Random Matrices in Wireless Flexible Networks

Romain Couillet
Alcatel-Lucent Chair on Flexible Radio, Supélec, Gif sur Yvette, FRANCE
romain.couillet@supelec.fr

Johannes Kepler Universitä t, Linz, Austria.

Outline

Large dimensional data

Let $\mathbf{w}_1, \mathbf{w}_2 \ldots \in \mathbb{C}^N$ be independently drawn from an N-variate process of mean zero and covariance $\mathbf{R} = \mathrm{E}[\mathbf{w}_1 \mathbf{w}_1^{\mathsf{H}}] \in \mathbb{C}^{N \times N}$.

Law of large numbers

As $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i^{\mathsf{H}} = \mathbf{W}\mathbf{W}^{\mathsf{H}} \xrightarrow{\text{a.s.}} \mathbf{R}$$

In reality, one cannot afford $n \to \infty$.

- if $n \gg N$,

$$\mathbf{R}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i^{\mathsf{H}}$$

  is a “good” estimate of $\mathbf{R}$.
- if $N/n = O(1)$, and if both (n, N) are large, we can still say, for all (i, j),

$$(\mathbf{R}_n)_{ij} \xrightarrow{\text{a.s.}} (\mathbf{R})_{ij}$$

  What about the global behaviour? What about the eigenvalue distribution?

Empirical and limit spectra of Wishart matrices

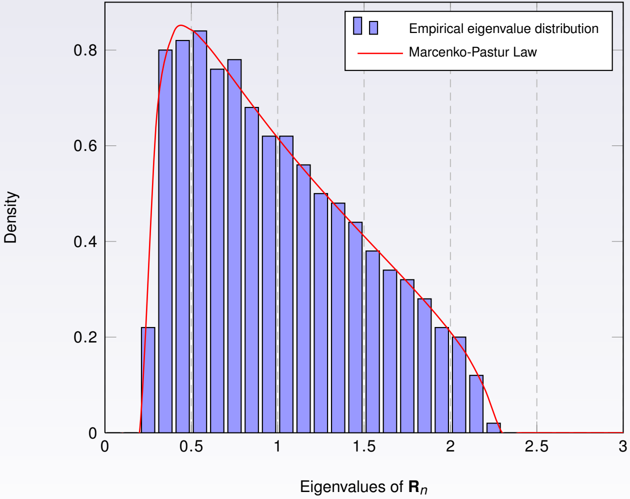


Figure: Histogram of the eigenvalues of $\mathbf{R}_n$ for $n = 2000$, $N = 500$, $\mathbf{R} = \mathbf{I}_N$

The Marcenko-Pastur Law

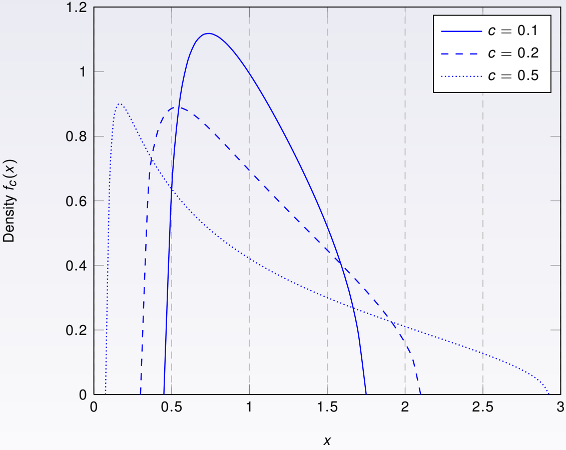
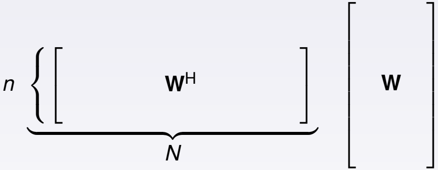


Figure: Marcenko-Pastur law for different limit ratios $c = \lim N/n$.

The Marcenko-Pastur law

Let $\mathbf{W} \in \mathbb{C}^{N \times n}$ have i.i.d. elements, of zero mean and variance $1/n$.
Eigenvalues of the matrix



when $N, n \to \infty$ with $N/n \to c$ **IS NOT IDENTITY!**

Remark: If the entries are Gaussian, the matrix is called a Wishart matrix with n degrees of freedom. The **exact** distribution is known in the finite case.

The birth of large dimensional random matrix theory

E. Wigner, “Characteristic vectors of bordered matrices with infinite dimensions,” The annals of mathematics, vol. 62, pp. 546-564, 1955.

$$\mathbf{X}_N = \frac{1}{\sqrt{N}} \begin{bmatrix} 0 & +1 & +1 & +1 & -1 & -1 & \cdots \\ +1 & 0 & -1 & +1 & +1 & +1 & \cdots \\ +1 & -1 & 0 & +1 & +1 & +1 & \cdots \\ +1 & +1 & +1 & 0 & +1 & +1 & \cdots \\ -1 & +1 & +1 & +1 & 0 & -1 & \cdots \\ -1 & +1 & +1 & +1 & -1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

As the matrix dimension increases, what can we say about the eigenvalues (energy levels)?

Semi-circle law, Full circle law...

- If $\mathbf{X}_N \in \mathbb{C}^{N\times N}$ is Hermitian with i.i.d. entries of mean 0, variance $1/N$ above the diagonal, then $F^{\mathbf{X}_N} \xrightarrow{\text{a.s.}} F$ where F has density f the semi-circle law

$$f(x) = \frac{1}{2\pi}\sqrt{(4-x^2)^+}$$

- Shown from the method of moments

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{tr} \mathbf{X}_N^{2k} = \frac{1}{k+1} C_k^{2k}$$

which are exactly the moments of $f(x)$!

- If $\mathbf{X}_N \in \mathbb{C}^{N\times N}$ has i.i.d. 0 mean, variance $1/N$ entries, then asymptotically its complex eigenvalues distribute uniformly on the complex unit circle.

Semi-circle law

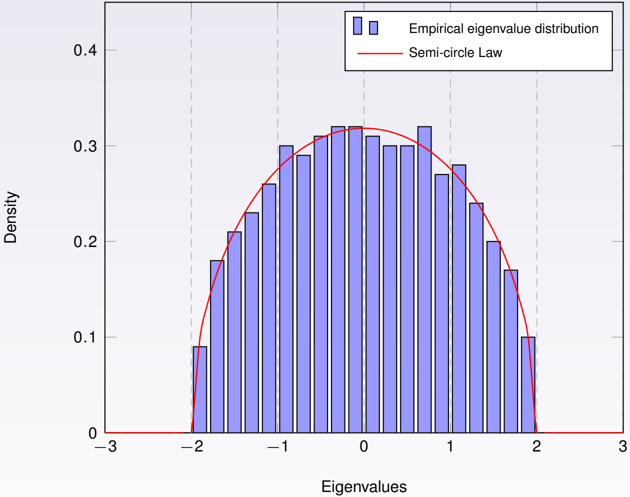


Figure: Histogram of the eigenvalues of Wigner matrices and the semi-circle law, for $N = 500$

Circular law

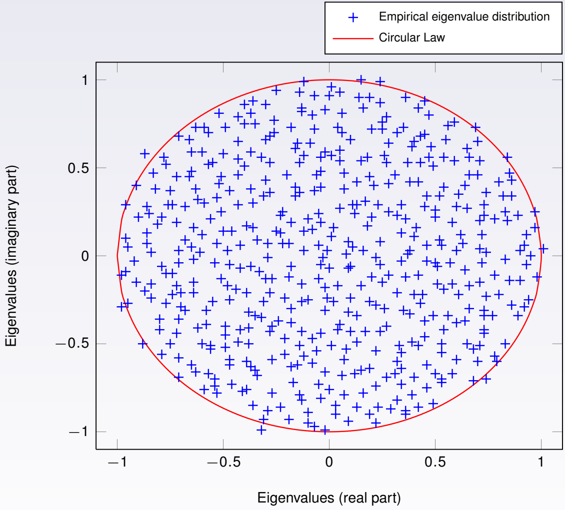


Figure: Eigenvalues of $\mathbf{X}_N$ with i.i.d. standard Gaussian entries, for $N = 500$.

More involved matrix models

- much study has surrounded the Marcenko-Pastur law, the Wigner semi-circle law etc.
- for practical purposes, we often need more general matrix models
  - products and sums of random matrices
  - i.i.d. models with correlation/variance profile
  - distribution of inverses etc.
- for these models, it is often impossible to have a closed-form expression of the limiting distribution.
- sometimes we do not have a limiting convergence.

To study these models, the method of moments is not enough!
A consistent powerful mathematical framework is required.

Eigenvalue distribution and moments

- The Hermitian matrix $\mathbf{R}_N \in \mathbb{C}^{N\times N}$ has successive *empirical* moments M_k^N, $k = 1, 2, \ldots$,

$$M_k^N = \frac{1}{N}\sum_{i=1}^{N} \lambda_i^k$$

- In classical probability theory, for A, B independent,

$$c_k(A+B) = c_k(A) + c_k(B)$$

with $c_k(X)$ the cumulants of X. The cumulants c_k are connected to the moments m_k by,

$$m_k = \sum_{\pi\in\mathcal{P}(k)} \prod_{V\in\pi} c_{|V|}$$

A natural extension of classical probability for non-commutative random variables exist, called

Free Probability

Free probability

- To connect the moments of $\mathbf{A}+\mathbf{B}$ to those of $\mathbf{A}$ and $\mathbf{B}$, independence is not enough. $\mathbf{A}$ and $\mathbf{B}$ must be asymptotically free,
  - two Gaussian matrices are free
  - a Gaussian matrix and any deterministic matrix are free
  - unitary (Haar distributed) matrices are free
  - a Haar matrix and a Gaussian matrix are free etc.
- Similarly as in classical probability, we define free cumulants C_k,

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

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

- Combinatorial description by non-crossing partitions,

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

Non-crossing partitions

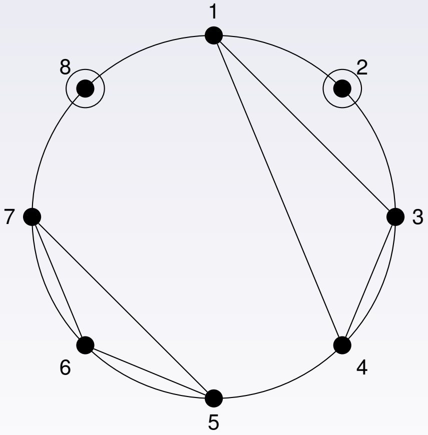


Figure: Non-crossing partition $\pi = \{\{1, 3, 4\}, \{2\}, \{5, 6, 7\}, \{8\}\}$ of $NC(8)$.

Moments of sums and products of random matrices

- Combinatorial calculus of all moments

Theorem

For free random matrices **A** *and* **B**, *we have the relationship,*

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

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

in conjunction with free moment-cumulant formula, gives all moments of sum and product.

Theorem

If F is a compactly supported distribution function, then F is determined by its moments.

- In the absence of support compactness, it is impossible to retrieve the distribution function from moments. This is in particular the case of Vandermonde matrices.

Free convolution

- In classical probability theory, for independent A, B,

$$\mu_{A+B}(x) = \mu_A(x) * \mu_B(x) \stackrel{\Delta}{=} \int \mu_A(t)\mu_B(x-t)dt$$

- In free probability, for free **A**, **B**, we use the notations

$$\mu_{\mathbf{A}+\mathbf{B}} = \mu_{\mathbf{A}} \boxplus \mu_{\mathbf{B}}, \ \mu_{\mathbf{A}} = \mu_{\mathbf{A}+\mathbf{B}} \boxminus \mu_{\mathbf{B}}, \ \mu_{\mathbf{AB}} = \mu_{\mathbf{A}} \boxtimes \mu_{\mathbf{B}}, \ \mu_{\mathbf{A}} = \mu_{\mathbf{A}+\mathbf{B}} \boxslash \mu_{\mathbf{B}}$$

Ø. Ryan, M. Debbah, "Multiplicative free convolution and information-plus-noise type matrices," Arxiv preprint math.PR/0702342, 2007.

