Gram Matrices and Statistical Signal Processing

Imperial College London, July 2016

Khalil Elkhalil
khalil.elkhalil@kaust.edu.sa

Computer, Electrical, Mathematical Sciences and Engineering (CEMSE)

جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

A work done with

- Dr Abla Kammoun (Research scientist, KAUST).
- Prof. Tareq Y. Al-Naffouri (Associate Professor, KAUST).

- Prof. Mohamed-Slim Alouini (Professor, KAUST).

Outline

- All started with the work of E. Wigner in the 1950's in nuclear physics.

- High dimensional probability and statistics (Tao, Marcenko, Pastur...)
- Electrical Eng. at the end of the 1990's (Verdu, Tse, Debbah,...)
- Recently applied in the field of statistical signal processing and robust estimation (Mestre, Loubaton, Couillet,...)
- Applied in Machine Learning (2016, Couillet and Kammoun).
- The journey continues ...

Random matrices

$$\mathbf{X}_N = \begin{bmatrix} X_{1,1} & \cdots & X_{1,N} \\ \vdots & & \vdots \\ X_{N,1} & \cdots & X_{N,N} \end{bmatrix}.$$

where the entries $X_{i,j}$ are random variables.

Usually we're interested in the following quantities

- The **spectrum** (The distribution of the λ_i's, λ_{min} and λ_{max}).
- Some useful statistics involving the eigenvalues of $\mathbf{X}_N$

$$\mathbb{E} \sum_i \mathbf{f}(\lambda_\mathbf{i})$$

$$\mathbf{f}(\mathbf{x}) : \frac{\mathbf{1}}{\mathbf{x}^\mathbf{p}} (\mathbf{p} \in \mathbb{Z}), \log \mathbf{x}, \ldots$$

- Eigenvectors, $\cdots$

Two approaches to obtain the statistics of interest

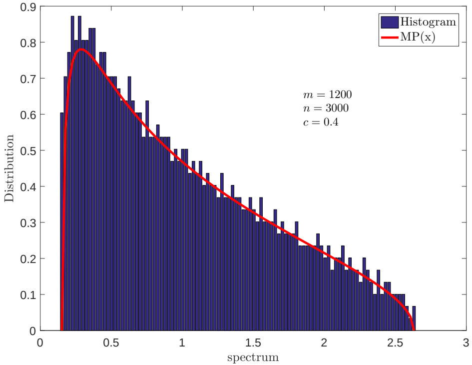
- Exact Approach (usually hard ☹).
- Asymptotic Approach (usually feasible and leads to simple expressions ☺).

$$m, n \to \infty, \quad \frac{m}{n} \to c \in (0, \infty).$$

Marcenko-Pastur's Theorem (Wishart matrices)

Let $\mathbf{X} \in \mathbb{C}^{m \times n}$ a Gaussian random matrix with i.i.d zero mean unit variance entries and $\mathbf{W} = \frac{1}{n}\mathbf{X}\mathbf{X}^*$ with spectral measure

$$\mathcal{E}(\mathbf{W}) = \frac{1}{m} \sum_i \delta_{\lambda_i(\mathbf{W})}.$$



Let $\lambda^- = (1 - \sqrt{c})^2$ and $\lambda^+ = (1 + \sqrt{c})^2$. Then as $(m, n) \to \infty$ with $\frac{m}{n} \to c \in (0, \infty)$

$$\mathcal{E}(\mathbf{W}) \xrightarrow[m,n \to +\infty]{d} \mathbb{P}_{MP}(dx) = \left(1 - \frac{1}{c}\right)^+ \delta_0(dx) + \frac{\sqrt{(\lambda^+ - x)(x - \lambda^-)}}{2\pi x c} \mathbf{1}_{[\lambda^-, \lambda^+]}(x) dx.$$

Gram random matrices

$$\begin{aligned}\mathbf{S} &= \mathbf{H}^*\mathbf{H}\\ &= \left(\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{X}\right)^*\left(\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{X}\right)\\ &= \mathbf{X}^*\mathbf{\Lambda}\mathbf{X}.\end{aligned}$$

where $\mathbf{X} \in \mathbb{C}^{n\times m}$ with i.i.d zero mean unit variance Gaussian entries and $\mathbf{\Lambda}$ is positive definite matrix with distinct eigenvalues $\theta_1, \theta_2, \cdots, \theta_n$.
This kind of matrices may arise in the following context

$$\mathbf{y} = \mathbf{H}\mathbf{v} + \mathbf{z}, \quad m < n.$$

The error covariance matrix after applying least squares (LS)

$$\mathbf{S}^{-1} = \left(\mathbf{H}^*\mathbf{H}\right)^{-1}.$$

The mean square error is given by

$$\begin{aligned}\mathbf{MSE} &= \mathbb{E}\ \text{trace}\left(\mathbf{H}^*\mathbf{H}\right)^{-1}\\ &= \mathbb{E}\sum_i \lambda_i^{-1}\left(\mathbf{H}^*\mathbf{H}\right)\\ \mathbf{f}\left(\mathbf{x}\right) &= \frac{\mathbf{1}}{\mathbf{x}}.\end{aligned}$$

Can we derive the MSE in closed form ? (Exact Approach)

