result

Ø. Ryan and M. Debbah, "Free deconvolution for signal processing applications", submitted to IEEE Transactions on Information Theory.

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

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

Estimation of power and the number of users

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

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 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.

Estimation of power and the number of users

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.

$$\boldsymbol{\Theta} = \mathbb{E}\left(\mathbf{y}_i \mathbf{y}_i^H\right) = \mathbf{W}\mathbf{P}\mathbf{W}^H + \sigma^2 \mathbf{I} \tag{4}$$

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:

$$\hat{\boldsymbol{\Theta}} = \frac{1}{L} \sum_{i=1}^{L} \mathbf{y}_i \mathbf{y}_i^H \tag{5}$$

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:

$$\hat{\boldsymbol{\Theta}} - \sigma^2 \mathbf{I}$$

Estimation of power and the number of users

The sample covariance matrix is related to the true covariance matrix $\mathbf{\Theta} = \mathbb{E}\left(\mathbf{y}_i\mathbf{y}_i^H\right)$ by:

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

with

$$\mathbf{\Theta} = \mathbf{W}\mathbf{P}\mathbf{W}^H + \sigma^2\mathbf{I}$$

and $\mathbf{X}$ is a $n \times L$ i.i.d Gaussian zero mean matrix. Combining (6), (6), 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 equation

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

Simulations

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

We use a 36×36 ($N = 36$) 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

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

Simulations

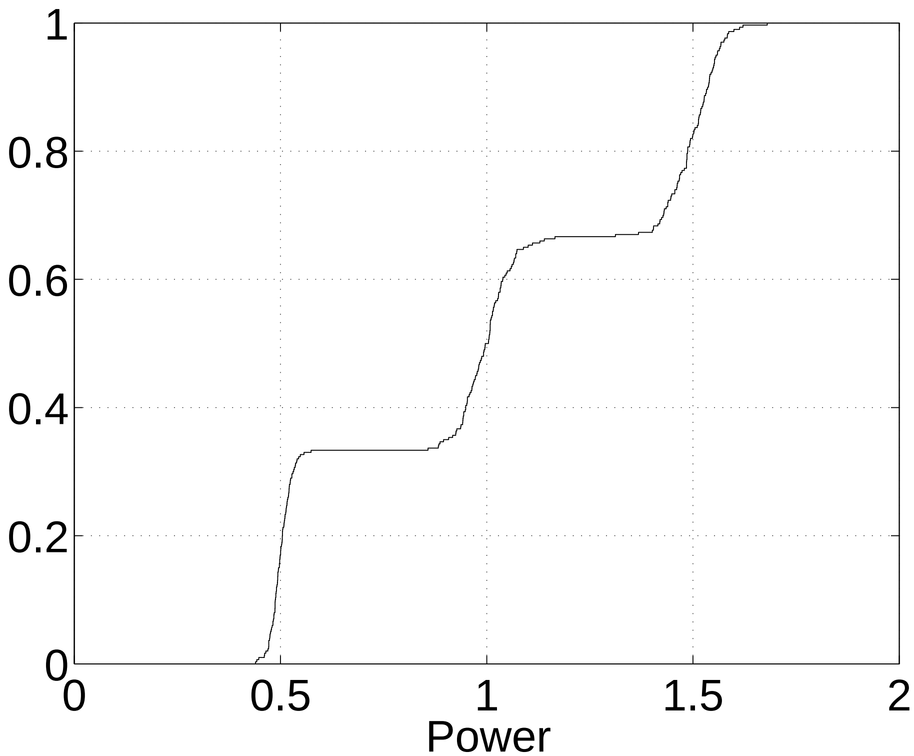

Figure 1: Distribution of the powers estimated from multiplicative free deconvolution from sample covariance matrices with $L = 2048$.