Theorem

Convolution of the information-plus-noise model Let $\mathbf{W}_N \in \mathbb{C}^{N\times n}$ *have i.i.d. Gaussian entries of mean* 0 *and variance* 1*,* $\mathbf{A}_N \in \mathbb{C}^{N\times n}$*, such that* $\mu_{\frac{1}{n}\mathbf{A}_N\mathbf{A}_N^{\mathrm{H}}} \Rightarrow \mu_A$*, as* $n/N \to c$*. Then the eigenvalue distribution of*

$$\mathbf{B}_N = \frac{1}{n}\left(\mathbf{A}_N + \sigma\mathbf{W}_N\right)\left(\mathbf{A}_N + \sigma\mathbf{W}_N\right)^{\mathrm{H}}$$

converges weakly and almost surely to μ_B *such that*

$$\mu_B = \left(\left(\mu_A \boxslash \mu_c\right) \boxplus \delta_{\sigma^2}\right) \boxtimes \mu_c$$

with μ_c *the Marcenko-Pastur law with ratio c.*

Similarities between classical and free probability

	Classical Probability	Free probability
Moments	$m_k = \int x^k dF(x)$	$M_k = \int x^k dF(x)$
Cumulants	$m_n = \sum_{\pi \in \mathcal{P}(n)} \prod_{V \in \pi} c_{\|V\|}$	$M_n = \sum_{\pi \in \mathcal{NC}(n)} \prod_{V \in \pi} C_{\|V\|}$
Independence	classical independence	freeness
Additive convolution	$f_{A+B} = f_A * f_B$	$\mu_{\mathbf{A}+\mathbf{B}} = \mu_{\mathbf{A}} \boxplus \mu_{\mathbf{B}}$
Multiplicative convolution	f_{AB}	$\mu_{\mathbf{AB}} = \mu_{\mathbf{A}} \boxtimes \mu_{\mathbf{B}}$
Sum Rule	$c_k(A+B) = c_k(A) + c_k(B)$	$C_k(\mathbf{A}+\mathbf{B}) = C_k(\mathbf{A}) + C_k(\mathbf{B})$
Central Limit	$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \to \mathcal{N}(0,1)$	$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \Rightarrow$ semi-circle law

Bibliography on Free Probability related work

- D. Voiculescu, “Addition of certain non-commuting random variables,” Journal of functional analysis, vol. 66, no. 3, pp. 323-346, 1986.
- R. Speicher, “Combinatorial theory of the free product with amalgamation and operator-valued free probability theory,” Mem. A.M.S., vol. 627, 1998.
- R. Seroul, D. O’Shea, “Programming for Mathematicians,” Springer, 2000.
- H. Bercovici, V. Pata, “The law of large numbers for free identically distributed random variables,” The Annals of Probability, pp. 453-465, 1996.
- A. Nica, R. Speicher, “On the multiplication of free N-tuples of noncommutative random variables,” American Journal of Mathematics, pp. 799-837, 1996.
- Ø. Ryan, M. Debbah, “Multiplicative free convolution and information-plus-noise type matrices,” Arxiv preprint math.PR/0702342, 2007.
- N. R. Rao, A. Edelman, “The polynomial method for random matrices,” Foundations of Computational Mathematics, vol. 8, no. 6, pp. 649-702, 2008.
- Ø. Ryan, M. Debbah, “Asymptotic Behavior of Random Vandermonde Matrices With Entries on the Unit Circle,” IEEE Trans. on Information Theory, vol. 55, no. 7, pp. 3115-3147, 2009.

The Stieltjes transform

Definition

Let F be a real distribution function. The Stieltjes transform m_F of F is the function defined, for $z \in \mathbb{C} \setminus \mathbb{R}$, as

$$m_F(z) = \int \frac{1}{\lambda - z} dF(\lambda)$$

For $a < b$ real, denoting $z = x + iy$, we have the inverse formula

$$F'(x) = \lim_{y \to 0} \frac{1}{\pi} \Im[m_F(x + iy)]$$

Knowing the Stieltjes transform is knowing the eigenvalue distribution!

Remark on the Stieltjes transform

- If F is the eigenvalue distribution of a Hermitian matrix $\mathbf{X}_N \in \mathbb{C}^{N\times N}$, we might denote $m_{\mathbf{X}} \triangleq m_F$, and

$$m_{\mathbf{X}}(z) = \int \frac{1}{\lambda - z} dF(\lambda) = \frac{1}{N} \operatorname{tr} (\mathbf{X}_N - z\mathbf{I}_N)^{-1}$$

- For compactly supported eigenvalue distribution,

$$m_F(z) = -\frac{1}{z} \int \frac{1}{1 - \frac{\lambda}{z}} = -\sum_{k=0}^{\infty} M_k^N z^{-k-1}$$

The Stieltjes transform is doubly more powerful than the moment approach!

  - conveys more information than any K-finite sequence $M_1, \ldots, M_K$.
  - is not handicapped by the support compactness constraint.

- however, Stieltjes transform methods, while stronger, are more painful to work with.

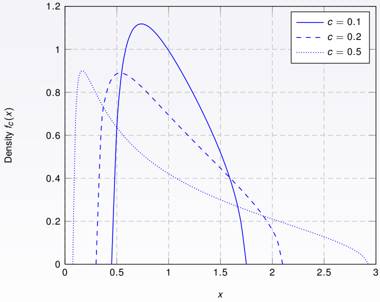
The Marcenko-Pastur law

Theorem

Let $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ have i.i.d. zero mean variance $1/n$ entries with finite eighth order moments. As $n, N \to \infty$ with $\frac{N}{n} \to c \in (0,\infty)$, the e.s.d. of $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$ converges almost surely to a nonrandom distribution function F_c with density f_c given by

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

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



Diagonal entries of the resolvent

Since we want an expression of m_F, we start by identifying the diagonal entries of the resolvent $(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N)^{-1}$ of $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$. Denote

$$\mathbf{X}_N = \begin{bmatrix} \mathbf{y}^{\mathsf{H}} \\ \mathbf{Y} \end{bmatrix}$$

Now, for $z \in \mathbb{C}^+$, we have

$$\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1} = \begin{bmatrix} \mathbf{y}^{\mathsf{H}}\mathbf{y} - z & \mathbf{y}^{\mathsf{H}}\mathbf{Y}^{\mathsf{H}} \\ \mathbf{Y}\mathbf{y} & \mathbf{Y}\mathbf{Y}^{\mathsf{H}} - z\mathbf{I}_{N-1} \end{bmatrix}^{-1}$$

Consider the first diagonal element of $(\mathbf{R}_N - z\mathbf{I}_N)^{-1}$. From the matrix inversion lemma,

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{pmatrix}$$

which here gives

$$\left[\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\right]_{11} = \frac{1}{-z - z\mathbf{y}^{\mathsf{H}}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}\mathbf{y}}$$

Trace Lemma

Z. Bai, J. Silverstein, “Spectral Analysis of Large Dimensional Random Matrices”, Springer Series in Statistics, 2009.

To go further, we need the following result,

Theorem

Let $\{\mathbf{A}_N\} \in \mathbb{C}^{N\times N}$*. Let* $\{\mathbf{x}_N\} \in \mathbb{C}^N$*, be a random vector of i.i.d. entries with zero mean, variance* $1/N$ *and finite* 8^{th} *order moment, independent of* $\mathbf{A}_N$*. Then*

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

For large N, we therefore have approximately

$$\left[\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\right]_{11} \simeq \frac{1}{-z - z\frac{1}{N}\operatorname{tr}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}}$$

Rank-1 perturbation lemma

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.

It is somewhat intuitive that adding *a single column* to $\mathbf{Y}$ won't affect the trace in the limit.

Theorem

Let $\mathbf{A}$ *and* $\mathbf{B}$ *be* $N \times N$ *with* $\mathbf{B}$ *Hermitian positive definite, and* $\mathbf{v} \in \mathbb{C}^N$. *For* $z \in \mathbb{C} \setminus \mathbb{R}^-$,

$$\left| \frac{1}{N} \operatorname{tr} \left((\mathbf{B} - z\mathbf{I}_N)^{-1} - (\mathbf{B} + \mathbf{v}\mathbf{v}^{\mathsf{H}} - z\mathbf{I}_N)^{-1} \right) \mathbf{A} \right| \leq \frac{1}{N} \frac{\|\mathbf{A}\|}{\operatorname{dist}(z, \mathbb{R}^+)}$$

with $\|\mathbf{A}\|$ *the spectral norm of* $\mathbf{A}$, *and* $\operatorname{dist}(z, A) = \inf_{y \in A} \|y - z\|$.

Therefore, for large N, we have approximately,

$$\begin{aligned} \left[\left(\mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N \right)^{-1} \right]_{11} &\simeq \frac{1}{-z - z\frac{1}{N} \operatorname{tr}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}} \\ &\simeq \frac{1}{-z - z\frac{1}{N} \operatorname{tr}(\mathbf{X}_N^{\mathsf{H}}\mathbf{X}_N - z\mathbf{I}_n)^{-1}} \\ &= \frac{1}{-z - z\frac{n}{N} m_{\underline{F}}(z)} \end{aligned}$$

in which we recognize the Stieltjes transform $m_{\underline{F}}$ of the l.s.d. of $\mathbf{X}_N^{\mathsf{H}}\mathbf{X}_N$.

End of the proof

We have again the relation

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

hence

$$\left[\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\right]_{11} \simeq \frac{1}{\frac{n}{N} - 1 - z - zm_F(z)}$$

Note that the choice $(1,1)$ is irrelevant here, so the expression is valid for all pair (i,i). Summing over the N terms and averaging, we finally have

$$m_F(z) = \frac{1}{N}\operatorname{tr}\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1} \simeq \frac{1}{c - 1 - z - zm_F(z)}$$

which solve a polynomial of second order. Finally

$$m_F(z) = \frac{c-1}{2z} - \frac{1}{2} + \frac{\sqrt{(c-1-z)^2 - 4z}}{2z}.$$

From the inverse Stieltjes transform formula, we then verify that m_F is the Stieltjes transform of the Marcenko-Pastur law.

Asymptotic results using the Stieltjes transform

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.

Theorem

Let $\underline{\mathbf{B}}_N = \mathbf{X}_N \mathbf{T}_N \mathbf{X}_N^{\mathrm{H}} \in \mathbb{C}^{N\times N}$, $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ *has i.i.d. entries of mean* 0 *and variance* $1/N$, $F^{\mathbf{T}_N} \Rightarrow F^T$, $n/N \to c$. *Then,* $F^{\underline{\mathbf{B}}_N} \Rightarrow \underline{F}$ *almost surely,* $\underline{F}$ *having Stieltjes transform*

$$m_{\underline{F}}(z) = \left(c \int \frac{t}{1 + t m_{\underline{F}}(z)} dF^T(t) - z \right)^{-1} = \left[\frac{1}{N} \operatorname{tr} \mathbf{T}_N \left(m_{\underline{F}}(z) \mathbf{T}_N + \mathbf{I}_N \right)^{-1} - z \right]^{-1}$$

which has a unique solution $m_{\underline{F}}(z) \in \mathbb{C}^+$ *if* $z \in \mathbb{C}^+$, *and* $m_{\underline{F}}(z) > 0$ *if* $z < 0$.

- in general, no explicit expression for $\underline{F}$.
- Stieltjes transform of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^{\mathrm{H}} \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$ with asymptotic distribution F,

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

Spectrum of the sample covariance matrix model $\mathbf{B}_N = \sum_{i=1}^n \mathbf{x}_i' \mathbf{x}_i'^{\mathrm{H}}$, with $\mathbf{X}_N'^{\mathrm{H}} = [\mathbf{x}_1', \ldots, \mathbf{x}_n']$, $\mathbf{x}_i'$ i.i.d. with zero mean and covariance $\mathbf{T}_N = \mathrm{E}[\mathbf{x}_1' \mathbf{x}_1'^{\mathrm{H}}]$.

Getting F' from m_F

- Remember that, for $a < b$ real,

$$f(x) \stackrel{\Delta}{=} F'(x) = \lim_{y \to 0} \frac{1}{\pi} \Im[m_F(x + iy)]$$

- to plot the density $f(x)$, span $z = x + iy$ on the line $\{x \in \mathbb{R}, y = \varepsilon\}$ parallel but close to the real axis, solve $m_F(z)$ for each z, and plot $\Im[m_F(z)]$.

Example (Sample covariance matrix)

For N multiple of 3, let $dF^T(x) = \frac{1}{3}\delta(x-1) + \frac{1}{3}\delta(x-3) + \frac{1}{3}\delta(x-K)$ and let $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^{\mathsf{H}} \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$ with $F^{\mathbf{B}_N} \to F$, then

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

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

We take $c = 1/10$ and alternatively $K = 7$ and $K = 4$.

Spectrum of the sample covariance matrix

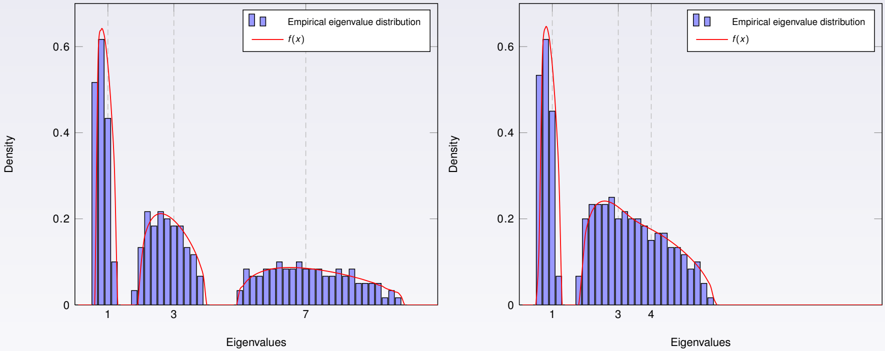


Figure: Histogram of the eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^{\mathsf{H}} \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$, $N = 3000$, $n = 300$, with $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in (i) 1, 3 and 7 on top, (ii) 1, 3 and 4 at bottom.

