Applications of Random Matrix Theory to Multi-Antenna Signal Processing

Xavier Mestre
xavier.mestre@cttc.cat
Co-authors: Francisco Rubio, Miguel A. Lagunas

Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)
Parc Mediterrani de la Tecnologia, Castelldefels, Barcelona (Spain), http://www.cttc.es/

NEWCOM Dept. 1 Meeting, Toulouse, May 15, 2006
Outline

- Introduction to random matrix theory.
- The sample covariance matrix: asymptotic properties.
- Modeling the finite sample size effect using random matrix theory.
- Determination of the optimum loading factor in MVDR beamformers.
- Subspace-based detection of Directions-of-Arrival (DoA): G-MUSIC.
- Practical implementation of reduced-rank MVDR filters.
- Summary and conclusions.
Introduction to random matrix theory

Random Matrix Theory studies the asymptotic behavior of the eigenvalues and eigenvectors of random matrices when their dimensions increase without bound. Given a generic $M \times M$ Hermitian random matrix $M$ ($M = M^H$), with eigenvalues $\lambda_1 \ldots \lambda_M$, we define the empirical distribution function (e.d.f.) of its eigenvalues as $F_M(\lambda) = \frac{1}{M} \# \{\lambda_m \leq \lambda, \ m = 1 \ldots M\}$.

For many random matrix models, it turns out that $F_M(\lambda) \rightarrow F(\lambda)$ almost surely, where $F(\lambda)$ is a deterministic probability distribution.
The sample covariance matrix

We assume that we collect $N$ independent samples (snapshots) from an array of $M$ antennas:

$$\hat{R} = \frac{1}{N} \sum_{n=1}^{N} y(n)y^H(n).$$
The sample covariance matrix: asymptotic properties

When both $M, N \to \infty$, $M/N \to c$, $0 < c < \infty$, the e.d.f. of the eigenvalues of $\hat{R}$ tends to a deterministic density function. Example: $R$ has 4 eigenvalues $\{1, 2, 3, 7\}$ with equal multiplicity.
Modeling the finite sample size effect using random matrix theory

Random Matrix Theory offers the possibility of analyzing the behavior of different quantities depending on $\hat{R}$ when the sample size and the number of sensors/antennas have the same order of magnitude.

Assume that we want to analyze the asymptotic behavior of a certain scalar function of $\hat{R}$, namely $f\left(\hat{R}\right)$.

- **Traditional Approach:** Assuming that the number of samples is high, we might establish that $f\left(\hat{R}\right) \to f\left(R\right)$ in some stochastic sense as $N \to \infty$ while $M$ remains fixed.

- **New Approach:** In order to characterize the situation where $M, N$ have the same order of magnitude, one might consider the limit $N, M \to \infty$, $M/N \to c$, $0 < c < \infty$. Note that, in general,

$$f\left(\hat{R}\right) \to f\left(R\right), \quad N, M \to \infty, M/N \to c$$

For example:

$$\frac{1}{M} \text{tr} \left[ \hat{R}^{-1} \right] \to (1 - c)^{-1} \frac{1}{M} \text{tr} \left[ R^{-1} \right], \quad c < 1.$$
$M, N$-consistent versus $N$-consistent estimators

When designing an estimator of a certain scalar function of $\mathbf{R}$, namely $f(\mathbf{R})$, one can distinguish between:

- **Traditional $N$-consistency:** Consistency when $N \to \infty$ while $M$ remains fixed.
- **$M, N$-consistency:** Consistency when $M, N \to \infty$ at the same rate.

We observe that $M, N$-consistency guarantees a good behavior when the number of samples $N$ has the same order of magnitude as the observation dimension $M$.

The objective of G-estimation (V.L. Girko) is to provide a systematic approach for the derivation of $M, N$-consistent estimators of different scalar functions of the true covariance matrix. For example, the G-estimator of $\frac{1}{M} \text{tr} [\mathbf{R}^{-1}]$ will be

$$\frac{(1 - c)}{M} \text{tr} \left[ \hat{\mathbf{R}}^{-1} \right]$$
Determination of the optimum loading factor in MVDR beamformers (i)

