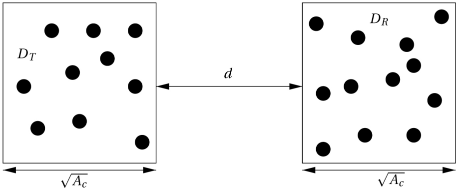
Spectral decay of finite Fourier transforms and related random matrices.

Aline Bonami,
Université d'Orléans, France

joint work with Abderrazek Karoui
Université de Carthage, Tunisia

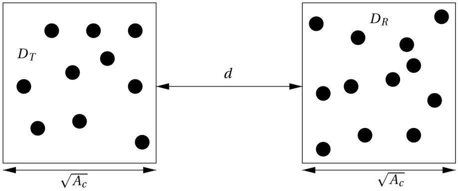
Transmission in a wireless MIMO network.

Considered by Desgroseilliers, Lévêque and Preissmann, EPFL (2013) :



Transmission in a wireless MIMO network.

Considered by Desgroseilliers, Lévêque and Preissmann, EPFL (2013) :



Matrix used to describe the *phase fading* between the emission and the reception :

$$a_{jk} = \frac{e^{2i\pi r_{jk}/\lambda}}{r_{jk}}.$$

Here r_{jk} is the distance between the nodes j and k, which are chosen randomly in two squares and at large distance.

Their questions on the channel fading matrix.

$$a_{jk} = \frac{e^{2i\pi r_{jk}/\lambda}}{r_{jk}}.$$

Here r_{jk} is the distance between the nodes j and k, which are in two squares and at large distance. We assume that the number of transmisssion nodes and reception nodes is the same, n.

Their questions on the channel fading matrix.

$$a_{jk} = \frac{e^{2i\pi r_{jk}/\lambda}}{r_{jk}}.$$

Here r_{jk} is the distance between the nodes j and k, which are in two squares and at large distance. We assume that the number of transmisssion nodes and reception nodes is the same, n.

- Number of degrees of freedom, that is, the number of significative singular values.
- For a value p, which is the total power of the input signal, uniformly distributed among the n nodes, want to have an approximation of the Shannon capacity of the system, given by

$$C(p) = \log_2(\det(I + \frac{p}{n} A^* A)) := \sum_j \log_2(1 + \frac{p}{n}\lambda_j(A)^2).$$

Their questions on the channel fading matrix.

$$a_{jk} = \frac{e^{2i\pi r_{jk}/\lambda}}{r_{jk}}.$$

Here r_{jk} is the distance between the nodes j and k, which are in two squares and at large distance. We assume that the number of transmisssion nodes and reception nodes is the same, n.

- Number of degrees of freedom, that is, the number of significative singular values.
- For a value p, which is the total power of the input signal, uniformly distributed among the n nodes, want to have an approximation of the Shannon capacity of the system, given by

$$C(p) = \log_2(\det(I + \frac{p}{n} A^* A)) := \sum_j \log_2(1 + \frac{p}{n}\lambda_j(A)^2).$$

Nodes are chosen randomly. We want to know this with large probability.

The mathematical question.

After approximation, simplification and normalization, the same authors reduce to the study of $n \times n$ matrices

$$A := \frac{\sqrt{m}}{n} \begin{pmatrix} e^{2i\pi m Z_1 Y_1} & \dots & e^{2i\pi m Z_1 Y_n} \\ \vdots & & \vdots \\ e^{2i\pi m Z_n Y_1} & \dots & e^{2i\pi m Z_n Y_n} \end{pmatrix}.$$

Here Y_j, Z_k independent and uniformly distributed in $I := (-1/2, +1/2)$.

The mathematical question.

After approximation, simplification and normalization, the same authors reduce to the study of $n \times n$ matrices

$$A := \frac{\sqrt{m}}{n} \begin{pmatrix} e^{2i\pi m Z_1 Y_1} & \dots & e^{2i\pi m Z_1 Y_n} \\ \vdots & & \vdots \\ e^{2i\pi m Z_n Y_1} & \dots & e^{2i\pi m Z_n Y_n} \end{pmatrix}.$$

Here Y_j, Z_k independent and uniformly distributed in $I := (-1/2, +1/2)$.
So A is a random matrix, with alea given by the Z_j's and alea given by the Y_k's.

By assumption $m \ll n$ is large.
Describe the singular values of A (i.e. the spectrum of A^*A or AA^*) for m large, m/n small.
Typically, $m \approx n^\delta$, with $1/2 \leq \delta < 1$.

Finite Fourier Transform.

Kind of discretization of the operator on $L^2(I)$ given by

$$\mathcal{F}_m(f)(z) := \sqrt{m} \int_{-1/2}^{+1/2} e^{2i\pi mzy} f(y)dy.$$

Finite Fourier Transform.