In this particular case: Yes, we can ☺

Mellin transform approach

$$\mathcal{M}_{f_\lambda}(s) \triangleq \int_0^\infty \xi^{s-1} f_\lambda(\xi)\, d\xi. \tag{1}$$

$$\mathbb{E}\ \mathrm{trace}\left(\mathbf{H}^*\mathbf{H}\right)^{-p} = \lim_{s\to 0} \mathcal{M}_{f_\lambda}(s-p+1)$$

Lemma 1[1]

$$\begin{aligned}\mathcal{M}_{f_\lambda}(s) = L \sum_{j=1}^{m}\sum_{i=1}^{m} \mathcal{D}(i,j)\,\Gamma(s+j-1)\Bigg(&\theta_{n-m+i}^{n-m+s+j-2} \\ &- \sum_{l=1}^{n-m}\sum_{k=1}^{n-m}\left[\boldsymbol{\Psi}^{-1}\right]_{k,l}\theta_l^{n-m+s+j-2}\theta_{n-m+i}^{k-1}\Bigg),\end{aligned} \tag{2}$$

with $L = \frac{\det(\boldsymbol{\Psi})}{m\prod_{k<l}^{n}(\theta_l-\theta_k)\prod_{l=1}^{m-1} l!}$, $\Gamma(.)$ the Gamma function and $\boldsymbol{\Psi}$ is the $(n-m)\times(n-m)$ Vandermonde matrix given by

$$\boldsymbol{\Psi} = \begin{bmatrix} 1 & \theta_1 & \cdots & \theta_1^{n-m-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \theta_{n-m} & \cdots & \theta_{n-m}^{n-m-1}\end{bmatrix}$$

and $\mathcal{D}(i,j)$ is the (i,j)−cofactor of the $(m\times m)$ matrix $\mathcal{C}$ whose (l,k)−th entry is given by

$$[\mathcal{C}]_{l,k} = (k-1)!\left(\theta_{n-m+l}^{n-m+k-1} - \sum_{p=1}^{n-m}\sum_{q=1}^{n-m}\left[\boldsymbol{\Psi}^{-1}\right]_{p,q}\theta_{n-m+l}^{p-1}\theta_q^{n-m+k-1}\right).$$

[1] G. Alfano *et al* "Capacity of MIMO Channels with One-Sided Correlation," *ISSSTA*, 2004.

- $p \leq 0$: straightforward.
- $p > 0$: singularity issues arise.

→ Divide and conquer

$$\mathcal{M}_{f_\lambda}(s-p+1) = \mathcal{M}_1(s-p+1) + \mathcal{M}_2(s-p+1), \tag{3}$$

where

$$\mathcal{M}_1(s) = L\sum_{j=1}^{p}\sum_{i=1}^{m}\mathcal{D}(i,j)\,\Gamma(s+j-1)\Bigg(\theta_{n-m+i}^{n-m+s+j-2} - \sum_{l=1}^{n-m}\sum_{k=1}^{n-m}\left[\boldsymbol{\Psi}^{-1}\right]_{k,l}\theta_l^{n-m+s+j-2}\theta_{n-m+i}^{k-1}\Bigg).$$

$$\mathcal{M}_2(s) = L\sum_{j=p+1}^{m}\sum_{i=1}^{m}\mathcal{D}(i,j)\,\Gamma(s+j-1)\Bigg(\theta_{n-m+i}^{n-m+s+j-2} - \sum_{l=1}^{n-m}\sum_{k=1}^{n-m}\left[\boldsymbol{\Psi}^{-1}\right]_{k,l}\theta_l^{n-m+s+j-2}\theta_{n-m+i}^{k-1}\Bigg).$$

Good news

$$\lim_{s\to 0} \mathcal{M}_2(s-p+1) = 0.$$

Proof

$$\lim_{s\to 0} \mathcal{M}_2\left(s-p+1\right)$$

$$= L \sum_{j=p+1}^{m} \sum_{i=1}^{m} \mathcal{D}\left(i,j\right) \Gamma\left(-p+j\right)$$

$$\times \left(\theta_{n-m+i}^{n-m-p+j-1} - \sum_{l=1}^{n-m} \sum_{k=1}^{n-m} \left[\boldsymbol{\Psi}^{-1}\right]_{k,l} \theta_{l}^{n-m-p+j-1} \theta_{n-m+i}^{k-1} \right)$$

$$= L \sum_{j=p+1}^{m} \sum_{i=1}^{m} \left[\mathcal{D}\right]_{i,j} \left[\mathcal{C}\right]_{i,j-p}$$

$$= L \sum_{j=p+1}^{m} \left[\mathcal{D}^t \mathcal{C}\right]_{j,j-p},$$

where $\mathcal{D}$ and $\mathcal{C}$ are as defined in Lemma 1. Since $\mathcal{D}$ is the cofactor of $\mathcal{C}$, then $\mathcal{D}^t\mathcal{C} = \det\left(\mathcal{C}\right)\mathbf{I}_m$. Therefore, $\left[\mathcal{D}^t\mathcal{C}\right]_{j,j-p} = 0$ for $j = p+1, \cdots, m$.