The Shannon Transform

A. M. Tulino, S. Verdù, “Random matrix theory and wireless communications,” Now Publishers Inc., 2004.

Definition

Let F be a probability distribution, m_F its Stieltjes transform, then the Shannon-transform $\mathcal{V}_F$ of F is defined as

$$\mathcal{V}_F(x) \triangleq \int_0^\infty \log(1 + x\lambda) dF(\lambda) = \int_x^\infty \left(\frac{1}{t} - m_F(-t) \right) dt$$

If F is the distribution function of the eigenvalues of $\mathbf{X}\mathbf{X}^{\mathsf{H}} \in \mathbb{C}^{N \times N}$,

$$\mathcal{V}_F(x) = \frac{1}{N} \log\det \left(\mathbf{I}_N + x\mathbf{X}\mathbf{X}^{\mathsf{H}} \right).$$

Note that this last relation is fundamental to wireless communication purposes!

Models studied with analytic tools

- Models involving i.i.d. matrices
  - sample covariance matrix models, $\mathbf{XTX}^{\mathrm{H}}$ and $\mathbf{T}^{\frac{1}{2}}\mathbf{X}^{\mathrm{H}}\mathbf{XT}^{\frac{1}{2}}$
  - doubly correlated models, $\mathbf{R}^{\frac{1}{2}}\mathbf{XTX}^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}}$. With $\mathbf{X}$ Gaussian, Kronecker model.
  - doubly correlated models with external matrix, $\mathbf{R}^{\frac{1}{2}}\mathbf{XTX}^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}} + \mathbf{A}$.
  - variance profile, $\mathbf{XX}^{\mathrm{H}}$, where $\mathbf{X}$ has i.i.d. entries with mean 0, variance $\sigma_{i,j}^2$.
  - Ricean channels, $\mathbf{XX}^{\mathrm{H}} + \mathbf{A}$, where $\mathbf{X}$ has a variance profile.
  - sum of doubly correlated i.i.d. matrices, $\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k^{\mathrm{H}}\mathbf{R}_k^{\frac{1}{2}}$.
  - information-plus-noise models $(\mathbf{X}+\mathbf{A})(\mathbf{X}+\mathbf{A})^{\mathrm{H}}$
  - frequency-selective doubly-correlated channels $(\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k\mathbf{R}_k^{\frac{1}{2}})(\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k\mathbf{R}_k^{\frac{1}{2}})$
  - sum of frequency-selective doubly-correlated channels $\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{H}_k\mathbf{T}_k\mathbf{H}_k^{\mathrm{H}}\mathbf{R}_k^{\frac{1}{2}}$, where $\mathbf{H}_k = \sum_{l=1}^{L} \mathbf{R}'^{\frac{1}{2}}_{kl}\mathbf{X}_{kl}\mathbf{T}'_{kl}\mathbf{X}_{kl}^{\mathrm{H}}\mathbf{R}'^{\frac{1}{2}}_{kl}$.
- Models involving a column subset $\mathbf{W}$ of unitary matrices
  - sum of doubly correlated Haar matrices $\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{W}_k\mathbf{T}_k\mathbf{W}_k^{\mathrm{H}}\mathbf{R}_k^{\frac{1}{2}}$
  - sum of products of correlated i.i.d. and/or Haar matrices $\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{W}_k^{\mathrm{H}}\mathbf{S}_k\mathbf{W}_k^{\mathrm{H}}\mathbf{T}_k\mathbf{X}_k\mathbf{R}_k^{\frac{1}{2}}$

In most cases, **T** and **R** can be taken random, but independent of **X**. More involved random matrices, such as Vandermonde matrices, were not yet studied.

Models studied with moments/free probability

- asymptotic results
  - most of the above models with Gaussian $\mathbf{X}$.
  - products $\mathbf{V}_1\mathbf{V}_1^{\mathrm{H}}\mathbf{T}_1\mathbf{V}_2\mathbf{V}_2^{\mathrm{H}}\mathbf{T}_2\ldots$ of Vandermonde and deterministic matrices
  - *conjecture:* any probability space of matrices invariant to row or column permutations.
- marginal studies, not yet fully explored
  - rectangular free convolution: singular values of rectangular matrices
  - finite size models. Instead of almost sure convergence of $m_{\mathbf{X}_N}$ as $N \to \infty$, we can study finite size behaviour of $\mathrm{E}[m_{\mathbf{X}_N}]$.

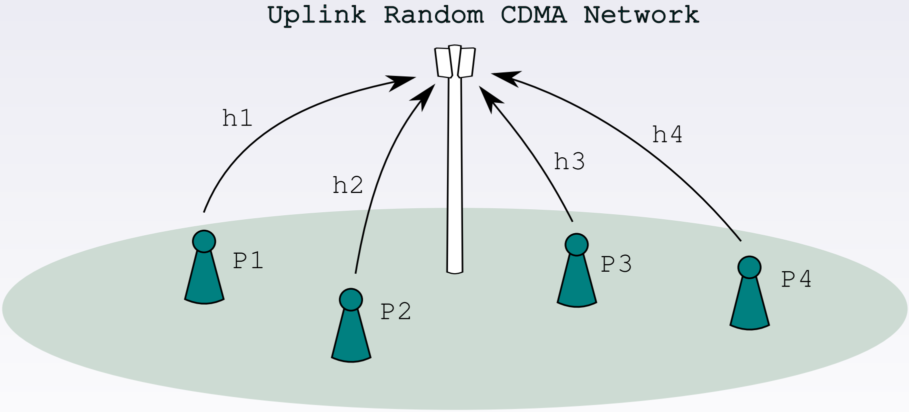
Related bibliography

- R. B. Dozier, J. W. Silverstein, “On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices,” Journal of Multivariate Analysis, vol. 98, no. 4, pp. 678-694, 2007.
- 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.
- J. W. Silverstein, S. Choi “Analysis of the limiting spectral distribution of large dimensional random matrices” Journal of Multivariate Analysis, vol. 54, no. 2, pp. 295-309, 1995.
- F. Benaych-Georges, “Rectangular random matrices, related free entropy and free Fisher’s information,” Arxiv preprint math/0512081, 2005.
- Ø. Ryan, M. Debbah, “Multiplicative free convolution and information-plus-noise type matrices,” Arxiv preprint math.PR/0702342, 2007.
- V. L. Girko, “Theory of Random Determinants,” Kluwer, Dordrecht, 1990.
- R. Couillet, M. Debbah, J. W. Silverstein, “A deterministic equivalent for the capacity analysis of correlated multi-user MIMO channels,” *submitted to IEEE Trans. on Information Theory*.
- V. L. Girko, “Theory of Random Determinants,” Kluwer, Dordrecht, 1990.
- W. Hachem, Ph. Loubaton, J. Najim, “Deterministic Equivalents for Certain Functionals of Large Random Matrices”, Annals of Applied Probability, vol. 17, no. 3, 2007.
- M. J. M. Peacock, I. B. Collings, M. L. Honig, “Eigenvalue distributions of sums and products of large random matrices via incremental matrix expansions,” IEEE Trans. on Information Theory, vol. 54, no. 5, pp. 2123, 2008.
- D. Petz, J. Réffy, “On Asymptotics of large Haar distributed unitary matrices,” Periodica Math. Hungar., vol. 49, pp. 103-117, 2004.
- Ø. Ryan, A. Masucci, S. Yang, M. Debbah, “Finite dimensional statistical inference,” *submitted to IEEE Trans. on Information Theory, Dec. 2009.*

Technical Bibliography

- W. Rudin, “Real and complex analysis,” New York, 1966.
- P. Billingsley, “Probability and measure,” Wiley New York, 2008.
- P. Billingsley, “Convergence of probability measures,” Wiley New York, 1968.

Uplink random CDMA



Capacity of uplink random CDMA

- System model conditions,
  - uplink random CDMA
  - K mobile users, 1 base station
  - N chips per CDMA spreading code.
  - User k, $k \in \{1, \ldots, K\}$ has code $\mathbf{w}_k \sim \mathcal{CN}(0, \mathbf{I}_N)$
  - User k transmits the symbol s_k.
  - User k's channel is $h_k\sqrt{P_k}$, with P_k the power of user k
- The base station receives

$$\mathbf{y} = \sum_{k=1}^{K} h_k \mathbf{w}_k \sqrt{P_k} s_k + \mathbf{n}$$

- This can be written in the more compact form

$$\mathbf{y} = \mathbf{W}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{s} + \mathbf{n}$$

with

  - $\mathbf{s} = [s_1, \ldots, s_K]^{\mathsf{T}} \in \mathbb{C}^K$,
  - $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_K] \in \mathbb{C}^{N \times K}$,
  - $\mathbf{P} = \mathrm{diag}(P_1, \ldots, P_K) \in \mathbb{C}^{K \times K}$,
  - $\mathbf{H} = \mathrm{diag}(h_1, \ldots, h_K) \in \mathbb{C}^{K \times K}$.

MMSE decoder

- Consists into taking

$$r_k = \mathbf{w}_k^{\mathrm{H}} \left(\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} + \sigma^2 \mathbf{I}_N \right)^{-1} \mathbf{y}$$

as symbol for user k.

- The SINR for user's k signal is

$$\gamma_k^{(\mathrm{MMSE})} = P_k|h_k|^2 \mathbf{w}_k^{\mathrm{H}} \Big(\sum_{\substack{1 \leq i \leq K \\ i \neq k}} P_i |h_i|^2 \mathbf{w}_i \mathbf{w}_i^{\mathrm{H}} + \sigma^2 \mathbf{I}_N \Big)^{-1} \mathbf{w}_k \tag{1}$$

$$= P_k|h_k|^2 \mathbf{w}_k^{\mathrm{H}} (\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k|h_k|^2 \mathbf{w}_k \mathbf{w}_k^{\mathrm{H}} + \sigma^2 \mathbf{I}_N)^{-1} \mathbf{w}_k. \tag{2}$$

- Now we have the following result

Theorem (Trace Lemma)

If $\mathbf{x} \in \mathbb{C}^N$ is i.i.d. with entries of zero mean, variance $1/N$, and $\mathbf{A} \in \mathbb{C}^{N \times N}$ is independent of $\mathbf{x}$, then

$$\mathbf{x}^{\mathrm{H}}\mathbf{A}\mathbf{x} = \sum_{i,j} x_i^* x_j A_{ij} \xrightarrow{\text{a.s.}} \frac{1}{N} \operatorname{tr} \mathbf{A}.$$

- Applying this result, for N large,

$$\mathbf{w}_k^{\mathrm{H}}(\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k|h_k|^2\mathbf{w}_k\mathbf{w}_k^{\mathrm{H}} + \sigma^2\mathbf{I}_N)^{-1}\mathbf{w}_k - \frac{1}{N}\operatorname{tr}(\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k|h_k|^2\mathbf{w}_k\mathbf{w}_k^{\mathrm{H}} + \sigma^2\mathbf{I}_N)^{-1} \xrightarrow{\text{a.s.}} 0.$$

MMSE decoder

$$\mathbf{w}_k^{\mathsf{H}}(\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}} - P_k|h_k|^2\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}} + \sigma^2\mathbf{I}_N)^{-1}\mathbf{w}_k - \frac{1}{N}\operatorname{tr}(\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}} - P_k|h_k|^2\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}} + \sigma^2\mathbf{I}_N)^{-1} \xrightarrow{\text{a.s.}} 0.$$

- Second important result,

Theorem (Rank 1 perturbation Lemma)

Let $\mathbf{A} \in \mathbb{C}^{N\times N}$, $\mathbf{x} \in \mathbb{C}^N$, $t > 0$, *then*

$$\left| \frac{1}{N}\operatorname{tr}(\mathbf{A} + t\mathbf{I}_N)^{-1} - \frac{1}{N}\operatorname{tr}(\mathbf{A} + \mathbf{x}\mathbf{x}^{\mathsf{H}} + t\mathbf{I}_N)^{-1} \right| \leq \frac{1}{tN}$$

- As N grows large,

$$\frac{1}{N}\operatorname{tr}\left(\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}} - P_k|h_k|^2\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}} + \sigma^2\mathbf{I}_N\right)^{-1} - \frac{1}{N}\operatorname{tr}\left(\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}} + \sigma^2\mathbf{I}_N\right)^{-1} \to 0,$$

- The RHS is the Stieltjes transform of $\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}}$ in $z = -\sigma^2$!

$$m_{\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}}}(-\sigma^2)$$

MMSE decoder

- From previous result,

$$m_{\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}}}(-\sigma^2) - m_N(-\sigma^2) \xrightarrow{\text{a.s.}} 0$$

with $m_N(-\sigma^2)$ the unique positive solution of

$$m = \left[\frac{1}{N}\operatorname{tr}\mathbf{HPH}^{\mathrm{H}}\left(m\mathbf{HPH}^{\mathrm{H}} + \mathbf{I}_K\right)^{-1} + \sigma^2\right]^{-1}$$

independent of k!