Kind of discretization of the operator on $L^2(I)$ given by

$$\mathcal{F}_m(f)(z) := \sqrt{m}\int_{-1/2}^{+1/2} e^{2i\pi mzy} f(y)dy.$$

The compact operator $\mathcal{F}_m^*\mathcal{F}_m$ is given by

$$\mathcal{Q}_m(f)(x) = \mathcal{F}_m^*\mathcal{F}_m(f)(x) = \int_{-1/2}^{+1/2} \frac{\sin m\pi(x-y)}{\pi(x-y)} f(y)dy.$$

$\mathcal{Q}_m$ is the *time frequency limiting operator.*
$\frac{\sin \pi x}{\pi x}$ *is known as the sinc kernel.*

Eigenvalues of $\mathcal{Q}_m$ are ordered in a decreasing sequence, starting from 0.

The main theorem for random matrices.

Theorem. The singular values of A are close to the singular values of $\mathcal{F}_m$. More precisely, with probability $1 - 4e^2 e^{-\xi^2/2}$,

$$\left(\sum_{j \geq 0} (\lambda_j(A)^2 - \lambda_j(\mathcal{F}_m)^2)^2 \right)^{1/2} \leq \frac{(2\xi + 1.5)m}{\sqrt{n}}.$$

$\sum_j \lambda_j(A)^2 = \sum_j \lambda_j(\mathcal{F}_m)^2 = m$. So the right hand side is an error if $\frac{m}{n}$ is small.

Notation : $\lambda(M) = \{\lambda_0(M), \cdots \lambda_j(M), \cdots\}$ is the spectrum (or sequence of singular values) of a symmetric positive semi-definite matrix (or symmetric positive semi-definite operator), in decreasing order.

The asymptotic behaviour of $\lambda_j(\mathcal{Q}_m)$.

(Landau and Widom 1980) m tending to ∞ :
$N(\alpha) = \#\{\lambda_j(\mathcal{Q}_m);\ \lambda_j(\mathcal{Q}_m) > \alpha\}$ is such that

$$N(\alpha) = m + \left[\frac{1}{\pi^2}\log\left(\frac{1-\alpha}{\alpha}\right)\right]\log(m) + o(\log(m)).$$

The asymptotic behaviour of $\lambda_j(\mathcal{Q}_m)$.

(Landau and Widom 1980) m tending to ∞ :
$N(\alpha) = \#\{\lambda_j(\mathcal{Q}_m);\ \lambda_j(\mathcal{Q}_m) > \alpha\}$ is such that

$$N(\alpha) = m + \left[\frac{1}{\pi^2} \log\left(\frac{1-\alpha}{\alpha}\right)\right] \log(m) + o(\log(m)).$$

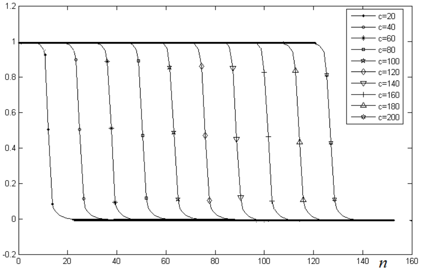


FIGURE : Graph of the $\lambda_j(\mathcal{Q}_m)$ for different values of $m = \frac{2c}{\pi}$

Non asymptotic results for $\lambda_j(\mathcal{Q}_m)$.

Landau (1991) : $\lambda_{\lceil m \rceil} \leq 1/2 \leq \lambda_{[m]-1}$.

Non asymptotic results for $\lambda_j(\mathcal{Q}_m)$.

Landau (1991) : $\lambda_{\lceil m \rceil} \leq 1/2 \leq \lambda_{[m]-1}$.

A lucky incident (Slepian et al, Bell labs in the sixties) : The eigenfunctions $\psi_{s,m}$ (which are called Prolate Spheroidal Wave Functions) are also eigenfunctions of an explicit Sturm-Liouville operator on I

$$\mathcal{L}_m \psi := -\frac{d}{dx}\left[(1-4x^2)\psi'\right] + \pi^2 m^2 x^2 \psi.$$

Can be used to study the behavior of the eigenvalues of $\mathcal{Q}_m$. Many recent works (Osipov, Rokhlin-Xiao, Bonami-Karoui, ...).

Non asymptotic results for $\lambda_j(\mathcal{Q}_m)$.

Landau (1991) : $\lambda_{\lceil m \rceil} \leq 1/2 \leq \lambda_{[m]-1}$.

A lucky incident (Slepian et al, Bell labs in the sixties) : The eigenfunctions $\psi_{s,m}$ (which are called Prolate Spheroidal Wave Functions) are also eigenfunctions of an explicit Sturm-Liouville operator on I

$$\mathcal{L}_m \psi := -\frac{d}{dx}\left[(1-4x^2)\psi'\right] + \pi^2 m^2 x^2 \psi.$$

Can be used to study the behavior of the eigenvalues of $\mathcal{Q}_m$. Many recent works (Osipov, Rokhlin-Xiao, Bonami-Karoui, ...).
In particular :

- ([BK2015]) for $\delta > 1$ there exists a, C, such that, for $s > \delta m$,

$$\lambda_s(\mathcal{Q}_m) \leq Ce^{-as},$$

- For $\delta > 1$ there exists a such that, for $s > m + \delta \log m$,

$$\lambda_s(\mathcal{Q}_m) \leq e^{-a\left(\frac{s-m}{\log m}\right)}$$

Non asymptotic results for $\lambda_s(\mathcal{Q}_m)$.