The first moment is more involved!

$$\mathbf{a}_j = \left[\theta_1^{n-m-p+j-1}, \theta_2^{n-m-p+j-1}, \cdots, \theta_{n-m}^{n-m-p+j-1}\right]^t.$$

$$\mathbf{D}_i = \mathrm{diag}\left[\log\left(\frac{\theta_{n-m+i}}{\theta_1}\right), \log\left(\frac{\theta_{n-m+i}}{\theta_2}\right), \cdots, \log\left(\frac{\theta_{n-m+i}}{\theta_{n-m}}\right)\right].$$

$$\mathbf{b}_i \triangleq \left[1, \theta_{n-m+i}, \cdots, \theta_{n-m+i}^{n-m-1}\right]^t.$$

Proposition [1] Let $q = \min(m, n-m)$, then for $1 \le p \le q$ we have

$$\lim_{s\to 0} \mathcal{M}_1(s-p+1) = L \sum_{j=1}^{p} \sum_{i=1}^{m} \mathcal{D}(i,j) \frac{(-1)^{p-j}}{(p-j)!} \mathbf{b}_i^t \boldsymbol{\Psi}^{-1} \mathbf{D}_i \mathbf{a}_j.$$

[1] K. Elkhalil *et al* "Analytical Derivation of the Inverse Moments of One-Sided Correlated Gram Matrices," *IEEE Transactions on Signal Processing*, 2016.

Elements of the Proof

- Some useful notations

$$\boldsymbol{\Psi}_s \triangleq \begin{bmatrix} \theta_1^s & \theta_1^{1+s} & \cdots & \theta_1^{n-m+s-1} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{n-m}^s & \theta_{n-m}^{1+s} & \cdots & \theta_{n-m}^{n-m+s-1} \end{bmatrix}, \boldsymbol{\Psi}_s = \boldsymbol{\Psi} + o(s).$$

$$\mathbf{a}_{s,j} \triangleq \left[\theta_1^{n-m+s-p+j-1}, \theta_2^{n-m+s-p+j-1}, \cdots, \theta_{n-m}^{n-m+s-p+j-1}\right]^t.$$

$$\mathbf{b}_{s,i} \triangleq \left[\theta_{n-m+i}^s, \theta_{n-m+i}^{1+s}, \cdots, \theta_{n-m+i}^{n-m+s-1}\right]^t.$$

$$\mathbf{b}_i \triangleq \left[1, \theta_{n-m+i}, \cdots, \theta_{n-m+i}^{n-m-1}\right]^t.$$

$$\mathbf{e}_k \triangleq \left[\mathbf{0}_{1\times(n-m-k-1)}, 1, \mathbf{0}_{1\times k}\right]^t,$$

$$k = 0, \cdots, n-m-1,$$

Main idea: It starts with a simple observation

$$\boldsymbol{\Psi}_s \triangleq \begin{bmatrix} \theta_1^s & \theta_1^{1+s} & \cdots & \theta_1^{n-m+s-p+j-1} & \theta_1^{n-m+s-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \theta_{n-m}^s & \theta_{n-m}^{1+s} & \cdots & \theta_{n-m}^{n-m+s-p+j-1} & \theta_{n-m}^{n-m+s-1} \end{bmatrix},$$

$$\mathbf{a}_{s,j} \triangleq \left[\theta_1^{n-m+s-p+j-1}, \theta_2^{n-m+s-p+j-1}, \cdots, \theta_{n-m}^{n-m+s-p+j-1}\right]^t.$$

$$\mathbf{b}_{s,i} \triangleq \left[\theta_{n-m+i}^s, \theta_{n-m+i}^{1+s}, \cdots, \theta_{n-m+i}^{n-m+s-1}\right]^t.$$

- $\boldsymbol{\Psi}_s \mathbf{e}_{p-j} = \mathbf{a}_{s,j}$ and $\mathbf{b}_{s,i}^t \mathbf{e}_{p-j} = \theta_{n-m+i}^{n-m+s-p+j-1} \Longrightarrow \boldsymbol{\Psi}_s^{-1} \mathbf{a}_{s,j} = \mathbf{e}_{p-j}$ and consequently $\mathbf{b}_{s,i}^t \boldsymbol{\Psi}_s^{-1} \mathbf{a}_{s,j} = \theta_{n-m+i}^{n-m+s-p+j-1}$.

Therefore,

$$\theta_{n-m+i}^{n-m+s-p+j-1} - \mathbf{b}_{s,i}^t \boldsymbol{\Psi}_s^{-1} \mathbf{a}_{s,j} = 0.$$

$$\mathcal{M}_1(s-p+1) = L\sum_{j=1}^{p}\sum_{i=1}^{m}\mathcal{D}(i,j)\,\Gamma(s-p+j)\left(\theta_{n-m+i}^{n-m+s-p+j-1} - \sum_{l=1}^{n-m}\sum_{k=1}^{n-m}\left[\boldsymbol{\Psi}^{-1}\right]_{k,l}\theta_l^{n-m+s-p+j-1}\theta_{n-m+i}^{k-1}\right)$$
$$= L\sum_{j=1}^{p}\sum_{i=1}^{m}\mathcal{D}(i,j)\,\Gamma(s-p+j)\left(\theta_{n-m+i}^{n-m+s-p+j-1} - \mathbf{b}_i^t\boldsymbol{\Psi}^{-1}\mathbf{a}_{s,j}\right)$$