- This is also

$$m = \left[\sigma^2 + \frac{1}{N}\sum_{1\leq i\leq K}\frac{P_i|h_i|^2}{1+mP_i|h_i|^2}\right]^{-1}$$

- Finally,

$$\gamma_k^{(\mathrm{MMSE})} - P_k|h_k|^2 m_N(-\sigma^2) \xrightarrow{\text{a.s.}} 0$$

and the capacity reads

$$C^{(\mathrm{MMSE})}(\sigma^2) - \frac{1}{K}\sum_{k=1}^{K}\log_2(1+P_k|h_k|^2 m_N(-\sigma^2)) \xrightarrow{\text{a.s.}} 0.$$

MMSE decoder

$$C^{(\mathrm{MMSE})}(\sigma^2) - \frac{1}{K}\sum_{k=1}^{K}\log_2(1 + P_k|h_k|^2 m_N(-\sigma^2)) \xrightarrow{\mathrm{a.s.}} 0.$$

- AWGN channel, $P_k = P$, $h_k = 1$,

$$C^{(\mathrm{MMSE})}(\sigma^2) \xrightarrow{\mathrm{a.s.}} c\log_2\left(1 + \frac{-(\sigma^2 + (c-1)P) + \sqrt{(\sigma^2 + (c-1)P)^2 + 4P\sigma^2}}{2\sigma^2}\right)$$

- Rayleigh channel, $P_k = P$, $|h_k|$ Rayleigh,

$$m = \left[\sigma^2 + c\int \frac{Pt}{1 + Ptm} e^{-t} dt\right]^{-1}$$

and

$$C_{\mathrm{MMSE}}(\sigma^2) \xrightarrow{\mathrm{a.s.}} c\int \log_2\left(1 + Ptm(-\sigma^2)\right) e^{-t} dt.$$

Matched-Filter, Optimal decoder ...

R. Couillet, M. Debbah, J. W. Silverstein, “A Deterministic Equivalent for the Capacity Analysis of Correlated Multi-User MIMO Channels,” IEEE Trans. on Information Theory, *accepted*, on arXiv.

- Similarly, we can compute deterministic equivalents for the matched-filter performance,

$$C_{\rm MF}(\sigma^2) - \frac{1}{N}\sum_{k=1}^{K}\log_2\left(1+\frac{P_k|h_k|^2}{\frac{1}{N}\sum_{i=1}^{K}P_i|h_i|^2+\sigma^2}\right)\xrightarrow{\rm a.s.} 0$$

- AWGN case,

$$C_{\rm MF}(\sigma^2)\xrightarrow{\rm a.s.} c\log_2\left(1+\frac{P}{Pc+\sigma^2}\right)$$

- Rayleigh case,

$$C_{\rm MF}(\sigma^2)\xrightarrow{\rm a.s.} -c\log_2(e)e^{\frac{Pc+\sigma^2}{P}}{\rm Ei}\left(-\frac{Pc+\sigma^2}{P}\right)$$

- ... and the optimal joint-decoder performance

$$C_{\rm opt}(\sigma^2)-\log_2\left(1+\frac{1}{\sigma^2N}\sum_{k=1}^{K}\frac{P_k|h_k|^2}{1+cP_k|h_k|^2m_N(-\sigma^2)}\right)-\frac{1}{N}\sum_{k=1}^{K}\log_2\left(1+cP_k|h_k|^2m_N(-\sigma^2)\right)$$
$$-\log_2(e)\left(\sigma^2m_N(-\sigma^2)-1\right)\xrightarrow{\rm a.s.} 0.$$

  with $m_N(-\sigma^2)$ defined as previously.
- Similar expressions are obtained for the AWGN and Rayleigh cases.

Simulation results: AWGN case

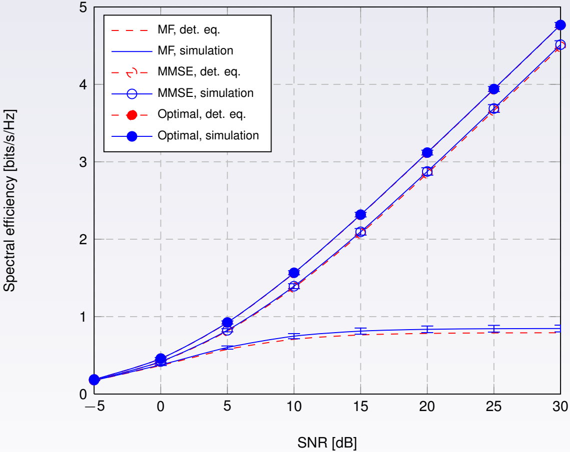


Figure: Spectral efficiency of random CDMA decoders, AWGN channels. Comparison between simulations and deterministic equivalents (det. eq.), for the matched-filter, the MMSE decoder and the optimal decoder, $K = 16$ users, $N = 32$ chips per code. Rayleigh channels. Error bars indicate two standard deviations.

Simulation results: Rayleigh case

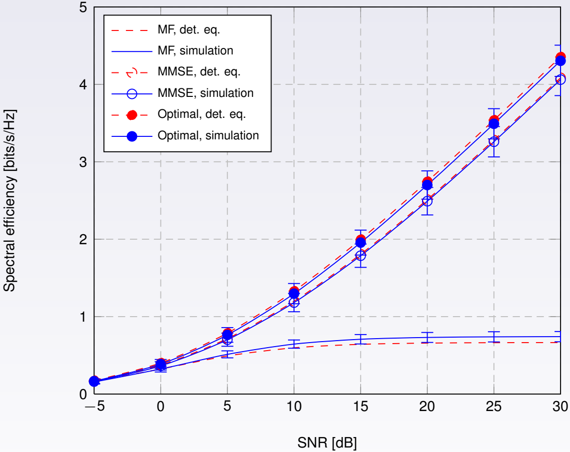


Figure: Spectral efficiency of random CDMA decoders, Rayleigh fading channels. Comparison between simulations and deterministic equivalents (det. eq.), for the matched-filter, the MMSE decoder and the optimal decoder, $K = 16$ users, $N = 32$ chips per code. Rayleigh channels. Error bars indicate two standard deviations.

Simulation results: Performance as a function of K/N

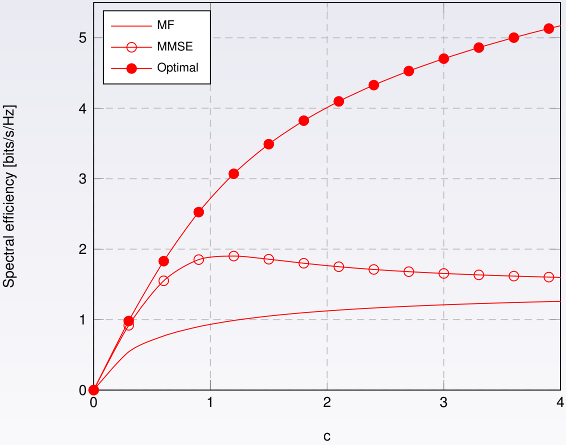


Figure: Spectral efficiency of random CDMA decoders, for different asymptotic ratios $c = K/N$, SNR=10 dB, AWGN channels. Deterministic equivalents for the matched-filter, the MMSE decoder and the optimal decoder. Rayleigh channels.

MIMO-MAC, SINR of the MMSE receiver

R. Couillet, M. Debbah, J. W. Silverstein, “A deterministic equivalent for the capacity analysis of correlated multi-user MIMO channels,” *submitted to IEEE Trans. on Information Theory*.

$$\mathbf{B}_N = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{H}_k^{\mathsf{H}}, \text{ with } \mathbf{H}_k = \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k^{\frac{1}{2}}$$

with $\mathbf{X}_k \in \mathbb{C}^{N \times n_k}$ with i.i.d. entries of zero mean, variance $1/n_k$, $\mathbf{R}_k$ Hermitian nonnegative definite, $\mathbf{T}_k$ diagonal. Denote $c_k = N/n_k$. Then, as all N and n_k grow large, with ratio c_k,

$$\frac{1}{N} \operatorname{tr}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \frac{1}{N} \operatorname{tr} \left(-z \left[\mathbf{I}_N + \sum_{k=1}^{K} \bar{e}_k(z) \mathbf{R}_k \right] \right)^{-1} \xrightarrow{\text{a.s.}} 0$$

where the set of functions $\{e_i(z)\}$ form the unique solution to the K equations

$$e_i(z) = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(-z \left[\mathbf{I}_N + \sum_{k=1}^{K} \bar{e}_k(z) \mathbf{R}_k \right] \right)^{-1}, \quad \bar{e}_i(z) = \frac{1}{n_i} \operatorname{tr} \mathbf{T}_i \left(-z \left[\mathbf{I}_{n_i} + c_i e_i(z) \mathbf{T}_i \right] \right)^{-1}$$

Hence, the SINR at the output of the MMSE receiver for user stream i of user k, γ_{ik}, satisfies

$$\gamma_{ik} = \mathbf{h}_{k,i}^{\mathsf{H}} \left(\mathbf{B}_N - \mathbf{h}_{k,i} \mathbf{h}_{k,i}^{\mathsf{H}} \right)^{-1} - t_{k,i} e_i(-\sigma^2) \xrightarrow{\text{a.s.}} 0.$$

Deterministic equivalent approach: guess work

We will use here the "guess-work" method to find the deterministic equivalent. Consider the simpler case $K = 1$.
Back to the original notations, we seek a matrix $\mathbf{D}$ such that

$$\frac{1}{N}\operatorname{tr}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \frac{1}{N}\operatorname{tr}\mathbf{D}^{-1} \xrightarrow{\text{a.s.}} 0$$

as $N \to \infty$.

Resolvent lemma

For invertible $\mathbf{A}$, $\mathbf{B}$ matrices,

$$\mathbf{A}^{-1} - \mathbf{B}^{-1} = -\mathbf{A}^{-1}(\mathbf{A} - \mathbf{B})\mathbf{B}^{-1}$$

Taking the matrix differences,

$$\mathbf{D}^{-1} - (\mathbf{B}_N - z\mathbf{I}_N)^{-1} = \mathbf{D}^{-1}(\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}} - z\mathbf{I}_N - \mathbf{D})(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$$

It seems convenient to take $\mathbf{D} = -z\mathbf{I}_N + \bar{e}_{\mathbf{B}_N}\mathbf{R}$ with $\bar{e}_{\mathbf{B}_N}$ left to be defined (the notation $\mathbf{B}_N$ in $\bar{e}_{\mathbf{B}_N}$ reminds that we do not look *yet* for a deterministic quantity).

Deterministic equivalent approach: guess work (2)

"Silverstein's" lemma

Let $\mathbf{A}$ be Hermitian invertible, then for any vector $\mathbf{x}$ and scalar τ such that $\mathbf{A} + \tau \mathbf{x}\mathbf{x}^{\mathsf{H}}$ is invertible

$$\mathbf{x}^{\mathsf{H}}(\mathbf{A} + \tau \mathbf{x}\mathbf{x}^{\mathsf{H}})^{-1} = \frac{\mathbf{x}^{\mathsf{H}}\mathbf{A}^{-1}}{1 + \tau \mathbf{x}\mathbf{A}^{-1}\mathbf{x}^{\mathsf{H}}}$$

With $\mathbf{D} = -z\mathbf{I}_N + \bar{e}_{\mathbf{B}_N}\mathbf{R}$,

$$\begin{aligned}
\mathbf{D}^{-1} - (\mathbf{B}_N - z\mathbf{I}_N)^{-1} &= \mathbf{D}^{-1}(\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}} - z\mathbf{I}_N - \mathbf{D})(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \\
&= \mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\left(\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\right)\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \bar{e}_{\mathbf{B}_N}\mathbf{D}^{-1}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \\
&= \mathbf{D}^{-1}\sum_{j=1}^{n}\tau_j\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \bar{e}_{\mathbf{B}_N}\mathbf{D}^{-1}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \\
&= \sum_{j=1}^{n}\tau_j\frac{\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}}{1 + \tau_j\mathbf{x}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j} - \bar{e}_{\mathbf{B}_N}\mathbf{D}^{-1}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}
\end{aligned}$$

Choice of $\bar{e}_{\mathbf{B}_N}$: $\bar{e}_{\mathbf{B}_N} = \frac{1}{n}\sum_{j=1}^{n}\frac{\tau_j}{1+\tau_j c\frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}} = \frac{1}{n}\operatorname{tr}\mathbf{T}\left(\mathbf{I}_n + \mathbf{T}\frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\right)^{-1}$

$$\frac{1}{N}\operatorname{tr}\mathbf{D}^{-1} - \frac{1}{N}\operatorname{tr}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} = \frac{1}{N}\sum_{j=1}^{n}\tau_j\left[\frac{\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1 + \tau_j\mathbf{x}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j} - \frac{\frac{1}{n}\operatorname{tr}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}}{1 + c\tau_j\frac{1}{N}\operatorname{tr}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}}\right]$$

Ergodic Mutual Information of MIMO-MAC

Remember now that

$$\int \log(1+xt)dF(t) = \int_{1/x}^{\infty} \left(\frac{1}{t} - m_F(-t) \right) dt$$

R. Couillet, M. Debbah, J. W. Silverstein, "A deterministic equivalent for the capacity analysis of correlated multi-user MIMO channels," *submitted to IEEE Trans. on Information Theory*.

Theorem

Under the previous model for $\mathbf{B}_N$, as N, n_k grow large,

$$\mathrm{E}\left[\frac{1}{N}\log\det(x\mathbf{B}_N + \mathbf{I}_N)\right] - \left[\frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{k=1}^{K} \bar{e}_k(-1/x)\mathbf{R}_k\right) + \sum_{k=1}^{K}\frac{1}{N}\log\det\left(\mathbf{I}_{n_k} + c_k e_k(-1/x)\mathbf{T}_k\right) \sum_{k=1}^{K}\frac{1}{N}\log\det\left(\mathbf{I}_{n_k} + c_k e_k(-1/x\right. - \frac{1}{x}\sum_{k=1}^{K}\bar{e}_k(-1/x)e_k(-1/x)\right] \xrightarrow{\text{a.s.}} 0$$

Ergodic sum rate capacity of MIMO-MAC

We look for $\mathbf{P}_1^\star, \ldots, \mathbf{P}_K^\star$ that achieve the optimal sum rate.