Landau (1991) : $\lambda_{\lceil m \rceil} \leq 1/2 \leq \lambda_{[m]-1}$.

A lucky incident (Slepian et al., Bell Lab in the sixties) : The eigenfunctions $\psi_{s,m}$ (which are called Prolate Spheroidal Wave Functions) are also eigenfunctions of an explicit Sturm-Liouville operator on I

$$\mathcal{L}_m \psi := -\frac{d}{dx}\left[(1-4x^2)\psi'\right] + \pi^2 m^2 x^2 \psi.$$

Can be used to study the behavior of the eigenvalues of $\mathcal{Q}_m$. Many recent works (Osipov, Rokhlin-Xiao, Bonami-Karoui, ...).
In particular :

- ([BK2015]) for $\delta > 1$ there exists a, C, such that, for $s > \delta m$,

$$\lambda_s(\mathcal{Q}_m) \leq Ce^{-as}, \qquad \sum_{j \geq s} \lambda_j(\mathcal{Q}_m) \leq C' e^{-as}.$$

- For $\delta > 1$ there exists a such that, for $s > m + \delta \log m$,

$$\lambda_s(\mathcal{Q}_m) \leq e^{-a\left(\frac{s-m}{\log m}\right)} \qquad \sum_{j \geq s} \lambda_j(\mathcal{Q}_m) \leq C'' \log m \, e^{-a\left(\frac{s-m}{\log m}\right)}.$$

Estimates obtained from min-max Principle.

Work in progress with A. Karoui and P. Jaming.

$$\lambda_n(\mathcal{Q}_m) = \min_{S_n} \max_{f \in S_n^\perp} \frac{\langle \mathcal{Q}_m f, f \rangle}{\|f\|^2_{L^2(I)}} = \min_{S_n} \max_{f \in S_n^\perp} \frac{\|\mathcal{F}_m f\|^2_{L^2(I)}}{\|f\|^2_{L^2(I)}},$$

where the S_n are $n+1-$dimensional subspaces of $L^2(I)$.

Estimates obtained from min-max Principle.

Work in progress with A. Karoui and P. Jaming.

$$\lambda_n(\mathcal{Q}_m) = \min_{S_n} \max_{f \in S_n^\perp} \frac{\langle \mathcal{Q}_m f, f \rangle}{\|f\|^2_{L^2(I)}} = \min_{S_n} \max_{f \in S_n^\perp} \frac{\|\mathcal{F}_m f\|^2_{L^2(I)}}{\|f\|^2_{L^2(I)}},$$

where the S_n are $n+1-$dimensional subspaces of $L^2(I)$.

- By taking for S_n the $n+1-$dimensional space generated by Legendre polynomials of degree $\leq n$, we find the estimate, valid for $n \geq \pi em/4$,

$$\lambda_n(m) \leq 2 \left(\frac{\pi em}{4(n+1)} \right)^{2n+1}.$$

Estimates obtained from min-max Principle.

Work in progress with A. Karoui and P. Jaming.

$$\lambda_n(\mathcal{Q}_m) = \min_{S_n} \max_{f \in S_n^\perp} \frac{\langle \mathcal{Q}_m f, f \rangle}{\|f\|^2_{L^2(I)}} = \min_{S_n} \max_{f \in S_n^\perp} \frac{\|\mathcal{F}_m f\|^2_{L^2(I)}}{\|f\|^2_{L^2(I)}},$$

where the S_n are $n+1-$dimensional subspaces of $L^2(I)$.

- By taking for S_n the $n+1-$dimensional space generated by Legendre polynomials of degree $\leq n$, we find the estimate, valid for $n \geq \pi e m/4$,

$$\lambda_n(m) \leq 2\left(\frac{\pi e m}{4(n+1)}\right)^{2n+1}.$$

- By taking the $1-$dimensional space generated by the projection of a Gaussian function, we find

$$\lambda_0(m) \geq 1 - e^{-m}.$$

Estimates obtained from min-max Principle.

Work in progress with A. Karoui and P. Jaming.

$$\lambda_n(\mathcal{Q}_m) = \min_{S_n} \max_{f \in S_n^\perp} \frac{\langle \mathcal{Q}_m f, f \rangle}{\|f\|^2_{L^2(I)}} = \min_{S_n} \max_{f \in S_n^\perp} \frac{\|\mathcal{F}_m f\|^2_{L^2(I)}}{\|f\|^2_{L^2(I)}},$$

where the S_n are $n+1-$dimensional subspaces of $L^2(I)$.