Substract and add $\mathbf{b}_i^t\boldsymbol{\Psi}_s^{-1}\mathbf{a}_{s,j}$

$$= L\sum_{j=1}^{p}\sum_{i=1}^{m}\mathcal{D}(i,j)\,\Gamma(s-p+j)\left(\theta_{n-m+i}^{n-m+s-p+j-1} - \mathbf{b}_i^t\boldsymbol{\Psi}_s^{-1}\mathbf{a}_{s,j}\right) + L\sum_{j=1}^{p}\sum_{i=1}^{m}\mathcal{D}(i,j)\,\Gamma(s-p+j)$$
$$\times\,\mathbf{b}_i^t\left(\boldsymbol{\Psi}_s^{-1} - \boldsymbol{\Psi}^{-1}\right)\mathbf{a}_{s,j}$$

Substract and add $\mathbf{b}_{s,i}^t\boldsymbol{\Psi}_s^{-1}\mathbf{a}_{s,j}$

$$= L\sum_{j=1}^{p}\sum_{i=1}^{m}\mathcal{D}(i,j)\,\Gamma(s-p+j)\left(\theta_{n-m+i}^{n-m+s-p+j-1} - \mathbf{b}_{s,i}^t\boldsymbol{\Psi}_s^{-1}\mathbf{a}_{s,j}\right) + L\sum_{j=1}^{p}\sum_{i=1}^{m}\mathcal{D}(i,j)\,\Gamma(s-p+j)$$
$$\times\,\underbrace{\left(\mathbf{b}_{s,i}^t - \mathbf{b}_i^t\right)}\boldsymbol{\Psi}_s^{-1}\mathbf{a}_{s,j} + L\sum_{j=1}^{p}\sum_{i=1}^{m}\mathcal{D}(i,j)\,\Gamma(s-p+j)\,\mathbf{b}_i^t\underbrace{\left(\boldsymbol{\Psi}_s^{-1} - \boldsymbol{\Psi}^{-1}\right)}\mathbf{a}_{s,j}$$

- $\theta^s - 1 \underset{s\to 0}{=} s\log\theta + o(s)$.

$$\mathbf{b}_{s,i} - \mathbf{b}_i = s\left[\log\left(\theta_{n-m+i}\right), \theta_{n-m+i}\log\left(\theta_{n-m+i}\right), \cdots, \theta_{n-m+i}^{n-m-1}\log\left(\theta_{n-m+i}\right)\right]^t + o\left(s\right)$$
$$= s\log\left(\theta_{n-m+i}\right)\mathbf{b}_i + o\left(s\right),$$

- For non positive arguments $-k,\ k = 0, 1, \cdots,$

$$\Gamma(s-p+j) \underset{s\to 0}{=} \frac{(-1)^{p-j}}{s(p-j)!} + o(s).$$

-

$$\Gamma\left(s-p+j\right)\left(\mathbf{b}_{s,i}^t - \mathbf{b}_i^t\right)\boldsymbol{\Psi}_s^{-1}\mathbf{a}_{s,j}$$
$$= \frac{(-1)^{p-j}\log\left(\theta_{n-m+i}\right)}{(p-j)!}\mathbf{b}_i^t\boldsymbol{\Psi}^{-1}\mathbf{a}_j + o(s)$$

- The resolvent identity

$$\boldsymbol{\Psi}_s^{-1} - \boldsymbol{\Psi}^{-1} = \boldsymbol{\Psi}_s^{-1}\left(\boldsymbol{\Psi} - \boldsymbol{\Psi}_s\right)\boldsymbol{\Psi}^{-1}.$$

- Finally,

$$\left(\boldsymbol{\Psi} - \boldsymbol{\Psi}_s\right) \underset{s\to 0}{=} -s\boldsymbol{\Phi}\boldsymbol{\Psi} + o(s),\ \ \boldsymbol{\Phi} = \operatorname{diag}\left(\log\theta_i, i = 1, \cdots n-m\right).$$

Possible Applications

- Performance of linear estimators such as the LS (correlated channel), or the best linear unbiased estimator (BLUE) when the noise in correlated.

$$\mathbf{y} = \mathbf{H}\mathbf{v} + \mathbf{z}, \tag{4}$$

$$\begin{aligned} \hat{\mathbf{v}}_{blue} &= \left(\mathbf{H}^* \boldsymbol{\Sigma}_z^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^* \boldsymbol{\Sigma}_z^{-1} \mathbf{y} \\ &= \mathbf{v} + \left(\mathbf{H}^* \boldsymbol{\Sigma}_z^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^* \boldsymbol{\Sigma}_z^{-1} \mathbf{z} \\ &= \mathbf{v} + \mathbf{e}_{blue}, \end{aligned} \tag{5}$$

$$\begin{aligned} \mathbb{E}_{\mathbf{H}}\{\|\hat{\mathbf{v}}_{blue} - \mathbf{v}\|^2\} &= \mathbb{E}_{\mathbf{H}} \mathrm{trace}\left(\boldsymbol{\Sigma}_{e,blue}\right) \\ &= \mathbb{E}_{\mathbf{H}} \mathrm{trace}\left(\left(\mathbf{H}^* \boldsymbol{\Sigma}_z^{-1} \mathbf{H}\right)^{-1}\right) \\ &= m\mu_{\boldsymbol{\Lambda}}(-1), \end{aligned} \tag{6}$$

where $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}_z^{-1}$.