Assume that we receive a signal with an antenna array of $M$ elements:

$$y(n) = s(n)s_d + n(n).$$

In this signal model it is customary to implement a linear filter (beamformer) to enhance the contribution of $s(n)$ and null out the noise term $n(n)$. The Minimum Variance solution for the filter takes a form proportional to

$$w = R^{-1}s_d \implies w = \hat{R}^{-1}s_d \implies w = \left(\hat{R} + \alpha I_M\right)^{-1}s_d$$

**Problem:** how to fix $\alpha$ if we now nothing about the scenario? Maximize SINR

$$\text{SINR} = \left(\frac{q(\alpha)}{P_s} - 1\right)^{-1} \quad q(\alpha) = \frac{s_d^H(\hat{R} + \alpha I_M)^{-1}R(\hat{R} + \alpha I_M)^{-1}s_d}{\left(s_d^H(\hat{R} + \alpha I_M)^{-1}s_d\right)^2}$$
Determination of the optimum loading factor in MVDR beamformers (ii)

1. Analyze the behavior when $M, N \to \infty$ at the same rate

$$\left| \text{SINR} - \left( \frac{\bar{q}(\alpha)}{P_s} - 1 \right)^{-1} \right| \xrightarrow{a.s.} 0 \quad \bar{q}(\alpha) = \frac{1}{1 - \frac{c}{M} \sum_{m=1}^{M} \left( \frac{\lambda_m}{\lambda_m + \gamma} \right)^2} \frac{s_d^H (\hat{R} + \gamma \mathbf{I}_M)^{-1} \hat{R} \left( \hat{R} + \gamma \mathbf{I}_M \right)^{-1} s_d}{\left( s_d^H (\hat{R} + \gamma \mathbf{I}_M)^{-1} s_d \right)^2}$$

where $\gamma = \alpha (1 + cb)$ and $b$ being the positive solution to

$$b = \frac{1}{M} \sum_{m=1}^{M} \frac{\lambda_m (1 + cb)}{\lambda_m + \alpha (1 + cb)}$$

2. Find an $M, N$-consistent estimator of $\bar{q}(\alpha)$:

$$\hat{q}(\alpha) = \frac{1}{\left( 1 - \frac{c}{M} \sum_{m=1}^{M} \frac{\hat{\lambda}_m}{\hat{\lambda}_m + \alpha} \right)^2} \frac{s_d^H \left( \hat{R} + \alpha \mathbf{I}_M \right)^{-1} \hat{R} \left( \hat{R} + \alpha \mathbf{I}_M \right)^{-1} s_d}{\left( s_d^H \left( \hat{R} + \alpha \mathbf{I}_M \right)^{-1} s_d \right)^2}$$
Determination of the optimum loading factor in MVDR beamformers (iii)

Cumulative distribution of the output SINR (signals coming from uniformly distributed DOAs)

![Graph 1: Cumulative distribution of the output SINR, $M=5, N=7, K=2+1$](image1)

![Graph 2: Cumulative distribution of the output SINR, $M=50, N=70, K=29+1$](image2)
Subspace-based detection of Directions-of-Arrival (DoA): G-MUSIC

Traditional subspace-based DoA detection exploits the orthogonality between signal and noise subspaces. The true spatial covariance matrix can be structured as

$$\mathbf{R} = \mathbf{S}(\Theta) \Phi \mathbf{S}(\Theta)^H + \sigma^2 \mathbf{I}_M$$

where $\mathbf{S}(\Theta)$ is an $M \times K$ matrix that contains the steering vectors corresponding to the $K$ different sources,

$$\mathbf{S}(\Theta) = \begin{bmatrix} \mathbf{s}(\theta_1) & \mathbf{s}(\theta_2) & \cdots & \mathbf{s}(\theta_K) \end{bmatrix}.$$  

The eigendecomposition of $\mathbf{R}$:

$$\mathbf{R} = \begin{bmatrix} \mathbf{E}_S & \mathbf{E}_N \end{bmatrix} \begin{bmatrix} \mathbf{A}_S & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I}_{M-K} \end{bmatrix} \begin{bmatrix} \mathbf{E}_S & \mathbf{E}_N \end{bmatrix}^H$$

and it turns out that $\mathbf{E}_N^H \mathbf{s}(\theta_k) = \mathbf{0}$. The MUSIC algorithm uses the sample noise eigenvectors:

$$\eta_{\text{MUSIC}}(\theta) = \frac{1}{\mathbf{s}^H(\theta) \hat{\mathbf{E}}_N \hat{\mathbf{E}}_N^H \mathbf{s}(\theta)}.$$
Subspace-based detection of DoAs: G-MUSIC (ii)

The MUSIC algorithm suffers from the breakdown effect. The performance breaks down when the number of samples or the SNR falls below a certain threshold. Cause: $\hat{E}_N$ is not a very good estimator of $E_N$ when $M, N$ have the same order of magnitude.