- By taking for S_n the $n+1-$dimensional space generated by Legendre polynomials of degree $\leq n$, we find the estimate, valid for $n \geq \pi e m/4$,

$$\lambda_n(m) \leq 2 \left(\frac{\pi e m}{4(n+1)} \right)^{2n+1}.$$

- By taking the $1-$dimensional space generated by the projection of a Gaussian function, we find

$$\lambda_0(m) \geq 1 - e^{-m}.$$

To be compared with asymptotic results of Widom and Fuchs.

Right plunge region and Remez Inequality.

Which $n+1-$dimensional space to be used for $m \approx n$?

$$\lambda_n(\mathcal{Q}_m) = \max_{S_n} \min_{f \neq 0, f \in S_n} \frac{\|\mathcal{F}_m f\|^2_{L^2(I)}}{\|f\|^2_{L^2(I)}}.$$

Right plunge region and Remez Inequality.

Which $n+1-$dimensional space to be used for $m \approx n$?

$$\lambda_n(\mathcal{Q}_m) = \max_{S_n} \min_{f \neq 0, f \in S_n} \frac{\|\mathcal{F}_m f\|^2_{L^2(I)}}{\|f\|^2_{L^2(I)}}.$$

Take for S_n the space of functions that may be written as a translate of $\sum_{j=0}^{n} a_j \varphi((n+1)x - k)$, with φ compactly supported in I.

To bound below the ratio, have to consider the Remez type best constant $\Gamma_2(n, \varepsilon)$ such that

$$\int_J |P(e^{it})|^2 \, \mathrm{d}t \geq \Gamma(n, \varepsilon) \int_0^{2\pi} |P(e^{it})|^2 \, \mathrm{d}t,$$

for any P is a Taylor polynomial of degree n and any J interval of length ε.

Right plunge region and Remez Inequality.

Which $n+1-$dimensional space to be used for $m \approx n$?

$$\lambda_n(\mathcal{Q}_m) = \max_{S_n} \min_{f \neq 0, f \in S_n} \frac{\|\mathcal{F}_m f\|^2_{L^2(I)}}{\|f\|^2_{L^2(I)}}.$$

Take for S_n the space of functions that may be written as a translate of $\sum_{j=0}^{n} a_j \varphi((n+1)x-k)$, with φ compactly supported in I.

To bound below the ratio, have to consider the Remez type best constant $\Gamma_2(n,\varepsilon)$ such that

$$\int_J |P(e^{it})|^2 \, dt \geq \Gamma(n,\varepsilon) \int_0^{2\pi} |P(e^{it})|^2 \, dt,$$

for any P is a Taylor polynomial of degree n and any J interval of length ε.

With $\varphi(x) = \mathbf{1}_{[-1/2,1/2]} \cos \pi x$, we find Landau's estimate $\lambda_{[m]-1} \geq 1/2$.

Right plunge region and Remez Inequality.

Which $n+1-$dimensional space to be used for $m \approx n$?

$$\lambda_n(\mathcal{Q}_m) = \max_{S_n} \min_{f\neq 0, f\in S_n} \frac{\|\mathcal{F}_m f\|^2_{L^2(I)}}{\|f\|^2_{L^2(I)}}.$$

Take for S_n the space of functions that may be written as a translate of $\sum_{j=0}^n a_j \varphi((n+1)x - k)$, with φ compactly supported in I.

To bound below the ratio, have to consider the Remez type best constant $\Gamma_2(n,\varepsilon)$ such that

$$\int_J |P(e^{it})|^2 \,\mathrm{d}t \geq \Gamma(n,\varepsilon) \int_0^{2\pi} |P(e^{it})|^2 \,\mathrm{d}t,$$

for any P is a Taylor polynomial of degree n and any J interval of length ε.

With $\varphi(x) = \mathbf{1}_{[-1/2,1/2]} \cos \pi x$, we find Landau's estimate $\lambda_{[m]-1} \geq 1/2$. Need only $\varepsilon = \pi$.

Gabriel Metsu


Gabriel Metsu

The main theorem for random matrices.

$$A := \frac{\sqrt{m}}{n} \begin{pmatrix} e^{2i\pi m Z_1 Y_1} & \dots & e^{2i\pi m Z_1 Y_n} \\ \vdots & & \vdots \\ e^{2i\pi m Z_n Y_1} & \dots & e^{2i\pi m Z_n Y_n} \end{pmatrix}.$$

with Y_j, Z_k independent and uniformly distributed in $(-1/2, +1/2)$.

The main theorem for random matrices.

$$A := \frac{\sqrt{m}}{n} \begin{pmatrix} e^{2i\pi m Z_1 Y_1} & \dots & e^{2i\pi m Z_1 Y_n} \\ \vdots & & \vdots \\ e^{2i\pi m Z_n Y_1} & \dots & e^{2i\pi m Z_n Y_n} \end{pmatrix}.$$

with Y_j, Z_k independent and uniformly distributed in $(-1/2, +1/2)$.