Simulations

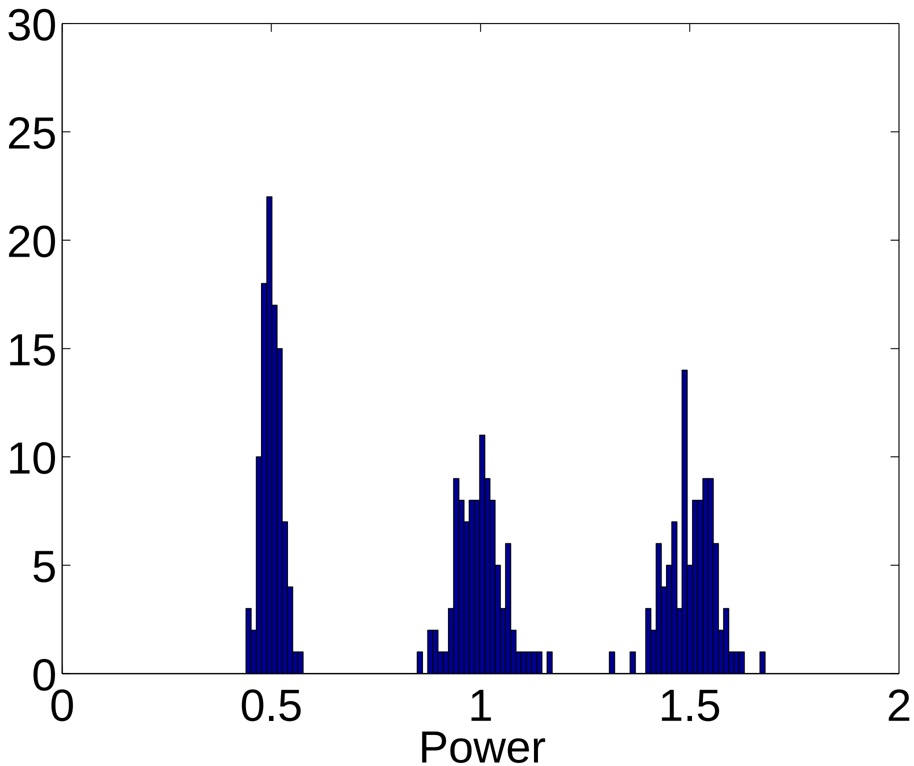

Figure 2: Distribution of the powers estimated from multiplicative free deconvolution from sample covariance matrices with $L = 2048$.

Simulations

We use a 36×36 ($N = 36$) diagonal matrix as our power matrix $\mathbf{P}$ with $\mu_{\mathbf{P}} = \delta_1$. We would like to determine the number of users.

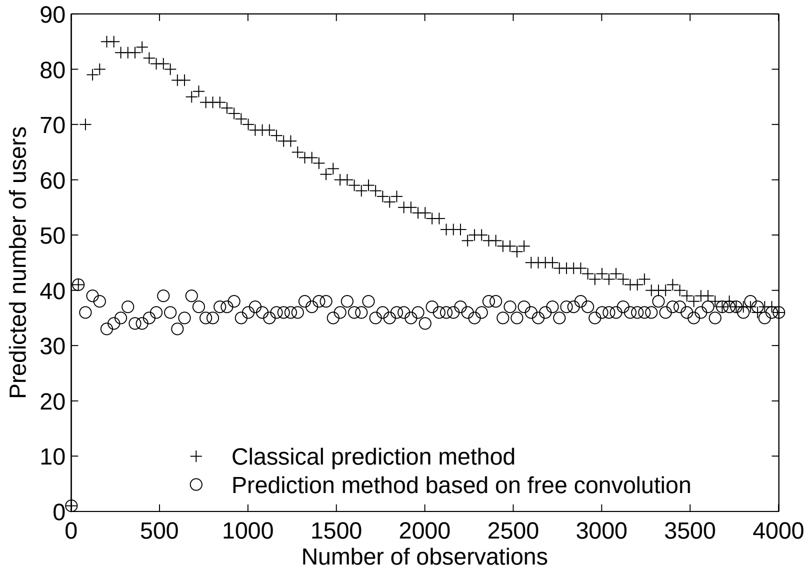

Figure 3: Estimation of the number of users with a classical method, and free convolution $L = 1024$ observations have been used.

Last Slide

THANK YOU!