- Approximation of the LMMSE error

$$\mathbb{E}_{\mathbf{H}}\{\|\hat{\mathbf{x}}_{lmmse} - \mathbf{x}\|^2\} = \mathbb{E}_{\mathbf{H}}\text{trace}\left(\mathbf{\Sigma}_{e,lmmse}\right) = \mathbb{E}_{\mathbf{H}}\text{trace}\left(\mathbf{\Sigma}_x^{-1} + \mathbf{H}^*\mathbf{\Sigma}_z^{-1}\mathbf{H}\right)^{-1}. \quad (7)$$

Theorem

Let $\mathbf{\Lambda} = \mathbf{\Sigma}_z^{-1}$ and $\mathbf{\Sigma}_x = \sigma_x^2\mathbf{I}$. Then, the LMMSE average estimation error at both the high SNR regime ($\sigma_x^2 \gg 1$) and the low SNR regime ($\sigma_x^2 \ll 1$) is given by

1. High SNR regime

$$\mathbb{E}_{\mathbf{H}}\{\|\hat{\mathbf{x}}_{lmmse} - \mathbf{x}\|^2\} = m\sum_{k=0}^{p}\frac{(-1)^k}{\sigma_x^{2k}}\mu_{\mathbf{\Lambda}}(-k-1) + o\left(\sigma_x^{-2p}\right), \quad (8)$$

where $p \le q - 1$ with $q = \min(m, n-m)$.

2. Low SNR regime

$$\mathbb{E}_{\mathbf{H}}\{\|\hat{\mathbf{x}}_{lmmse} - \mathbf{x}\|^2\} = m\sum_{k=0}^{\infty}(-1)^k\sigma_x^{2k+2}\mu_{\mathbf{\Lambda}}(k). \quad (9)$$

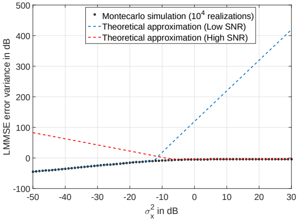


Figure: LMMSE mean square error with $\mathbf{\Sigma}_z$ modeled as Bessel correlation matrix: Montecarlo simulation versus theoretical approximations.

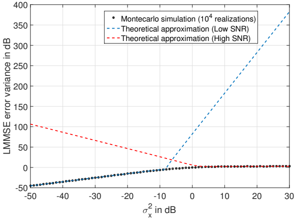


Figure: LMMSE mean square error with $\mathbf{\Sigma}_z$ modeled as random correlation matrix: Montecarlo simulation versus theoretical approximations.

- Sample Correlation Matrix (SCM)

$$\mathbf{R} = \mathbb{E}\left[\mathbf{u}\mathbf{u}^*\right],$$

where $\mathbf{u}(k) = \mathbf{R}^{\frac{1}{2}}\mathbf{x}(k)$ and $\mathbf{x}(k) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_m)$.
To estimate $\mathbf{R}$, we refer to the SCM as

$$\hat{\mathbf{R}}(n) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{u}(k)\mathbf{u}^*(k)$$

*The SCM is the maximum likelihood (ML) estimator of $\mathbf{R}$.
* For fixed m and large n

$$\left\|\mathbf{R} - \hat{\mathbf{R}}(n)\right\| \to 0, \quad a.s.$$

The SCM may not be robust for finite dimensions !
Exponentially weighted SCM

$$\hat{\mathbf{R}}(n) = (1-\lambda)\sum_{k=1}^{n} \lambda^{n-k}\mathbf{u}(k)\mathbf{u}^*(k)$$

$$= \mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{\Lambda}(n)\mathbf{X}^*\mathbf{R}^{\frac{1}{2}},$$

where $\mathbf{\Lambda}(n) = (1-\lambda)\mathrm{diag}\left(\lambda^{n-1}, \lambda^{n-2}, \cdots, 1\right)$.

$$Loss(n) \triangleq \mathbb{E}\left\|\mathbf{R}^{\frac{1}{2}}\hat{\mathbf{R}}^{-1}(n)\mathbf{R}^{\frac{1}{2}} - \mathbf{I}_m\right\|_{Fro}^2.$$

Lemma

Let $\mathbf{S}_n = \mathbf{X}\mathbf{\Lambda}(n)\mathbf{X}^*$. *Then, the loss can be expressed as a function of inverse moments of* $\mathbf{S}_n$ *as*

$$Loss(n) = m\left(1 + \mu_{\mathbf{\Lambda}(n)}(-2) - 2\mu_{\mathbf{\Lambda}(n)}(-1)\right). \qquad (10)$$

How to choose λ ?

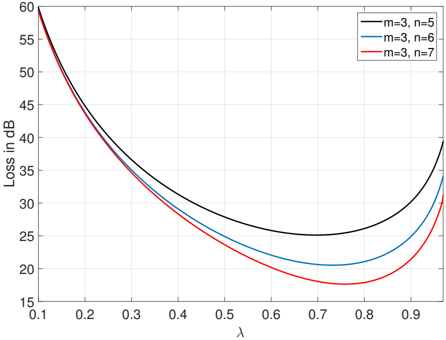


Figure: The estimation loss as a function of the forgetting factor λ.