$\mathcal{F}_m$ is the operator on $L^2(I)$ given by

$$\mathcal{F}_m(f)(z) := \sqrt{m} \int_{-1/2}^{+1/2} e^{2i\pi m z y} f(y) dy.$$

The main theorem for random matrices.

$$A := \frac{\sqrt{m}}{n} \begin{pmatrix} e^{2i\pi m Z_1 Y_1} & \dots & e^{2i\pi m Z_1 Y_n} \\ \vdots & & \vdots \\ e^{2i\pi m Z_n Y_1} & \dots & e^{2i\pi m Z_n Y_n} \end{pmatrix}.$$

with Y_j, Z_k independent and uniformly distributed in $(-1/2, +1/2)$.
$\mathcal{F}_m$ is the operator on $L^2(I)$ given by

$$\mathcal{F}_m(f)(z) := \sqrt{m} \int_{-1/2}^{+1/2} e^{2i\pi mzy} f(y) dy.$$

Theorem. The singular values of A are close to the singular values of $\mathcal{F}_m$. More precisely, with probability $1 - 4e^2 e^{-\xi^2/2}$,

$$\left(\sum_{j \geq 0} (\lambda_j(A)^2 - \lambda_j(\mathcal{F}_m)^2)^2 \right)^{1/2} \leq \frac{(2\xi + 1.5)m}{\sqrt{n}}.$$

What is expected ?

$$A^*A = \frac{m}{n^2} \sum_{\ell=1}^{n} \left(e^{2i\pi m Z_\ell (Y_k - Y_j)}\right)_{j,k=1,\cdots n}.$$

Take the expectation in Z.

$$H := \mathbb{E}_Z(A^*A) = \frac{1}{n} \left(\frac{\sin(\pi m(Y_k - Y_j))}{\pi(Y_k - Y_j)}\right)_{j,k=1,\cdots n} := \frac{1}{n} \left(\kappa_m(Y_j, Y_k)\right)_{j,k=1,\cdots n}.$$

$\kappa_m(x, y) = m\,\mathrm{sinc}(m(x - y))$ is a positive definite kernel.

What is expected ?

$$A^*A = \frac{m}{n^2}\sum_{\ell=1}^{n}\left(e^{2i\pi m Z_\ell(Y_k - Y_j)}\right)_{j,k=1,\cdots n}.$$

Take the expectation in Z.

$$H := \mathbb{E}_Z(A^*A) = \frac{1}{n}\left(\frac{\sin(\pi m(Y_k - Y_j))}{\pi(Y_k - Y_j)}\right)_{j,k=1,\cdots n} := \frac{1}{n}\left(\kappa_m(Y_j, Y_k)\right)_{j,k=1,\cdots n}.$$

$\kappa_m(x,y) = m\,\mathrm{sinc}(m(x-y))$ is a positive definite kernel.
Expected : the spectrum of A^*A is close to the spectrum of its expectation in Z, which is close to the spectrum of the integral operator $\mathcal{Q}_m$, that is, the operator with kernel κ_m with large probability.

McDiarmid 's Vectorial Concentration Inequality.

Theorem [Hayes]. Let Φ be a measurable function on $\mathbb{R}^n$ with values in a Hilbert space $\mathcal{H}$. Assume that

$$\|\Phi(z_1, \cdots, z_\ell, \cdots, z_n) - \Phi(z_1, \cdots, z'_\ell, \cdots, z_n)\|_{\mathcal{H}} \leq c$$

for each sequence $(z_j)_{j \neq \ell}, z_\ell, z'_\ell$. Then, if $Z_1, \cdots Z_n$ are independent,

$$\mathbb{P}\left(\|\Phi(Z_1, \cdots, Z_n) - \mathbb{E}\Phi(Z_1, \cdots, Z_n)\|_{\mathcal{H}} > \xi\right) \leq 2e^2 e^{-\frac{\xi^2}{2c^2 n}}.$$

For $n = 1$, called Mc Diarmid's Inequality, or Bounded Difference Inequality (Azuma-Hoeffding Inequality).

McDiarmid 's Vectorial Concentration Inequality.

Theorem [Hayes]. Let Φ be a measurable function on $\mathbb{R}^n$ with values in a Hilbert space $\mathcal{H}$. Assume that

$$\|\Phi(z_1, \cdots, z_\ell, \cdots, z_n) - \Phi(z_1, \cdots, z'_\ell, \cdots, z_n)\|_{\mathcal{H}} \leq c$$

for each sequence $(z_j)_{j \neq \ell}, z_\ell, z'_\ell$. Then, if $Z_1, \cdots Z_n$ are independent,

$$\mathbb{P}\left(\|\Phi(Z_1, \cdots, Z_n) - \mathbb{E}\Phi(Z_1, \cdots, Z_n)\|_{\mathcal{H}} > \xi\right) \leq 2e^2 e^{-\frac{\xi^2}{2c^2 n}}.$$

For $n = 1$, called Mc Diarmid's Inequality, or Bounded Difference Inequality (Azuma-Hoeffding Inequality).