R. Couillet, M. Debbah, J. W. Silverstein, "A deterministic equivalent for the capacity analysis of correlated multi-user MIMO channels," *submitted to IEEE Trans. on Information Theory*.

The deterministic-equivalent maximizing precoders $\mathbf{P}_1^\circ, \ldots, \mathbf{P}_K^\circ$ satisfy

$$\mathbf{P}_k^\circ = \mathbf{U}_k \operatorname{diag}(p_{k,1}^\circ, \ldots, p_{k,n_k}^\circ)\mathbf{U}_k^{\mathsf{H}}, \quad \text{where } \mathbf{T}_k = \mathbf{U}_k \operatorname{diag}(t_{k,1}, \ldots, t_{k,n_k})\mathbf{U}_k^{\mathsf{H}}$$

and $p_{k,i}$ defined by iterative water-filling as

$$p_{k,i}^\circ = \left(\mu_k - \frac{1}{e_k^\circ t_{k,i}}\right)^+$$

with μ_k such that $\operatorname{tr} \mathbf{P}_k^\circ = P_k$, and e_k° defined as e_k for $(\mathbf{P}_1, \ldots, \mathbf{P}_K) = (\mathbf{P}_1^\circ, \ldots, \mathbf{P}_K^\circ)$.
Moreover, under some conditions,

$$\|\mathbf{P}_k^\star - \mathbf{P}_k^\circ\| \to 0.$$

Variance profile

W. Hachem, Ph. Loubaton, J. Najim, "Deterministic equivalents for certain functionals of large random matrices," Annals of Applied Probability, vol. 17, no. 3, pp. 875-930, 2007.

Theorem

Let $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ *have independent entries with* $(i,j)^{th}$ *entry of zero mean and variance* $\frac{1}{n}\sigma_{ij}^2$*. Let* $\mathbf{A}_N \in \mathbb{R}^{N\times n}$ *be deterministic with uniformly bounded column norm. Then*

$$\frac{1}{N}\operatorname{tr}\left((\mathbf{X}_N+\mathbf{A}_N)(\mathbf{X}_N+\mathbf{A}_N)^{\mathsf{H}}-z\mathbf{I}_N\right)^{-1}-\frac{1}{N}\operatorname{tr}\mathbf{T}_N(z)\xrightarrow{\text{a.s.}}0$$

where $\mathbf{T}_N(z)$ *is the unique function that solves*

$$\mathbf{T}_N(z)=\left(\Psi^{-1}(z)-z\mathbf{A}_N\tilde{\Psi}(z)\mathbf{A}_N^{\mathsf{T}}\right)^{-1},\quad \tilde{\mathbf{T}}_N(z)=\left(\tilde{\Psi}^{-1}(z)-z\mathbf{A}_N^{\mathsf{T}}\Psi(z)\mathbf{A}_N\right)^{-1}$$

with $\Psi(z)=\operatorname{diag}(\psi_i(z))$*,* $\tilde{\Psi}(z)=\operatorname{diag}(\tilde{\psi}_i(z))$*, with entries defined as*

$$\psi_i(z)=-\left(z(1+\frac{1}{n}\operatorname{tr}\tilde{\mathbf{D}}_i\tilde{\mathbf{T}}(z))\right)^{-1},\quad \tilde{\psi}_j(z)=-\left(z(1+\frac{1}{n}\operatorname{tr}\mathbf{D}_j\mathbf{T}(z))\right)^{-1}$$

and $\mathbf{D}_j=\operatorname{diag}(\sigma_{ij}^2, 1\leq i\leq N)$*,* $\tilde{\mathbf{D}}_i=\operatorname{diag}(\sigma_{ij}^2, 1\leq j\leq n)$

Variance profile

W. Hachem, Ph. Loubaton, J. Najim, “Deterministic equivalents for certain functionals of large random matrices,” Annals of Applied Probability, vol. 17, no. 3, pp. 875-930, 2007.

Theorem

For the previous model, we also have that

$$\frac{1}{N}\mathrm{E}\log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}(\mathbf{X}_N + \mathbf{A}_N)(\mathbf{X}_N + \mathbf{A}_N)^{\mathsf{H}}\right)$$

has deterministic equivalent

$$\begin{aligned}&\frac{1}{N}\log\det\left[\frac{1}{\sigma^2}\Psi(-\sigma^2)^{-1} + \mathbf{A}_N\tilde{\Psi}(-\sigma^2)\mathbf{A}_N^{\mathsf{T}}\right]\\ &+\frac{1}{N}\log\det\frac{1}{\sigma^2}\Psi(-\sigma^2)^{-1} - \frac{\sigma^2}{nN}\sum_{i,j}\sigma_{ij}^2\mathbf{T}_{ii}(-\sigma^2)\tilde{\mathbf{T}}_{jj}(-\sigma^2).\end{aligned}$$

Haar random matrices

M. Debbah, W. Hachem, P. Loubaton, M. de Courville, "MMSE analysis of certain large isometric random precoded systems", IEEE Transactions on Information Theory, vol. 49, no. 5, pp. 1293-1311, 2003.

- Recent results were proposed when the matrices $\mathbf{X}_N$ are unitary and unitarily invariant (Haar matrices).
- The central result is the trace lemma

Lemma

Let $\mathbf{W} \in \mathbb{C}^{N\times n}$ *be* $n < N$ *columns of a Haar matrix and* $\mathbf{w}$ *a column of* $\mathbf{W}$. *Let* $\mathbf{B}_N \in \mathbb{C}^{N\times N}$ *a random matrix, function of all columns of* $\mathbf{W}$ *except* $\mathbf{w}$. *Then, assuming that, for growing* N, $c = \sup_n n/N < 1$ *and* $B = \sup_N \|\mathbf{B}_N\| < \infty$, *we have:*

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

Haar random matrices (2)

R. Couillet, J. Hoydis, M. Debbah, "Random beamforming over quasi-static and fading channels: a deterministic equivalent approach", to appear in IEEE Trans. on Inf. Theory.

Theorem

Let $\mathbf{T}_i \in \mathbb{C}^{n_i \times n_i}$ *be nonnegative diagonal and let* $\mathbf{H}_i \in \mathbb{C}^{N \times N_i}$. *Define* $\mathbf{R}_i \triangleq \mathbf{H}_i \mathbf{H}_i^{\mathrm{H}} \in \mathbb{C}^{N \times N}$, $c_i = \frac{n_i}{N_i}$ *and* $\bar{c}_i = \frac{N_i}{N}$. *Denote*

$$\mathbf{B}_N = \sum_{i=1}^{K} \mathbf{H}_i \mathbf{W}_i \mathbf{T}_i \mathbf{W}_i^{\mathrm{H}} \mathbf{H}_i^{\mathrm{H}}.$$

Then, as $N, N_1, \ldots, N_K, n_1, \ldots, n_K \to \infty$ *with ratios bounded* $\bar{c}_i$ *and* $0 \leq c_i \leq 1$ *for all* i*, almost surely*

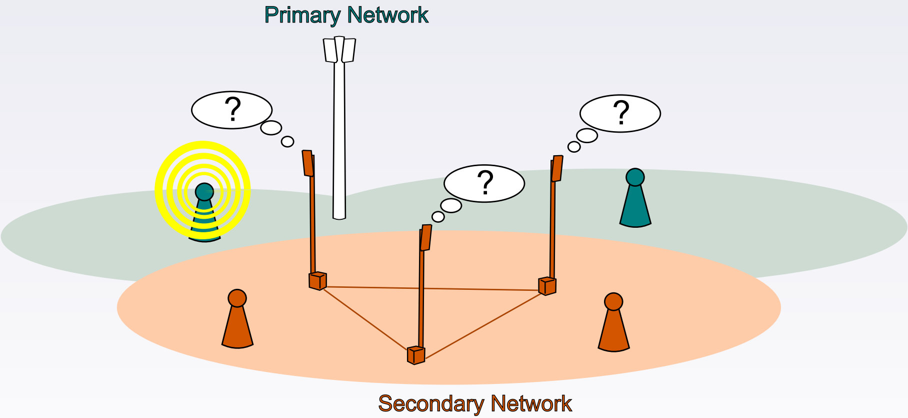
$$F^{\mathbf{B}_N} - F_N \Rightarrow 0, \quad \text{with } m_N(z) = \frac{1}{N} \operatorname{tr} \left(\sum_{i=1}^{K} \bar{e}_i(z) \mathbf{R}_i - z \mathbf{I}_N \right)^{-1}$$

where $(\bar{e}_1(z), \ldots, \bar{e}_K(z))$ *are the solutions (conditionally unique) of*

$$e_i(z) = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{j=1}^{K} \bar{e}_j(z) \mathbf{R}_j - z \mathbf{I}_N \right)^{-1}$$

$$\bar{e}_i(z) = \frac{1}{N} \operatorname{tr} \mathbf{T}_i \left(e_i(z) \mathbf{T}_i + [\bar{c}_i - e_i(z) \bar{e}_i(z)] \mathbf{I}_{n_i} \right)^{-1} \quad \text{(compare to i.i.d. case!)}$$

Signal Sensing in Cognitive Radios



Position of the Problem

Decide on presence of *informative signal* or *pure noise*.

Limited *a priori* Knowledge

- Known parameters: the prior information I
  - N sensors
  - L sampling periods
  - unit transmit power
  - unit channel variance
- Possibly unknown parameters
  - M signal sources
  - noise power equals σ^2

One situation, one solution

For a given prior information I, there must be a unique solution to the detection problem.

Problem statement

Signal detection is a typical hypothesis testing problem.

- $\mathcal{H}_0$: only background noise.

$$\mathbf{Y} = \sigma\boldsymbol{\Theta} = \sigma \begin{pmatrix} \theta_{11} & \cdots & \theta_{1L} \\ \vdots & \ddots & \vdots \\ \theta_{N1} & \cdots & \theta_{NL} \end{pmatrix}$$

- $\mathcal{H}_1$: informative signal plus noise.

$$\mathbf{Y} = \begin{pmatrix} h_{11} & \ldots & h_{1M} & \sigma & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{N1} & \ldots & h_{NM} & 0 & \cdots & \sigma \end{pmatrix} \begin{pmatrix} s_1^{(1)} & \cdots & \cdots & s_1^{(L)} \\ \vdots & \vdots & \vdots & \vdots \\ s_M^{(1)} & \cdots & \cdots & s_M^{(L)} \\ \theta_{11} & \cdots & \cdots & \theta_{1L} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{N1} & \cdots & \cdots & \theta_{NL} \end{pmatrix}$$

Solution

Solution of hypothesis testing is the function

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

If the receiver does not know if $\mathcal{H}_1$ is more likely than $\mathcal{H}_0$,

$$P_{\mathcal{H}_1} = P_{\mathcal{H}_0} = \frac{1}{2}$$

Therefore,

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

Odds for hypothesis $\mathcal{H}_0$

If the SNR is known then the maximum Entropy Principle leads to

$$P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y}) = \frac{1}{(\pi\sigma^2)^{NL}} e^{-\frac{1}{\sigma^2}\operatorname{tr}\mathbf{Y}\mathbf{Y}^H}$$

Odds for hypothesis $\mathcal{H}_1$

If known N, M, SNR only then

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

with

$$\begin{aligned} \boldsymbol{\Sigma} &= L \begin{pmatrix} h_{11} & \dots & h_{1M} & \sigma & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{N1} & \dots & h_{NM} & 0 & \cdots & \sigma \end{pmatrix} \begin{pmatrix} h_{11} & \dots & h_{1M} & \sigma & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{N1} & \dots & h_{NM} & 0 & \cdots & \sigma \end{pmatrix}^H \\ &= \mathbf{U} \left(L\boldsymbol{\Lambda} \right) \mathbf{U}^H \end{aligned}$$

Odds for hypothesis $\mathcal{H}_1$ (2)

Case $M = 1$.
Maximum Entropy distribution for **H** is Gaussian i.i.d channel. *Unordered* eigenvalue distribution for $\mathbf{\Sigma}$,

$$P_{\mathbf{\Lambda}}(\mathbf{\Lambda})d\mathbf{\Lambda} = \mathbf{1}_{(\lambda_1 > \sigma^2)} \frac{1}{N}(\lambda_1 - \sigma^2)^{N-1} \frac{e^{-(\lambda_1 - \sigma^2)}}{(N-1)!} \prod_{i=2}^{N} \delta(\lambda_i - \sigma^2) d\lambda_1 \ldots d\lambda_N$$

Maximum Entropy distribution for $\mathbf{Y}|\mathbf{\Sigma}\mathcal{H}_1$ is correlated Gaussian,

$$P_{\mathbf{Y}|\mathbf{\Sigma} I_1}(\mathbf{Y}, \mathbf{U}, L\mathbf{\Lambda}) = \frac{1}{\pi^{LN} \det(\mathbf{\Lambda})^L} e^{-\operatorname{tr}\left(\mathbf{Y}\mathbf{Y}^H \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^H\right)}$$

Neyman-Pearson Test

- $M = 1$,

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \frac{e^{\sigma^2 - \frac{1}{\sigma^2}\sum_{i=1}^N \lambda_i}}{N\pi^{LN}\sigma^{2(N-1)(L-1)}} \sum_{l=1}^{N} \frac{e^{\frac{\lambda_l}{\sigma^2}}}{\prod_{\substack{i=1\\ i\neq l}}^{N}(\lambda_l - \lambda_i)} J_{N-L-1}(\sigma^2, \lambda_l)$$

with $(\lambda_1, \ldots, \lambda_N) = \mathrm{eig}(\mathbf{Y}\mathbf{Y}^{\mathsf{H}})$ and

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

- From which we have the Neyman-Pearson test

$$C_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \frac{1}{N} \sum_{l=1}^{N} \frac{\sigma^{2(N+L-1)} e^{\sigma^2 + \frac{\lambda_l}{\sigma^2}}}{\prod_{\substack{i=1\\ i\neq l}}^{N}(\lambda_l - \lambda_i)} J_{N-L-1}(\sigma^2, \lambda_l)$$

Neyman-Pearson test only depends on the eigenvalues! But in an involved way!