Large RMT (Asymptotic Approach)

The Stieltjes Transform (ST)

For a probability measure $\mathbb{P}$, the associated ST is defined as

$$\mathcal{S}_{\mathbb{P}}(z) = \int_{\mathbb{R}} \frac{\mathbb{P}(d\lambda)}{\lambda - z}, \quad z \in \mathbb{C}^{+}$$

Spectral Measure

$$\mathcal{E}(\mathbf{A}) = \frac{1}{m} \sum_i \delta_{\lambda_i}.$$

Then,

$$\mathcal{S}_{\mathcal{E}}(z) = \frac{1}{m} \sum_i \frac{1}{\lambda_i - z} = \frac{1}{m} \text{trace}\left(\mathbf{A} - z\mathbf{I}\right)^{-1}.$$

It follows that,

$$\left(\frac{\partial^k \mathcal{S}_{\mathcal{E}}(z)}{\partial z^k}\right)_{z=0} = \frac{k!}{m} \text{tr}\left(\mathbf{A}^{-(k+1)}\right).$$

Theorem (Silverstein and Bai) Consider the Gram matrix $\mathbf{S} = \frac{1}{m}\mathbf{X}^*\mathbf{\Lambda}\mathbf{X}$ with ST $s(z)$.

- The entries of $\mathbf{X}$ are Gaussian i.i.d with zero mean and unit variance.
- $m, n \to \infty$ with $\frac{m}{n} \to c \in (0, \infty)$.
- $\mathbf{\Lambda}$ is nonnegative definite matrix with

$$\mathcal{E}(\mathbf{\Lambda}) \xrightarrow{n\to\infty} dH(\tau).$$

Then, the ST of $\mathbf{S}$, $s(z)$ satisfies

$$s(z) = \left(-z + c\int \frac{\tau dH(\tau)}{1+\tau s(z)}\right)^{-1} \quad (11)$$

$$\approx \left(-z + c.\text{trace}\left[\mathbf{\Lambda}\left(\mathbf{I} + s(z)\mathbf{\Lambda}\right)^{-1}\right]\right)^{-1} \quad (12)$$

$$= \frac{1}{-z + \frac{1}{m}\sum_{k=1}^{n}\frac{\lambda_k}{1+\lambda_k s(z)}}. \quad (13)$$

$$s_0^{(k)} = \left(\frac{\partial^k s(z)}{\partial z^k}\right)_{z=0} = \frac{k!}{m}\text{tr}\left(\mathbf{S}^{-(k+1)}\right).$$

How to obtain the higher order moments ?

Higher order moments

Let $p \geq 1$ and $f_k(z) = -\frac{1}{1+[\mathbf{D}]_{k,k} s(z)}$. Denote by $f_k^{(p)}$ the p-th derivative of $f_k(z)$ at $z = 0$. Then, the following relations hold true:

$$ps_0^{(p-1)} + \frac{s_0^{(p)}}{m} \sum_{k=1}^{n} \frac{[\mathbf{D}]_{k,k} f_k^{(0)}}{1 + [\mathbf{D}]_{k,k} s(0)} + \frac{1}{m} \sum_{k=1}^{n} \sum_{l=1}^{p-1} \binom{p}{l} \frac{[\mathbf{D}]_{k,k} s_0^{(l)} f_k^{(p-l)}}{1 + [\mathbf{D}]_{k,k} s(0)} = 0, \quad (14)$$

$$f_k^{(p)} + \frac{[\mathbf{D}]_{k,k} s_0^{(p)} f_k^{(0)}}{1 + [\mathbf{D}]_{k,k} s(0)} + \sum_{l=1}^{p-1} \binom{p}{l} \frac{[\mathbf{D}]_{k,k} s_0^{(l)} f_k^{(p-l)}}{1 + [\mathbf{D}]_{k,k} s(0)} = 0. \quad (15)$$

Algorithm 1 Asymptotic inverse moments computation

1: Compute $s(0)$ using (22)
2: Compute $f_k(0) = -\frac{1}{1+[\mathbf{D}]_{k,k} s(0)}$
3: **for** $i = 1 \to p$ **do**
4: compute $s_0^{(i)}$ using (14)
5: compute $f_k^{(i)}$ using (15)

Numerical validation

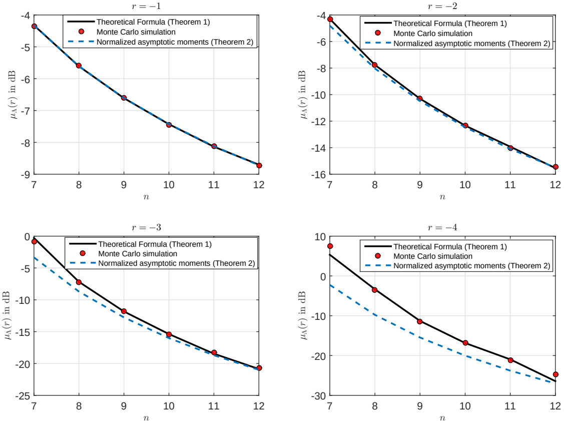


Figure: Inverse moments for $\boldsymbol{\Lambda}$ modeled as Bessel Correlation matrix: A comparison between theoretical result (Theorem 1), normalized asymptotic moments (Theroem 2) and Monte Carlo simulations (10^5 realizations).