Corollary. $\mathbb{P}_Z\left(\|A^*A - H\|_{\mathrm{HS}} > \xi \frac{m}{\sqrt{n}}\right) \leq 2e^2 e^{-\frac{\xi^2}{2}}$.

The first step.

Equivalent to say that, with a probability larger than $1 - 2e^2 e^{-\frac{\xi^2}{2}}$,

$$\|A^*A - H\|_{\mathrm{HS}} \leq \frac{\xi m}{\sqrt{n}}.$$

Recall the notation : $\lambda(M) = \{\lambda_0(M), \cdots \lambda_{n-1}(M)\}$ is the spectrum of a positive definite matrix, in decreasing order.

The first step.

Equivalent to say that, with a probability larger than $1 - 2e^2 e^{-\frac{\xi^2}{2}}$,

$$\|A^*A - H\|_{\mathrm{HS}} \leq \frac{\xi m}{\sqrt{n}}.$$

Recall the notation : $\lambda(M) = \{\lambda_0(M), \cdots \lambda_{n-1}(M)\}$ is the spectrum of a positive definite matrix, in decreasing order.
By Lindskii Inequality, this implies that

$$\|\lambda(A^*A) - \lambda(H)\|_{\ell^2} \leq \frac{\xi m}{\sqrt{n}}.$$

We want the left hand side to be small compared to $\|H\|_{\mathrm{HS}}$.

The first step.

Equivalent to say that, with a probability larger than $1 - 2e^2 e^{-\frac{\xi^2}{2}}$,

$$\|A^*A - H\|_{\mathrm{HS}} \leq \frac{\xi m}{\sqrt{n}}.$$

Recall the notation : $\lambda(M) = \{\lambda_0(M), \cdots \lambda_{n-1}(M)\}$ is the spectrum of a positive definite matrix, in decreasing order.
By Lindskii Inequality, this implies that

$$\|\lambda(A^*A) - \lambda(H)\|_{\ell^2} \leq \frac{\xi m}{\sqrt{n}}.$$

We want the left hand side to be small compared to $\|H\|_{\mathrm{HS}}$. But

$$\mathbb{E}_Y(\|H\|_{\mathrm{HS}}^2) = m + O(\log m) + O\left(\frac{m^2}{n}\right).$$

The relative error will be small if m is large and m is small compared to n.

The second step.

More generally, let κ be a positive definite kernel on $(-1/2, +1/2)$ and assume that $\sup \kappa(x, x) \leq m$. Then, with probability larger than $1 - 2e^2 e^{-\frac{\xi^2}{2}}$, the spectrum of the random matrix

$$H_\kappa := n^{-1} \left(\kappa(Y_j, Y_k)\right)_{j,k=1,\cdots n}$$

satisfies the inequality

$$\|\lambda(H_\kappa) - \mathbb{E}(\lambda(H_\kappa)\|_{\ell^2} \leq \frac{\xi m}{\sqrt{n}}.$$

The second step.

More generally, let κ be a positive definite kernel on $(-1/2, +1/2)$ and assume that $\sup \kappa(x, x) \leq m$. Then, with probability larger than $1 - 2e^2 e^{-\frac{\xi^2}{2}}$, the spectrum of the random matrix

$$H_\kappa := n^{-1} \left(\kappa(Y_j, Y_k)\right)_{j,k=1,\cdots n}$$

satisfies the inequality

$$\|\lambda(H_\kappa) - \mathbb{E}(\lambda(H_\kappa)\|_{\ell^2} \leq \frac{\xi m}{\sqrt{n}}.$$

Remark. Estimates for one eigenvalue are given by Shawe-Taylor, Cristianini and Kandola (see also Blanchard Bousquet and Wald). Such a matrix is called a kernel matrix or a kernel Gram matrix. For machine learning non linear principal component analysis (i. e. eigenvectors of such matrices).

Comparison with the integral operator.