Neyman-Pearson Test against energy detector, SNR known

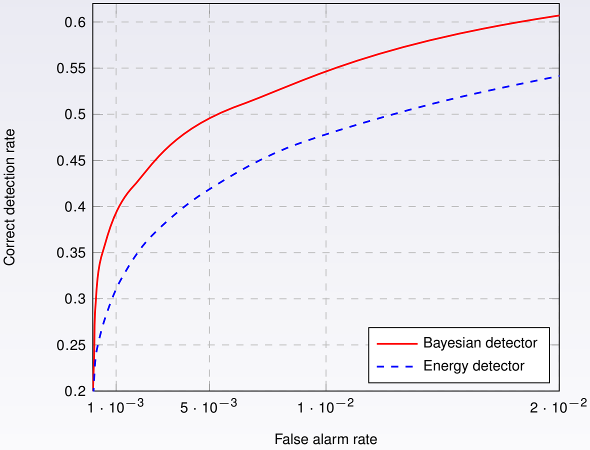


Figure: ROC curve for SIMO transmission, $M = 1$, $N = 4$, $L = 8$, SNR $= -3$ dB, FAR range of practical interest.

Neyman-Pearson Test, Unknown SNR

- We need to integrate out the prior for σ^2.
- This leads to

$$C(\mathbf{Y}) = \frac{\int P_{\mathbf{Y}|\sigma^2, I'_M}(\mathbf{Y}, \sigma^2) P_{\sigma^2}(\sigma^2) d\sigma^2}{\int P_{\mathbf{Y}|\sigma^2, \mathcal{H}_0}(\mathbf{Y}, \sigma^2) P_{\sigma^2}(\sigma^2) d\sigma^2}$$

- prior $P_{\sigma^2}(\sigma^2)$ is chosen to be
  - uniform over $[\sigma_-^2, \sigma_+^2]$
  - Jeffrey over $(0, \infty)$

Reminder: the Marcenko-Pastur Law

If $\mathcal{H}_0$, then the eigenvalues of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ asymptotically distribute as

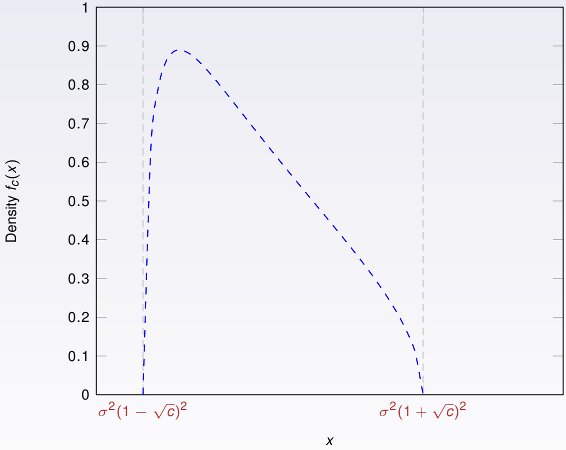


Figure: Marcenko-Pastur law with $c = \lim N/L$.

Alternative Tests in Large Random Matrix Theory

Z. D. Bai, J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices," The Annals of Probability, vol. 26, no.1 pp. 316-345, 1998.

Theorem

$$P(\textit{no eigenvalues outside } [\sigma^2(1-\sqrt{c})^2, \sigma^2(1+\sqrt{c})^2] \textit{ for all large } N) = 1$$

- If $\mathcal{H}_0$,

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

- independent of the SNR!

Conditioning Number Test

L. S. Cardoso, M. Debbah, P. Bianchi, J. Najim, "Cooperative spectrum sensing using random matrix theory," International Symposium on Wireless Pervasive Computing, Santorini, Greece, 2008.

- conditioning number test

$$C_{\mathrm{cond}}(\mathbf{Y}) = \frac{\lambda_{\max}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}})}{\lambda_{\min}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}})}$$

- if $C_{\mathrm{cond}}(\mathbf{Y}) > \tau$, presence of a signal.
- if $C_{\mathrm{cond}}(\mathbf{Y}) < \tau$, absence of signal.
- but this is *ad-hoc*! how good does it compare to optimal?
- can we find non *ad-hoc* approaches?

Alternative Tests in Large Random Matrix Theory (2)

Bianchi, J. Najim, M. Maida, M. Debbah, "Performance of Some Eigen-based Hypothesis Tests for Collaborative Sensing," Proceedings of IEEE Statistical Signal Processing Workshop, 2009.

Generalized Likelihood Ratio Test

- Alternative test to Neyman-Pearson,

$$C_{\mathrm{GLRT}}(\mathbf{Y}) = \frac{\sup_{\mathbf{H},\sigma^2} P_{\mathcal{H}_1|\mathbf{Y},\mathbf{H},\sigma^2}(\mathbf{Y})}{\sup_{\sigma^2} P_{\mathcal{H}_0|\mathbf{Y},\sigma^2}(\mathbf{Y})}$$

- based on ratios of maximum likelihood
- clearly sub-optimal but avoid the need for priors.

- GLRT test

$$C_{\mathrm{GLRT}}(\mathbf{Y}) = \left(\left(1 - \frac{1}{N}\right)^{N-1} \frac{\lambda_{\max}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}})}{\frac{1}{N}\sum_{i=1}^{N}\lambda_i} \left(1 - \frac{\lambda_{\max}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}})}{\sum_{i=1}^{N}\lambda_i}\right)^{N-1} \right)^{-L}.$$

- Contrary to the *ad-hoc* conditioning number test, GLRT based on

$$\frac{\lambda_{\max}}{\frac{1}{N}\operatorname{tr}(\mathbf{Y}\mathbf{Y}^{\mathsf{H}})}$$

Neyman-Pearson Test against Asymptotic Tests

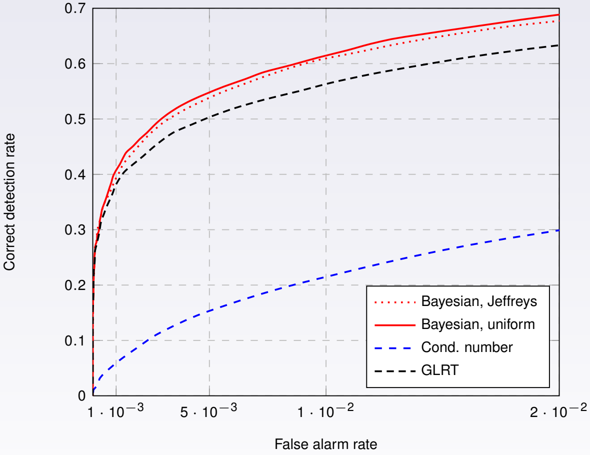


Figure: ROC curve for *a priori* unknown σ^2 of the Bayesian method, conditioning number method and GLRT method, $M = 1$, $N = 4$, $L = 8$, SNR $= 0$ dB. For the Bayesian method, both uniform and Jeffreys prior, with exponent $\alpha = 1$, are provided.

Position of the problem

- it has long been difficult to analytically invert the simplest $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ model to recover the diagonal entries of $\mathbf{T}_N$. Indeed, we only have the deterministic equivalent result

$$\underline{m}_N(z) = \left(-z + c\int \frac{t}{1+t\underline{m}_N(z)}dF^{\mathbf{T}_N}(t)\right)^{-1}$$

  with $\underline{m}_N$ the deterministic equivalent of the Stieltjes transform for $\underline{\mathbf{B}}_N = \mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N$.
- when $\mathbf{T}_N$ has eigenvalues $t_1,\ldots,t_K$ with multiplicity $n_1,\ldots,n_K$, this is

$$\underline{m}_N(z) = \left(-z + \frac{1}{N}\sum_{k=1}^{K} n_k \frac{t_k}{1+t_k\underline{m}_N(z)}\right)^{-1}$$

- an N, n-consistent estimator for the t_k's was never found until recently...
- however, moment-based methods and free probability approaches provide simple solutions to estimate consistently all moments of $F^{\mathbf{T}_N}$.

Free Deconvolution Approach: Reminders

- For free random matrices **A** and **B**, we have the cumulant/moment relationships,

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

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

- this allows one to compute all moments of sum and product distributions

$$\mu_{\mathbf{A}} \boxplus \mu_{\mathbf{B}}, \quad \mu_{\mathbf{A}} \boxtimes \mu_{\mathbf{B}}$$

- in addition, we have results for the information-plus-noise model

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

  whose e.s.d. converges weakly and almost surely to μ_B such that

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

  with μ_c the Mar**u**cenko-Pastur law and $\Gamma_N = \mathbf{R}_N \mathbf{R}_N^{\mathrm{H}}$.
- all basic matrix operations needed in wireless communications are accessible for convenient matrices (Gaussian, Vandermonde etc.)
- all operations are merely polynomial operations on the moments. As a consequence, for $\mathbf{B}_N = f(\mathbf{R}_N)$,

All moments of the l.s.d. of $\mathbf{B}_N$ are obtained as polynomials of those of the l.s.d. of $\mathbf{R}_N$

From free convolution to free deconvolution

Ø. Ryan, M. Debbah, "Multiplicative free convolution and information-plus-noise type matrices," Arxiv preprint math.PR/0702342, 2007.

- The k^{th} moment of the l.s.d. of $\mathbf{B}_N$ is a polynomial of the k-first moments of the l.s.d. of $\mathbf{R}_N$.
  - we can therefore invert the problem and express the k^{th} moment of $\mathbf{R}_N$ as the first k moments of $\mathbf{B}_N$. This entails deconvolution operations,

$$\mu_{\mathbf{A}} = \mu_{\mathbf{A}+\mathbf{B}} \boxminus \mu_{\mathbf{B}}$$

$$\mu_{\mathbf{A}} = \mu_{\mathbf{AB}} \boxslash \mu_{\mathbf{B}}$$

    and for the information-plus-noise model, $\mathbf{B}_N = \frac{1}{n}\left(\mathbf{R}_N + \sigma \mathbf{X}_N\right)\left(\mathbf{R}_N + \sigma \mathbf{X}_N\right)^{\mathsf{H}}$

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

- for more involved models, the polynomial relations can be iterated and even automatically generated.

Example of polynomial relation

- Consider the information-plus-noise model

$$\mathbf{Y} = \mathbf{D} + \mathbf{X}$$

with $\mathbf{Y} \in \mathbb{C}^{N\times n}$, $\mathbf{D} \in \mathbb{C}^{N\times n}$, $\mathbf{X} \in \mathbb{C}^{N\times n}$ with i.i.d. entries of mean 0 and variance 1. Denote

$$M_k = \lim_{n\to\infty} \frac{1}{n} \operatorname{tr}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}})^k$$

$$D_k = \lim_{n\to\infty} \frac{1}{n} \operatorname{tr}(\frac{1}{N}\mathbf{D}\mathbf{D}^{\mathrm{H}})^k$$

- For that model, we have the relations

$$\begin{aligned} M_1 &= D_1 + 1 \\ M_2 &= D_2 + (2+2c)D_1 + (1+c) \\ M_3 &= D_3 + (3+3c)D_2 + 3cD_1{}^2 + (1+3c+c^2) \end{aligned}$$

hence

$$\begin{aligned} D_1 &= M_1 - 1 \\ D_2 &= M_2 - (2+2c)M_1 + (1+c) \\ D_3 &= M_3 - (3+3c)M_2 - 3cM_1{}^2 + (6c^2+18c+6)M_1 - (4c^2+12c+4) \end{aligned}$$

Free deconvolution: Eigenvalue Inference

- For practical finite size applications, the deconvolved moments will exhibit errors. Different strategies are available,
- direct inversion with Newton-Girard formulas. Assuming perfect evaluation of $\frac{1}{K}\sum_{k=1}^{K} P_k^m$, $P_1, \ldots, P_K$ are given by the K solutions of the polynomial

$$X^K - \Pi_1 X^{K-1} + \Pi_2 X^{K-2} - \ldots + (-1)^K \Pi_K$$

where the Π_m's (known as the *elementary symmetric polynomials*) are iteratively defined as

$$(-1)^k k \Pi_k + \sum_{i=1}^{k} (-1)^{k+i} S_i \Pi_{k-i} = 0$$

where $S_k = \sum_{i=1}^{k} P_i^k$.