Consider

$$\mathbf{y} = \mathbf{Hx} + \mathbf{z}, \ \ \mathbf{z} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{R}\right).$$

The covariance of the estimation error vector is given by

$$\boldsymbol{\Sigma} = \left(\mathbf{H}^{*}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}.$$

Popular measures for the quality of estimation

- Mean square error

$$\mathrm{MSE} = \mathrm{trace}\left(\boldsymbol{\Sigma}\right), \ \mathbf{f}(\mathbf{x}) = \frac{\mathbf{1}}{\mathbf{x}}.$$

- The log volume of the concentration ellipsoid (VCE)

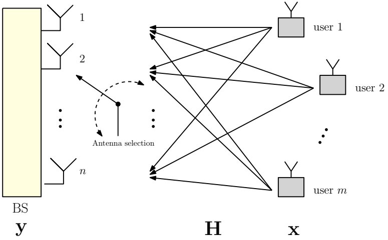
$$\mathrm{VCE} \propto -\log\det\left(\boldsymbol{\Sigma}\right), \ \mathbf{f}(\mathbf{x}) = -\log\mathbf{x}.$$

- Worst case error variance (WEV)

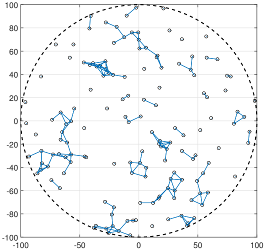
$$WEV = \max_{\|\mathbf{q}\|=1} \mathbf{q}^{T}\boldsymbol{\Sigma}\mathbf{q} = \max_{i} \lambda_i\left(\boldsymbol{\Sigma}\right) = \lambda_{max}\left(\boldsymbol{\Sigma}\right),$$

Measurement selection

- Antenna selection in massive MIMO systems.



- Sensor selection.



- ...

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}.$$

$$\mathbf{y}_{\mathcal{S}} = \mathbf{S}\mathbf{y}, \tag{16}$$

where $\mathbf{S}$ is defined as follows

$$[\mathbf{S}]_{i,j} = \begin{cases} 1 & j = \mathcal{S}[i] \\ 0 & \text{otherwise} \end{cases}, i = 1, \cdots, k. \tag{17}$$

Ee have the following properties

- $\mathbf{S}\mathbf{S}^T = \mathbf{I}_k$.
- $\mathbf{S}^T\mathbf{S} = \mathrm{diag}(\mathbf{s})$.

where $\mathbf{s} = \{s_i\}_{i=1,\cdots,n}$.
The resultant error covariance matrix, which we denote by $\mathbf{\Sigma}_{\mathcal{S}}$, easily writes as

$$\begin{aligned} \mathbf{\Sigma}_{\mathcal{S}} &= \left(\mathbf{H}^H\mathbf{S}^T\left(\mathbf{S}\mathbf{R}\mathbf{S}^T\right)^{-1}\mathbf{S}\mathbf{H}\right)^{-1} \\ &= \left(\mathbf{W}^H\mathbf{R}^{\frac{1}{2}}\mathrm{diag}(\mathbf{s})\,\mathbf{R}^{\frac{1}{2}}\mathbf{W}\right)^{-1}. \end{aligned} \tag{18}$$

How to choose $\mathbf{s}$?

Mathematically speaking, a general formulation of the selection problem can be illustrated as follows

$$\begin{aligned} \mathcal{S}^* &= \underset{\mathcal{S}}{\operatorname{argmin}} \quad f\left(\boldsymbol{\Sigma}_{\mathcal{S}}\right) \\ \text{s.t.} &\quad \mathcal{S} \subseteq \{1, \cdots, n\} \\ &\quad |\mathcal{S}| = k. \end{aligned} \tag{19}$$

where f = MSE, VCE, WEV.

→ NP hard problem (we don't know an algorithm that can solve the problem in a polynomial time).

Convex Relaxation (Joshi and Boyd)

$$\begin{aligned} \widehat{\mathbf{s}} &= \underset{\mathbf{s}}{\operatorname{argmin}} \quad f\left(\mathbf{s}\right) \\ \text{s.t.} &\quad \mathbf{1}^T \mathbf{s} = k \\ &\quad 0 \leq s_i \leq 1, i = 1, \cdots, n. \end{aligned} \tag{20}$$

→ polynomial complexity of $\mathcal{O}\left(n^3\right)$

What if the channel **H** is time-varying ? Best we can do is $N\mathcal{O}\left(n^3\right)$ so far!

Blind Measurement Selection via Large RMT

idea: Solve an approximate problem

We solve the following optimization problem

$$\begin{aligned} \widehat{\mathcal{S}} = \underset{\mathcal{S}}{\operatorname{argmin}} \quad & \overline{f\left(\mathbf{\Sigma}_{\mathcal{S}}\right)} \\ \text{s.t.} \quad & \mathcal{S} \subseteq \{1, \cdots, n\} \\ & |\mathcal{S}| = k. \end{aligned} \tag{21}$$

where $\overline{f\left(\mathbf{\Sigma}_{\mathcal{S}}\right)}$ is an approximation of $f\left(\mathbf{\Sigma}_{\mathcal{S}}\right)$ (Deterministic equivalent, DE).