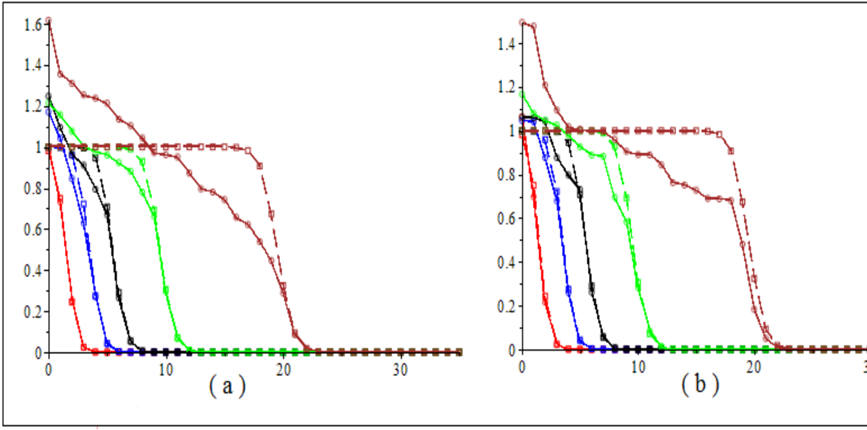


FIGURE : (a) Graphs of $\lambda(H)$ versus $\lambda(\mathcal{Q}_m)$ with $n = 200$ and for the various values of $m = 2, 4, 6, 10, 20$, (from the left to the right), (b) same as (a) with $\lambda(A^*A)$ instead of $\lambda(H)$.

Approximation by the integral operator.

We come back to

$$A^*A = \frac{m}{n^2}\sum_{\ell=1}^{n}\left(e^{2i\pi m Z_\ell (Y_k - Y_j)}\right)_{j,k=1,\cdots n}.$$

$$H := \mathbb{E}_Z(A^*A) = \frac{1}{n}\left(\frac{\sin(\pi m(Y_k - Y_j))}{\pi(Y_k - Y_j)}\right)_{j,k=1,\cdots n}.$$

Recall : want to prove that the spectrum of A^*A and H are close to the spectrum of $\mathcal{Q}_m$.
It remains to prove that the spectra of $\mathbb{E}(H)$ and $\mathcal{Q}_m$ are close.

A variant of Koltchinskii–Giné Theorem.

Theorem. *Let κ be a positive definite kernel on $(-1/2, +1/2)$ and T_κ be the integral operator with kernel κ and $\lambda(T_\kappa)$ its spectrum in decreasing order. Let*

$$H_\kappa := n^{-1} \left(\kappa(Y_j, Y_k)\right)_{j,k=1,\cdots n}.$$

Then, assuming that $m := \sup_y \kappa(y, y)$ is finite, one has the inequality

$$\mathbb{E}\left(\sum_{j=0}^{n-1} |\lambda_j(H_\kappa) - \lambda_j(T_\kappa)|^2\right) \leq \frac{2m^2}{n}.$$

One word on the proof of the theorem.

Let κ be a positive definite kernel on $(-1/2, +1/2)$, which is the kernel of the compact operator T. By Mercer Theorem it may be written as

$$\kappa(x, y) = \sum_j \lambda_j(T)\psi_j(x)\psi_j(y),$$

where ψ_j is an orthonormal basis of eigenfunctions.

The proof consists in cutting the sum into two parts.

One word on the proof of the theorem.

Let κ be a positive definite kernel on $(-1/2, +1/2)$, which is the kernel of the compact operator T. By Mercer Theorem it may be written as

$$\kappa(x, y) = \sum_j \lambda_j(T)\psi_j(x)\psi_j(y),$$

where ψ_j is an orthonormal basis of eigenfunctions.

The proof consists in cutting the sum into two parts.

Mercer's Formula is also used in machine learning to define a feature function, with values in a feature space. Namely the function $x \mapsto \Psi(x) = (\sqrt{\lambda_j(T)}\psi_j(x))_{j\geq 0}$. The feature space is the space ℓ^2 and

$$k(x, y) = \langle \Psi(x), \Psi(y) \rangle_{\ell^2}.$$

Conclusion for the spectra.

Putting together inequalities, , with probability larger than $1 - 2e^2 e^{-\frac{\xi^2}{2}}$,

$$\|\lambda(H) - \lambda(\mathcal{Q}_m)\|_{\ell^2} \leq \frac{(\xi + 1.5)m}{\sqrt{n}}.$$

Similarly, with probability larger than $1 - 2e^2 e^{-\frac{\xi^2}{2}}$,

$$\|\lambda(A^*A) - \lambda(\mathcal{Q}_m)\|_{\ell^2} \leq \frac{(2\xi + 1.5)m}{\sqrt{n}}.$$

Conclusion for the spectra.

Putting together inequalities, , with probability larger than $1 - 2e^2 e^{-\frac{\xi^2}{2}}$,

$$\|\lambda(H) - \lambda(\mathcal{Q}_m)\|_{\ell^2} \leq \frac{(\xi + 1.5)m}{\sqrt{n}}.$$

Similarly, with probability larger than $1 - 2e^2 e^{-\frac{\xi^2}{2}}$,

$$\|\lambda(A^*A) - \lambda(\mathcal{Q}_m)\|_{\ell^2} \leq \frac{(2\xi + 1.5)m}{\sqrt{n}}.$$

For m/n small, practical computations can be done on the integral operator.

Comparison with other results.