The performance breakdown effect can be easily analyzed using random matrix theory, especially under a noise eigenvalue separation assumption: $|\eta_{\text{MUSIC}}(\theta) - \bar{\eta}_{\text{MUSIC}}(\theta)| \to 0$

$$\bar{\eta}_{\text{MUSIC}}(\theta) = \left( s^H(\theta) \left( \sum_{k=1}^{M} w(k) e_k e_k^H \right) s(\theta) \right)^{-1}$$

$$w(k) = \begin{cases} 1 - \frac{1}{M-K} \sum_{r=M-K+1}^{M} \left( \frac{\mu_r \lambda_r - \sigma^2}{\lambda_r - \sigma^2} - \frac{\mu_1}{\lambda_r - \mu_1} \right) & k \leq M - K \\ \frac{\sigma^2}{\lambda_k - \sigma^2} - \frac{\mu_1}{\lambda_k - \mu_1} & k > M - K \end{cases}$$

where $\{\mu_r, r = 1, \ldots, M\}$ are the solutions to $\frac{1}{M} \sum_{r=1}^{M} \frac{\lambda_r}{\lambda_r - \mu} = \frac{1}{c}$. 
Subspace-based detection of DoAs: G-MUSIC (iii)

We consider a scenario with three sources impinging on a ULA \((d/\lambda_c = 0.5)\) from DoAs: \(-10^\circ, 35^\circ, 40^\circ\) (SNR=10dB).
Subspace-based detection of DoAs: G-MUSIC (iv)

We propose to use an $M, N$-consistent estimator of the MUSIC cost function:

$$\eta_{\text{G-MUSIC}}(\theta) = \left( S^H(\theta) \left( \sum_{k=1}^{M} \phi(k) \hat{e}_k \hat{e}_k^H \right) S(\theta) \right)^{-1}$$

$$\phi(k) = \begin{cases} 
1 + \sum_{r=M-K+1}^{M} \left( \frac{\hat{\lambda}_r}{\lambda_k - \lambda_1} - \frac{\hat{\mu}_r}{\lambda_k - \hat{\mu}_r} \right) & k \leq M - K \\
- \sum_{r=1}^{M-K} \left( \frac{\hat{\lambda}_r}{\lambda_k - \lambda_1} - \frac{\hat{\mu}_r}{\lambda_k - \hat{\mu}_r} \right) & k > M - K 
\end{cases}$$

where now $\hat{\mu}_1, \ldots, \hat{\mu}_M$ are the solutions to the equation

$$\frac{1}{M} \sum_{k=1}^{M} \frac{\hat{\lambda}_k}{\lambda_k - \hat{\mu}} = \frac{1}{c}.$$
Subspace-based detection of DoAs: G-MUSIC (v)

Comparative evaluation of MUSIC and G-MUSIC via simulations. Scenario with four sources \((-20^\circ, -10^\circ, 35^\circ, 37^\circ, \text{SNR}=10\text{dB})\), ULA \((M = 20, \frac{d}{\lambda_c} = 0.5)\).
Practical implementation of reduced-rank MVDR filters

It is often impractical to implement the MVDR filter using matrix inversion:

\[ w = R^{-1}s_d \implies w = \hat{R}^{-1}s_d \]

On the one hand \( \hat{R} \) might be complex to invert; on the other, \( \hat{R}^{-1} \) is not the best estimator of \( R^{-1} \).