Some useful DEs

- MSE

$$s(0) = \frac{1}{\frac{1}{m}\sum_{k=1}^{n}\frac{\lambda_k}{1+\lambda_k s(0)}}. \tag{22}$$

- VCE

$$\overline{\mathcal{I}} = \log\det\left(\mathbf{I}_n + \frac{c}{\overline{\delta}}\mathbf{D}\right) + m\log\left(\overline{\delta}\right) - nc, \quad \overline{\delta} = \frac{1}{n}\operatorname{tr}\left[\mathbf{D}\left(\mathbf{I}_n + \frac{c}{\overline{\delta}}\mathbf{D}\right)^{-1}\right].$$

$$\mathbf{D} = \operatorname{diag}\left(\theta_1, \cdots, \theta_n\right).$$

- WEV

$$\overline{\lambda} = \left(-\frac{1}{\underline{m}} + \int \frac{t}{1+c\underline{m}t}\nu\left(t\right)dt\right), \quad \underline{m}^2 = \frac{1}{c}\left(\int \frac{t^2}{\left(1+c\underline{m}t\right)^2}\nu\left(t\right)dt\right)^{-1}.$$

Low complexity search algorithm (Greedy approach)

Algorithm 2 Greedy Approach for Antenna Selection

```
Initialize S = randsample(n, k)                ▷ randomly generate a pattern of size k
Compute metric* = f(H, S)
for i = 1 → #iterations do
    S̄ = {1, ⋯, n} \ S
    j ← 1
    while j ≤ n − k do
        p ← S̄[j]
        I ← S
        table ← zeros(k, 1)
         for l = 1 → k do
            I[l] ← p
            table[l] ← f(H, I)
            I ← S
         if min(table) < metric* then
            metric* ← min(table)
            S[arg min(table)] ← p
```

Complexity of $\mathcal{O}\left(n^2\right)$.

Performance

- Antennas selection in Massive MIMO: $\mathbf{R}_{i,j} = \lambda^{|i-j|}$.

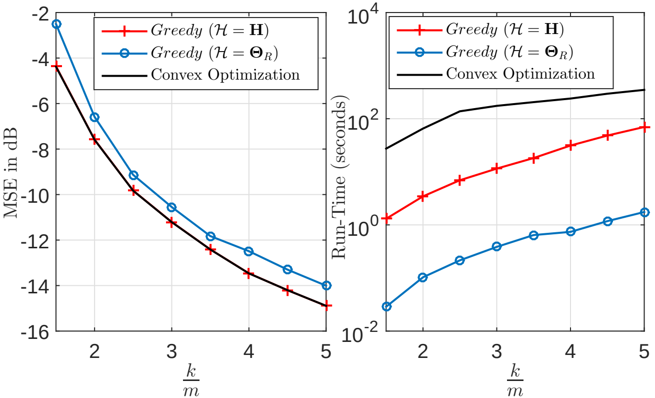
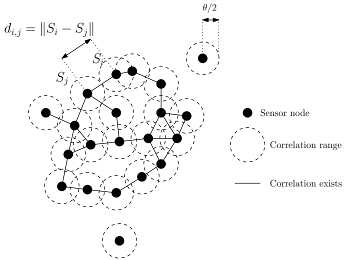
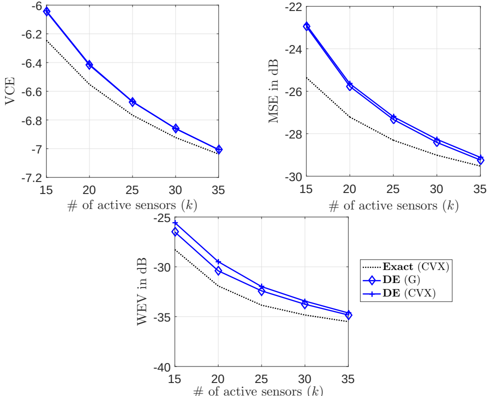


Figure: Average MSE (over $N = 100$ channel realizations) and Run-time vs. the ratio $\frac{k}{m}$ for $\mathbf{\Theta}_R$ = Toeplitz ($\lambda = 0.75$). $m = 10$ users, $n = 1.25k$ and $\rho = 10$ dB.

- Sensor selection:



$$
\mathbf{R}_{i,j} = \begin{cases} \dfrac{\cos^{-1}\left(\dfrac{d_{i,j}}{\theta}\right)}{\pi} - \dfrac{d_{i,j}}{\pi\theta^2}\sqrt{\theta^2 - d_{i,j}^2}, & 0 \leq d_{i,j} \leq \theta \\ 0 & d_{i,j} > \theta . \end{cases}
$$



- Exact derivation of the inverse moments of Gram matrices with one-sided correlation.
- Correlation can be exploited to figure out the potential measurements in a linear system.
- Good performances in some practical scenarios (massive MIMO and WSN).
- The case of identical eigenvalues is still an open question.

Thank you for your attention