For a given eigenvalue,

$$|\lambda_j(H) - \lambda_j(\mathcal{Q}_m)| \leq \|\lambda(H) - \lambda(\mathcal{Q}_m)\|_{\ell^2} \leq \frac{(\xi + 1.5)m}{\sqrt{n}},$$

See the work of Shawe-Taylor, Cristianini and Kandola, and also Blanchard Bousquet and Wald... where such inequalities for one eigenvalue are developed for kernel matrices. But here all eigenvalues at the same time.

Comparison with other results.

For a given eigenvalue,

$$|\lambda_j(H) - \lambda_j(\mathcal{Q}_m)| \leq \|\lambda(H) - \lambda(\mathcal{Q}_m)\|_{\ell^2} \leq \frac{(\xi + 1.5)m}{\sqrt{n}},$$

See the work of Shawe-Taylor, Cristianini and Kandola, and also Blanchard Bousquet and Wald... where such inequalities for one eigenvalue are developed for kernel matrices. But here all eigenvalues at the same time.

Other results of these authors :

- $$\mathbb{E}(\sum_{j \geq s} \lambda_j(H)) \leq \sum_{j \geq s} \lambda_j(\mathcal{Q}_m).$$
- With large probability
$$|\sum_{j \geq s} \lambda_j(H) - \mathbb{E}(\sum_{j \geq s} \lambda_j(H))| \leq \frac{\xi m}{\sqrt{n}}.$$

Reconstruction error.

Because of the min-max principle

$$\sum_{j<s} \lambda_j(A)^2 = \sum_{j<s} \lambda_j(A^*A) = \sup \sum_{i=1}^{s} \langle A^*Av_i, v_i \rangle,$$

where the supremum is taken over all orthonormal system of s vectors.

Reconstruction error.

Because of the min-max principle

$$\sum_{j<s} \lambda_j(A)^2 = \sum_{j<s} \lambda_j(A^*A) = \sup \sum_{i=1}^{s} \langle A^*Av_i, v_i \rangle,$$

where the supremum is taken over all orthonormal system of s vectors. Equivalently, this quantity is

$$\sum_{j\geq s} \lambda_j(A)^2 = \min \|A - PAP\|_{\mathrm{HS}}^2$$

where the supremum is taken on all orthogonal projections P on subspaces of dimension s. This is the *reconstruction error* when replacing A by its action of a subspace of dimension s.

Reconstruction error.

Because of the min-max principle

$$\sum_{j<s} \lambda_j(A)^2 = \sum_{j<s} \lambda_j(A^*A) = \sup \sum_{i=1}^{s} \langle A^*Av_i, v_i \rangle,$$

where the supremum is taken over all orthonormal system of s vectors. Equivalently, this quantity is

$$\sum_{j \geq s} \lambda_j(A)^2 = \min \|A - PAP\|_{\mathrm{HS}}^2$$

where the supremum is taken on all orthogonal projections P on subspaces of dimension s. This is the *reconstruction error* when replacing A by its action of a subspace of dimension s.

$$\mathbb{E}_Z(\sum_{j \geq s} \lambda_j(A)^2) \leq \sum_{j \geq s} \lambda_j(H).$$

$$\mathbb{E}(\sum_{j \geq s} \lambda_j(A)^2) \leq \sum_{j \geq s} \lambda_j(\mathcal{Q}_m).$$

Comparison with the integral operator.

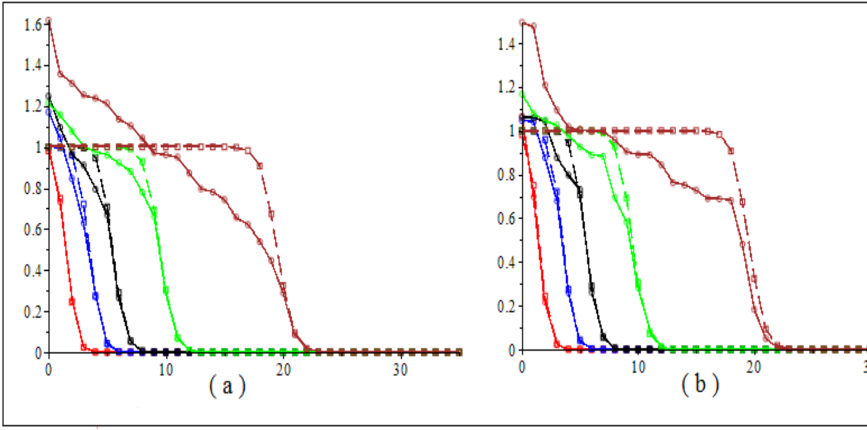


FIGURE : (a) Graphs of $\lambda(H)$ versus $\lambda(\mathcal{Q}_m)$ with $n = 200$ and for the various values of $m = 2, 4, 6, 10, 20$, (from the left to the right), (b) same as (a) with $\lambda(A^*A)$ instead of $\lambda(H)$.