**Idea:** approximate \( R^{-1} \) with powers of the sample covariance matrix \( \{ \hat{R}, \hat{R}^2, \ldots, \hat{R}^D \} \), namely

\[ w = R^{-1}s_d \implies w = \sum_{k=1}^{D} \alpha_k \hat{R}^k s_d \]

**Problem:** designing the weights \( \alpha_k, k = 1 \ldots D \), in order to optimize the performance of the filter.

It should be observed that, by virtue of the Cayley-Hamilton theorem,

\[ R^{-1} = \sum_{k=1}^{M} \beta_k R^k \]

for certain weights \( \beta_k, k = 1 \ldots M \). In practice, \( D \ll M \) is enough for good performance.
Practical implementation of reduced-rank MVDR filters (ii)

A possible design criterion for the weights $\alpha = [\alpha_1, \ldots, \alpha_D]$ is the optimization of the output SINR

$$\text{SINR} = \left( \frac{p(\alpha)}{P_s} - 1 \right)^{-1}$$

$$p(\alpha) = \frac{\sum_{m=1}^{D} \sum_{n=1}^{D} \alpha_m \alpha_n s_d^H \hat{R}^m \hat{R}^n s_d}{\left( \sum_{k=1}^{D} \alpha_k s_d^H \hat{R}^k s_d \right)^2}$$

**Asymptotic approximation:** we analyze the quantity SINR when $M, N \to \infty$ at the same rate,

$$\left| \text{SINR} - \left( \frac{\bar{p}(\alpha)}{P_s} - 1 \right)^{-1} \right| \xrightarrow{a.s.} 0$$

$$\bar{p}(\alpha) = \frac{\sum_{m=1}^{D} \sum_{n=1}^{D} \alpha_m \alpha_n \int \int \varphi(z_1, z_2) \, dz_1 \, dz_2}{\left( \sum_{k=1}^{D} \alpha_k \int z^{k-1} f(z) s_d^H (\hat{R} - f(z) I_M)^{-1} s_d \, dz \right)^2}$$

where

$$\varphi(z_1, z_2) = z_1^{m-1} z_2^{n-1} \frac{f(z_1) - f(z_2)}{z_1 - z_2} f(z_1) f(z_2) s_d^H (\hat{R} - f(z_1) I_M)^{-1} R (\hat{R} - f(z_2) I_M)^{-1} s_d$$

$$f(z) = \frac{z}{1 - c - czb(z)}$$
Practical implementation of reduced-rank MVDR filters (iii)

We propose to use an $M, N$-consistent estimator of the asymptotic output $\text{SINR} = \left( \frac{\hat{p}(\alpha)}{P_s} - 1 \right)^{-1}$, where

$$\hat{p}(\alpha) = \frac{\alpha^H A \alpha}{|\alpha^H b|^2}$$

$$A_{ij} = \frac{1}{c^2} \sum_{k=1}^{M} \sum_{l=1}^{M} \frac{\hat{\mu}_k \hat{\mu}_l}{\hat{\eta}_k \hat{\eta}_l} s_d^H \left( \hat{R} - \hat{\mu}_M I_M \right)^{-1} \hat{R} \left( \hat{R} - \hat{\mu}_M I_M \right)^{-1} s_d, \quad b_k = s_d^H \hat{R}^k s_d$$

where $\hat{\mu}_1, \ldots, \hat{\mu}_M$ are the solutions to the equation

$$\frac{1}{M} \sum_{k=1}^{M} \frac{\hat{\lambda}_k}{\hat{\lambda}_k - \mu} = \frac{1}{c} \quad \text{and} \quad \hat{\eta}_k = \frac{1}{M} \sum_{r=1}^{M} \frac{\hat{\lambda}_r}{(\hat{\lambda}_r - \hat{\mu}_k)^2}.$$

The final solution for the asymptotically optimum weights takes the form

$$\hat{\alpha}_{opt} = A^{-1} b.$$
Practical implementation of reduced-rank MVDR filters (iv)

Comparison in terms of output SINR versus rank selection ($M = 192$, $K = 24$, $N = 240, 156$).
Summary and Conclusions

• Traditional definition of consistency is not very useful to characterize estimators in situations where the number of available observations is low.

• Consistency when both the number of observations and their dimension go to infinity is not guaranteed with the sample covariance matrix.

• We propose estimators that are consistent when both the number of observations and their dimension go to infinity: good properties when the sample size is low.

• Our work has fundamentally concentrated on three applications: design of the optimum loading factor in a MVDR beamformer, GMUSIC and design of RR MVDR filters.

• The estimators have a good performance when the number of samples and the observation dimension have the same order of magnitude, and it improve as the number of antennas grows large.
NEWCOM Summer School

Random Matrix Theory for Wireless Communications

June 19-23, 2006

Centre Tecnològic de Telecomunicacions de Catalunya
Parc Mediterrani de la Tecnologia, Castelldefels, Barcelona (Spain)

http://www.cttc.cat/newcom/

Thank you for your attention!!!