  - may lead to non-real solutions!
  - does not minimize any conventional error criterion
  - convenient for one-shot power inference
  - when multiple realizations are available, statistical solutions are preferable

Free deconvolution: inferring powers

- alternative approach: estimators that minimize conventional error metrics

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

- for the model $\mathbf{Y} = \mathbf{T}^{\frac{1}{2}}\mathbf{X}$, an asymptotic central limit result is known for the moments, i.e. for $M_k^{(N)}$ the order k empirical moment of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$ and M_k its limit, as $N \to \infty$,

$$N\left(M_k^{(N)} - M_k\right) \Rightarrow X$$

  where X is a central Gaussian random variable.
- then maximum-likelihood or MMSE estimators can then be used to find the moments of $\mathbf{T}$.

The Stieltjes Transform Method

- Consider the *sample covariance matrix* model

$$\mathbf{Y} \triangleq \mathbf{T}^{\frac{1}{2}}\mathbf{X} \in \mathbb{C}^{N\times n},\ \mathbf{B}_N = \frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathrm{H}} \in \mathbb{C}^{N\times N},\ \underline{\mathbf{B}}_N = \frac{1}{n}\mathbf{Y}^{\mathrm{H}}\mathbf{Y} \in \mathbb{C}^{n\times n}$$

where $\mathbf{T} \in \mathbb{C}^{N\times N}$ has eigenvalues $t_1, \ldots, t_K$, t_k with multiplicity N_k and $\mathbf{X} \in \mathbb{C}^{N\times n}$ is i.i.d. zero mean, variance 1.

J. W. Silverstein, Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," J. of Multivariate Analysis, vol. 54, no. 2, pp. 175-192, 1995.

- If $F^{\mathbf{T}} \Rightarrow T$, then $m_{F^{\mathbf{B}_N}}(z) = m_{\mathbf{B}_N}(z) \xrightarrow{\text{a.s.}} m_F(z)$ such that

$$m_T\left(-1/m_{\underline{F}}(z)\right) = -z m_{\underline{F}}(z) m_F(z)$$

with $m_{\underline{F}}(z) = c m_F(z) + (c-1)\frac{1}{z}$ and $N/n \to c$.

Complex integration

- From Cauchy integral formula, denoting $\mathcal{C}_k$ a contour enclosing **only** t_k,

$$t_k = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{\omega}{\omega - t_k} d\omega = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{1}{N_k}\sum_{j=1}^{K} N_j \frac{\omega}{\omega - t_j} d\omega = \frac{N}{2\pi i N_k}\oint_{\mathcal{C}_k} \omega m_T(\omega) d\omega.$$

- After the variable change $\omega = -1/m_{\underline{F}}(z)$,

$$t_k = \frac{N}{N_k}\frac{1}{2\pi i}\oint_{\mathcal{C}_{\underline{F},k}} z m_F(z) \frac{m'_{\underline{F}}(z)}{m^2_{\underline{F}}(z)} dz,$$

- When the system dimensions are large,

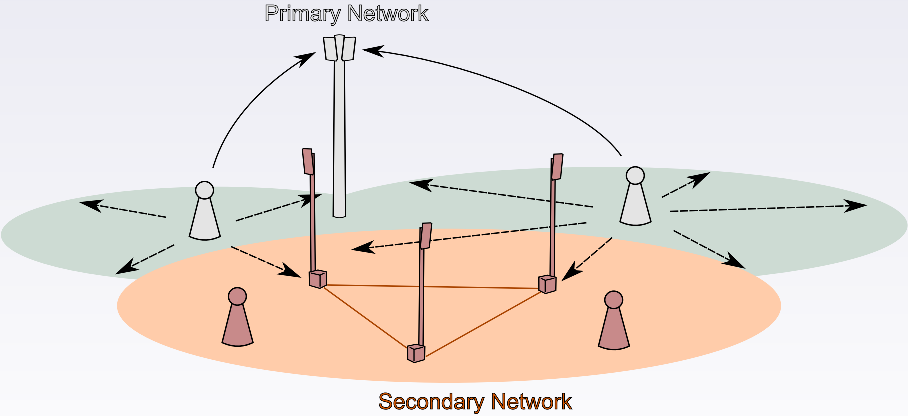
$$m_F(z) \simeq m_{\mathbf{B}_N}(z) \triangleq \frac{1}{N}\sum_{k=1}^{N} \frac{1}{\lambda_k - z}, \quad \text{with} \quad (\lambda_1, \ldots, \lambda_N) = \mathrm{eig}(\mathbf{B}_N) = \mathrm{eig}(\mathbf{Y}\mathbf{Y}^{\mathsf{H}}).$$

- Dominated convergence arguments then show

$$t_k - \hat{t}_k \xrightarrow{\text{a.s.}} 0 \quad \text{with} \quad \hat{t}_k = \frac{N}{N_k}\frac{1}{2\pi i}\oint_{\mathcal{C}_{\underline{F},k}} z m_{\mathbf{B}_N}(z) \frac{m'_{\underline{\mathbf{B}}_N}(z)}{m^2_{\underline{\mathbf{B}}_N}(z)} dz = \frac{n}{N_k}\sum_{m\in\mathcal{N}_k} (\lambda_m - \mu_m)$$

with $\mathcal{N}_k$ the indexes of cluster k and $\mu_1 < \ldots < \mu_N$ are the ordered eigenvalues of the matrix $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{n}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)^{\mathsf{T}}$.

Application Context: Coverage range in Femtocells



Problem statement

- a device embedded with N antennas receives a signal
  - originating from multiple sources
  - number of sources K is not necessarily known
  - source k is equipped with n_k antennas (ideally $n_k >> 1$)
  - signal k goes through unknown MIMO channel $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$
  - the variance σ^2 of the additive noise is not necessarily known
- the problem is to infer
  - $P_1, \ldots, P_K$ knowing $K, n_1, \ldots, n_K$
  - $P_1, \ldots, P_K$ and $n_1, \ldots, n_K$ knowing K
  - $K, P_1, \ldots, P_K$ and $n_1, \ldots, n_K$
- we will regard the problem under the angle of
  - free deconvolution: i.e. from the moments of the receive $\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$, infer those of $\mathbf{P}$, and infer on $\mathbf{P}$
  - Stieltjes transform: i.e. from analytical formulas on the asymptotic eigenvalue distribution of $\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$, we derive consistent estimates of each P_k.

System model

- at time t, source k transmit signal $\mathbf{x}_k^{(t)} \in \mathbb{C}^{n_k}$ with i.i.d. entries of zero mean and variance 1.
- we denote P_k the power emitted by user k
- the channel $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$ from user k to the receiver has i.i.d. entries of zero mean and variance $1/N$.
- at time t, the additive noise is denoted $\sigma \mathbf{w}^{(t)}$, with $\mathbf{w}^{(t)} \in \mathbb{C}^N$ with i.i.d. entries of zero mean and variance 1.
- hence the receive signal $\mathbf{y}^{(t)}$ at time t,

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

Gathering M time instant into $\mathbf{Y} = [\mathbf{y}^{(1)} \dots \mathbf{y}^{(M)}] \in \mathbb{C}^{N \times M}$, this can be written

$$\mathbf{Y} = \sum_{k=1}^{K} \mathbf{H}_k \sqrt{P_k} \mathbf{X}_k + \sigma \mathbf{W} = \mathbf{H} \mathbf{P}^{\frac{1}{2}} \mathbf{X} + \sigma \mathbf{W}$$

with $\mathbf{H} = [\mathbf{H}_1 \dots \mathbf{H}_K] \in \mathbb{C}^{N \times n}$, $n = \sum_{k=1}^{K} n_k$,
$\mathbf{P} = \mathrm{diag}(P_1, \dots, P_1, P_2, \dots, P_2, \dots, P_K, \dots, P_K)$ where P_k has multiplicity n_k on the diagonal, $\mathbf{X}^{\mathsf{H}} = [\mathbf{X}_1^{\mathsf{H}} \dots \mathbf{X}_K^{\mathsf{H}}]^{\mathsf{H}} \in \mathbb{C}^{n \times M}$, $\mathbf{X}_k = [\mathbf{x}_k^{(1)} \dots \mathbf{x}_k^{(M)}] \in \mathbb{C}^{n_k \times M}$, $\mathbf{W}$ defined similarly.

Reminder on free deconvolution

- Free probability provides tools to compute

$$d_k = \frac{1}{K}\sum_{i=1}^{K}\lambda(\mathbf{P})^k = \frac{1}{K}\sum_{i=1}^{K}P_i^k$$

as a function of

$$m_k = \frac{1}{N}\sum_{i=1}^{N}\lambda\left(\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^k$$

- One can obtain all the successive sum powers of $P_1, \ldots, P_K$.
- From that, we can infer on the values of each P_k!
- The tools come from the relations,
  - cumulant to moment (and also moment to cumulant),

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

  - Sums of cumulants for *asymptotically free* **A** and **B** (of measure $\mu_A \boxplus \mu_B$),

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

  - Products of cumulants for *asymptotically free* **A** and **B** (of measure $\mu_A \boxtimes \mu_B$),

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

  - Moments of information plus noise models $\mathbf{B}_N = \frac{1}{n}\left(\mathbf{A}_N + \sigma\mathbf{W}_N\right)\left(\mathbf{A}_N + \sigma\mathbf{W}_N\right)^{\mathrm{H}}$,

$$\mu_B = \left((\mu_A \boxslash \mu_c) \boxplus \delta_{\sigma^2}\right) \boxtimes \mu_c$$

  with μ_c the Marcenko-Pastur law with ratio c.

Free deconvolution approach

- one can deconvolve $\mathbf{YY}^{\mathsf{H}}$ in three steps,
  - an information-plus-noise model with "deterministic matrix" $\mathbf{HP}^{\frac{1}{2}}\mathbf{XX}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}$,

$$\mathbf{YY}^{\mathsf{H}} = (\mathbf{HP}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W})(\mathbf{HP}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W})^{\mathsf{H}}$$

  - from $\mathbf{HP}^{\frac{1}{2}}\mathbf{XX}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}$, up to a Gram matrix commutation, we can deconvolve the signal $\mathbf{X}$,

$$\mathbf{P}^{\frac{1}{2}}\mathbf{HH}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{XX}^{\mathsf{H}}$$

  - from $\mathbf{P}^{\frac{1}{2}}\mathbf{HH}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}$, a new matrix commutation allows one to deconvolve $\mathbf{HH}^{\mathsf{H}}$

$$\mathbf{PHH}^{\mathsf{H}}$$

Free deconvolution operations

In terms of free probability operations, this is

- noise deconvolution

$$\mu_{\frac{1}{M}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}} = \left(\left(\mu_{\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}} \boxslash \mu_c\right) \boxminus \delta_{\sigma^2}\right) \boxtimes \mu_c$$

  with μ_c the Marcenko-Pastur law and $c = N/M$.

- signal deconvolution

$$\mu_{\frac{1}{M}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}} = \frac{N}{n}\mu_{\frac{1}{M}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}} + \left(1 - \frac{N}{n}\right)\delta_0$$

- channel deconvolution

$$\mu_{\mathbf{P}} = \mu_{\mathbf{P}\frac{1}{n}\mathbf{H}^{\mathsf{H}}\mathbf{H}} \boxslash \mu_{\eta_{c_1}}$$

  with $c_1 = n/N$

Free deconvolution: moments

- from the three previous steps (plus addition of null eigenvalues), the moments of $\mathbf{P}$ can be computed from those of $\mathbf{YY}^{\mathrm{H}}$.
- this process can be automatized by combinatorics softwares
- finite size formulas are also available
- the first moments m_k of $\frac{1}{M}\mathbf{YY}^{\mathrm{H}}$ as a function of the first moments d_k of $\mathbf{P}$ read

$$
\begin{aligned}
m_1 &= N^{-1}nd_1 + 1 \\
m_2 &= \left(N^{-2}M^{-1}n + N^{-1}n\right) d_2 + \left(N^{-2}n^2 + N^{-1}M^{-1}n^2\right) d_1^2 \\
&\quad + \left(2N^{-1}n + 2M^{-1}n\right) d_1 + \left(1 + NM^{-1}\right) \\
m_3 &= \left(3N^{-3}M^{-2}n + N^{-3}n + 6N^{-2}M^{-1}n + N^{-1}M^{-2}n + N^{-1}n\right) d_3 \\
&\quad + \left(6N^{-3}M^{-1}n^2 + 6N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 3N^{-1}M^{-1}n^2\right) d_2 d_1 \\
&\quad + \left(N^{-3}M^{-2}n^3 + N^{-3}n^3 + 3N^{-2}M^{-1}n^3 + N^{-1}M^{-2}n^3\right) d_1^3 \\
&\quad + \left(6N^{-2}M^{-1}n + 6N^{-1}M^{-2}n + 3N^{-1}n + 3M^{-1}n\right) d_2 \\
&\quad + \left(3N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 9N^{-1}M^{-1}n^2 + 3M^{-2}n^2\right) d_1^2 \\
&\quad + \left(3N^{-1}M^{-2}n + 3N^{-1}n + 9M^{-1}n + 3NM^{-2}n\right) d_1
\end{aligned}
$$

where

$$
m_k = \frac{1}{N}\sum_{i=1}^{N} \lambda\left(\frac{1}{M}\mathbf{YY}^{\mathrm{H}}\right)^k \text{ and } d_k = \frac{1}{K}\sum_{i=1}^{K} \lambda(\mathbf{P})^k = \frac{1}{K}\sum_{i=1}^{K} P_i^k
$$

Free deconvolution: inferring powers

Direct inversion with Newton-Girard formulas. Assuming perfect evaluation of $\frac{1}{K}\sum_{k=1}^{K} P_k^m$, $P_1, \ldots, P_K$ are given by the K solutions of the polynomial

$$X^K - \Pi_1 X^{K-1} + \Pi_2 X^{K-2} - \ldots + (-1)^K \Pi_K$$

where the Π_m's (known as the *elementary symmetric polynomials*) are iteratively defined as

$$(-1)^k k \Pi_k + \sum_{i=1}^{k} (-1)^{k+i} S_i \Pi_{k-i} = 0$$

where $S_k = \sum_{i=1}^{k} P_i^k$.

- may lead to non-real solutions!
- does not minimize any conventional error criterion
- convenient for one-shot power inference
- when multiple realizations are available, statistical solutions are preferable

Stieltjes transform approach

- Remember the matrix model

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

with $\mathbf{W}, \mathbf{Y} \in \mathbb{C}^{N\times M}$, $\mathbf{H} \in \mathbb{C}^{N\times n}$, $\mathbf{X} \in \mathbb{C}^{n\times M}$, and $\mathbf{P} \in \mathbb{C}^{n\times n}$ diagonal.

- this can be written in the following way

$$\mathbf{Y} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix} \in \mathbb{C}^{N\times M}$$

and extend it into the matrix

$$\mathbf{Y}_{\mathrm{ext}} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix} \in \mathbb{C}^{(N+n)\times M}$$

which is a sample covariance matrix model.

- the population covariance matrix is

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

itself a sample covariance matrix.

Asymptotic spectrum

R. Couillet, J. W. Silverstein, Z. Bai, M. Debbah, "Eigen-inference Energy Estimation of Multiple Sources", IEEE Trans. on Information Theory, 2010, *submitted*.

- the asymptotic spectrum of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ has Stietljes transform $m(z)$, $z \in \mathbb{C}^+$, such that

$$m(z) = \frac{M}{N}\underline{m}(z) + \frac{M-N}{N}\frac{1}{z}$$

where $\underline{m}(z)$ is the unique solution in $\mathbb{C}^+$ of

$$\frac{1}{\underline{m}(z)} = -\sigma^2 + \frac{1}{f(z)} - \frac{1}{N}\sum_{k=1}^{K}\frac{n_k P_k}{1 + P_k f(z)}$$

where $f(z)$ is given by

$$f(z) = \frac{M-N}{N}\underline{m}(z) - \frac{M}{N}z\underline{m}(z)^2$$

and we want to determine P_k.

Asymptotic spectrum of $\frac{1}{M}\mathbf{YY}^{\mathsf{H}}$

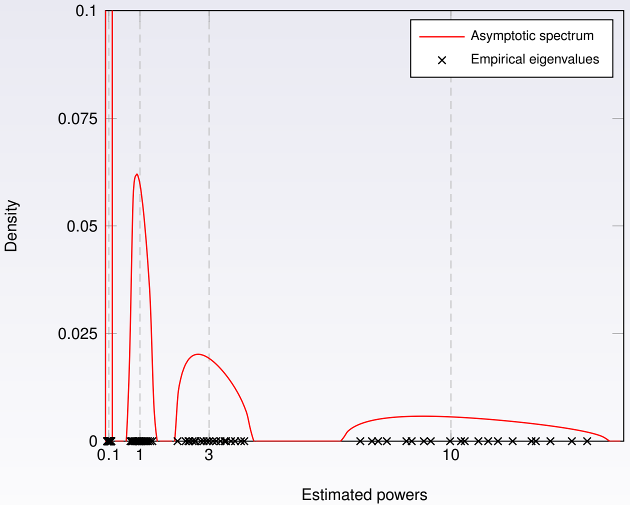


Figure: Empirical and asymptotic eigenvalue distribution of $\frac{1}{M}\mathbf{YY}^{\mathsf{H}}$ when $\mathbf{P}$ has three distinct entries $P_1 = 1$, $P_2 = 3$, $P_3 = 10$, $n_1 = n_2 = n_3$, $N/n = 10$, $M/N = 10$, $\sigma^2 = 0.1$. Empirical test: $n = 60$.

Stieltjes transform approach: final result

R. Couillet, J. W. Silverstein, Z. Bai, M. Debbah, “Eigen-inference Energy Estimation of Multiple Sources”, IEEE Trans. on Information Theory, 2010, *submitted*.

Theorem

Let $\mathbf{B}_N = \frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}} \in \mathbb{C}^{N\times N}$*, with* $\mathbf{Y}$ *defined as previously. Denote its ordered eigenvalues vector* $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$*,* $\lambda_1 < \ldots, \lambda_N$*. Further assume asymptotic spectrum separability. Then, for* $k \in \{1, \ldots, K\}$*, as N, n, M grow large, we have*

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

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

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

with $\mathcal{N}_k = \{N - \sum_{i=k}^{K} n_i + 1, \ldots, N - \sum_{i=k+1}^{K} n_i\}$ *the set of indexes matching the cluster corresponding to* P_k*,* $(\eta_1, \ldots, \eta_N)$ *the ordered eigenvalues of* $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$ *and* $(\mu_1, \ldots, \mu_N)$ *the ordered eigenvalues of* $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{M}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$*.*

Comments on the result

- very compact formula
- low computational complexity
- assuming cluster separation, it allows also to infer the number of eigenvalues, as well as the multiplicity of each eigenvalue.
- however, strong requirement on cluster separation
- if separation is not true, the mean of the eigenvalues instead of the eigenvalues themselves is computed.
- it is possible to infer K, all n_k and all P_k using the Stieltjes transform method.

Multi-Source Power Estimation: Performance Comparison

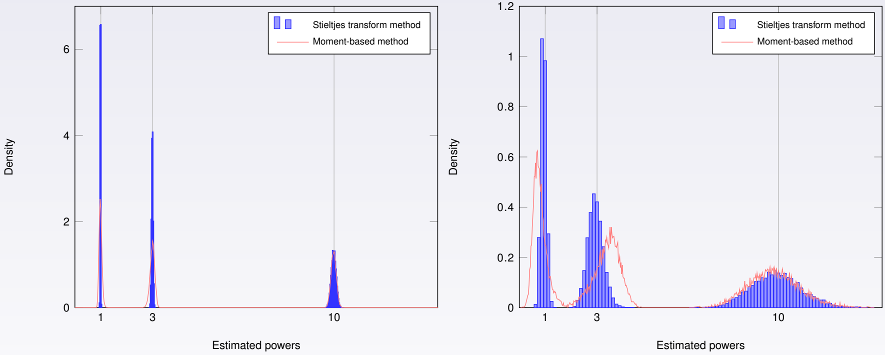


Figure: Multi-source power estimation, for $K = 3$, $P_1 = 1$, $P_2 = 3$, $P_3 = 10$, $n_1/n = n_2/n = n_3/n = 1/3$,$n/N = N/M = 1/10$, $\mathrm{SNR} = 10$ dB, for $10,000$ simulation runs; Top $n = 60$, bottom $n = 6$.

Multi-Source Power Estimation: Performance Comparison

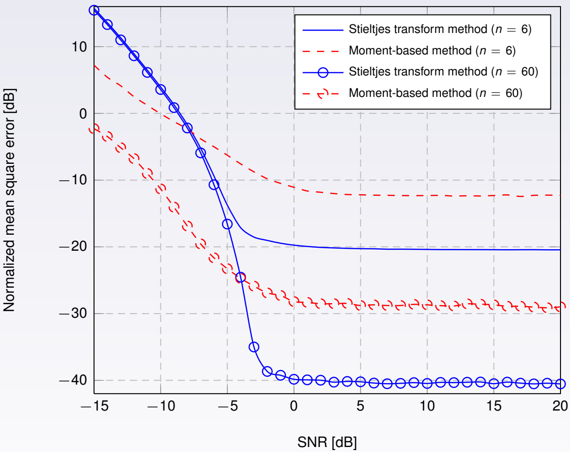


Figure: Normalized mean square error of the vector $(\hat{P}_1, \hat{P}_2, \hat{P}_3)$, $P_1 = 1$, $P_2 = 3$, $P_3 = 10$, $n_1/n = n_2/n = n_3/n = 1/3$,$n/N = N/M = 1/10$, for 10, 000 simulation runs.

General comments and steps left to fulfill

- up to this day
  - the moment approach is much simpler to derive
  - it does not require any cluster separation
  - the finite size case is treated in the mean, which the Stieltjes transform approach cannot do.
  - however, the Stieltjes transform approach makes full use of the spectral knowledge, when the moment approach is limited to a few moments.
  - the results are more natural, and more “telling”
- in the future, it is expected that the cluster separation requirement can be overtaken.
- a natural general framework attached to the Stieltjes transform method could arise
- central limit results on the estimates is expected

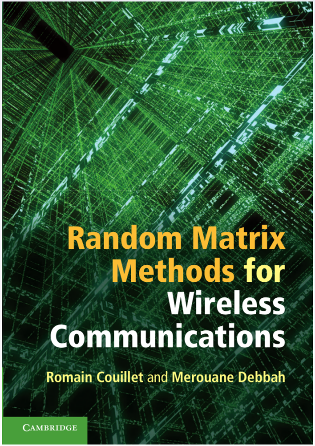
Related bibliography

- N. El Karoui, “Spectrum estimation for large dimensional covariance matrices using random matrix theory,” Annals of Statistics, vol. 36, no. 6, pp. 2757-2790, 2008.
- N. R. Rao, J. A. Mingo, R. Speicher, A. Edelman, “Statistical eigen-inference from large Wishart matrices,” Annals of Statistics, vol. 36, no. 6, pp. 2850-2885, 2008.
- R. Couillet, M. Debbah, “Free deconvolution for OFDM multicell SNR detection”, PIMRC 2008, Cannes, France.
- 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.
- R. Couillet, J. W. Silverstein, M. Debbah, “Eigen-inference for multi-source power estimation”, *submitted to ISIT 2010*.
- Z. D. Bai, J. W. Silverstein, “No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices,” The Annals of Probability, vol. 26, no.1 pp. 316-345, 1998.
- Z. D. Bai, J. W. Silverstein, “CLT of linear spectral statistics of large dimensional sample covariance matrices,” Annals of Probability, vol. 32, no. 1A, pp. 553-605, 2004.
- 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.
- Ø. Ryan, M. Debbah, “Free Deconvolution for Signal Processing Applications,” IEEE International Symposium on Information Theory, pp. 1846-1850, 2007.

Selected authors' recent bibliography

- Articles in Journals,
  - R. Couillet, J. W. Silverstein, Z. Bai, M. Debbah, "Eigen-Inference for Energy Estimation of Multiple Sources," IEEE Trans. on Information Theory, 2010, *submitted*.
  - R. Couillet, J. W. Silverstein, M. Debbah, "A Deterministic Equivalent for the Capacity Analysis of Correlated Multi-User MIMO Channels," IEEE Trans. on Information Theory, 2010, *accepted for publication*.
  - P. Bianchi, J. Najim, M. Maida, M. Debbah, "Performance of Some Eigen-based Hypothesis Tests for Collaborative Sensing," IEEE Trans. on Information Theory, 2010, *accepted for publication*.
  - R. Couillet, M. Debbah, "A Bayesian Framework for Collaborative Multi-Source Signal Sensing," IEEE Trans. on Signal Processing, 2010, *accepted for publication*.
  - S. Wagner, R. Couillet, M. Debbah, D. T. M. Slock, "Large System Analysis of Linear Precoding in MISO Broadcast Channels with Limited Feedback," IEEE Trans. on Information Theory, 2010, *submitted*.
  - A. Masucci, Ø. Ryan, S. Yang, M. Debbah, "Gaussian Finite Dimensional Statistical Inference," IEEE Trans. on Information Theory, 2009, *submitted*.
  - Ø. Ryan, M. Debbah, "Asymptotic Behaviour of Random Vandermonde Matrices with Entries on the Unit Circle," IEEE Trans. on Information Theory, vol. 55, no. 7 pp. 3115-3148, July 2009.
  - M. Debbah, R. Müller, "MIMO channel modeling and the principle of maximum entropy," IEEE Trans. on Information Theory, vol. 51, no. 5, pp. 1667-1690, 2005.
- Articles in International Conferences
  - **R. Couillet, S. Wagner, M. Debbah, A. Silva, "The Space Frontier: Physical Limits of Multiple Antenna Information Transfer", Inter-Perf 2008, Athens, Greece. BEST STUDENT PAPER AWARD.**
  - R. Couillet, M. Debbah, V. Poor, "Self-organized spectrum sharing in large MIMO multiple access channels", submitted to ISIT 2010.
  - R. Couillet, M. Debbah, "Uplink capacity of self-organizing clustered orthogonal CDMA networks in flat fading channels", ITW 2009 Fall, Taormina, Sicily.

Available in bookstores!



Detailed outline

Romain Couillet, Mérouane Debbah, *Random Matrix Methods for Wireless Communications*.