Matrices de Fourier aléatoires, matrices de Gram

Aline Bonami,
Université d'Orléans, France

travail en collaboration avec Abderrazek Karoui
avec une partie en cours avec Zaineb Aloui

Hammamet,
le 2 novembre 2019

Random Fourier Matrices

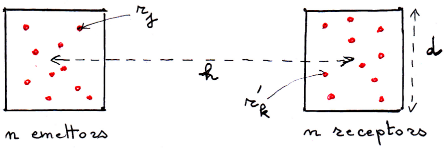
$$A := \frac{1}{n} \begin{pmatrix} e^{2i\pi m Z_1 \cdot Y_1} & \dots & e^{2i\pi m Z_1 \cdot Y_n} \\ \vdots & & \vdots \\ e^{2i\pi m Z_n \cdot Y_1} & \dots & e^{2i\pi m Z_n \cdot Y_n} \end{pmatrix}.$$

Here Y_j, Z_k independent samples of the laws P and Q on $\mathbb{R}^d$.
So A is a random matrix, with alea given by Z_j's and Y_k's.

Remark. It is equivalent to erase m from the exponent and replace the law P by the law P_m of mY_1.

Remark. When $d = 1, m = 1$, if Y_j and Z_k are replaced by the constants j/n and k/n, it is the Fourier matrix used in FFT.

We assume $m \ll n$ large. Typically, $m \approx n^\delta$ with $0 \leq \delta < 1$.



MIMO transmission between two squares of side d, separated by a distance h, with n nodes in each squares.
The *phase fading* between the emission and the reception is given by the matrix

$$c_{jk} = \frac{e^{2i\pi|r_j - r'_k|/\lambda}}{|r_j - r'_k|}.$$

Here r_j is the position of the $j-$th emitter node while r'_k is the position of the $k-$th receptor node. Both $n-$tuples of nodes are chosen randomly.

Questions.

The physics is well-known : the phase fading between emitters and receptors is given by the Helmoltz equation.

Desgroseilliers, Lévêque and Preissmann, EPFL (2013) consider this model, with two questions on the singular values of the *channel matrix*

$$c_{jk} = \frac{e^{2i\pi|r_j - r'_k|/\lambda}}{|r_j - r'_k|} :$$

- Number of degrees of freedom, that is, the number of significant singular values.
- Approximation of the Shannon capacity of the system, given by

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

We want to know this with large probability.

The simplified model (Desgroseilliers et al.).

For the study of the spectrum of the matrix

$$c_{jk} = \frac{e^{2i\pi|r_j - r'_k|/\lambda}}{|r_j - r'_k|},$$

it can be replaced (up to a small error) by the matrix with entries

$$\widetilde{c}_{jk} = \frac{\exp(2i\pi m y_j y'_k)}{h},$$

with $m = \frac{d^2}{h\lambda}$, (recall λ is the wave length).

Notation. $\lambda(M) = (\lambda_0(M), \cdots \lambda_j(M), \cdots)$ is the sequence of *singular values* of a matrix M, in decreasing order. When M is a symmetric positive semi-definite matrix (resp. operator), it is the spectrum of the matrix. In general,

$$\lambda_j(M) = \sqrt{\lambda_j(M^* M)}.$$

Two tools for spectra.

Let U be unitary. Then $\lambda(UA) = \lambda(AU) = \lambda(A)$.

A fundamental inequality : Hoffman–Wielandt inequality for singular values of matrices or operators, or

$$\sum_{j=1}^{n} (\lambda_j(D) - \lambda_j(D'))^2 \leq \|D - D'\|_{HS}^2.$$

Used to have approximation results in terms of the ℓ_n^2 norm. As a consequence ([BK] 2019)

$$\|\lambda(C) - \lambda(\widetilde{C})\|_{\ell_n^2} = \frac{n}{h} O(\frac{d^2}{\lambda h})$$

which gives a small relative error if $\frac{d^3}{\lambda h^2}$ is small.

Stationary Random Gram matrices.

$$H := \frac{1}{n} \begin{pmatrix} \kappa(m(Y_1, Y_1)) & \cdots & \kappa(m(Y_1, Y_n)) \\ \vdots & & \vdots \\ \kappa(m(Y_n, Y_1)) & \cdots & \kappa(m(Y_n, Y_n)) \end{pmatrix},$$

where κ is a positive definite function on $\mathbb{R}^d$, which is stationary : $\kappa(x, y) = \kappa(x - y)$. Here $Y_1, ..., Y_n$ is a sample of the probability law P.

We assume κ even and $\kappa(0) = 1$.
Bochner's Theorem : $\kappa = \mathcal{F}Q$, where Q is a symmetric probability law.

$$\mathcal{F}(Q)(y) := \int_{\mathbb{R}^d} e^{2i\pi y \cdot z} dQ(z).$$

Same kind of questions on the spectrum of H : asymptotic properties for m, n large.

A tool for big data.

Estimates for one eigenvalue of H are given by Shawe-Taylor, Cristianini and Kandola (see also Blanchard Bousquet and Zwald). ℓ^2 estimates have been given by Adamczak et al., Rosasco et al.,... 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) : one looks for the eigenfunctions and eigenvectors of a kernel matrix, for which scalar products are replaced by $\kappa(x, y)$, with κ a positive definite kernel.

The link between matrices

We have defined

- the random Fourier Matrix A with entries $\frac{1}{n}e^{2i\pi Z_j \cdot Y_k}$, with (Y_k) a sample of the law P, while (X_j) is a sample of the law Q,
- the random Gram Matrix H with entries $\frac{1}{n}\kappa(Y_j - Y_k)$, with (Y_k) a sample of the law P and $\kappa = \mathcal{F}Q$.

$$\mathbb{E}_Z(A^*A)_{jk} = \frac{1}{n}E_Z\left(\sum_l e^{2i\pi(Z_l \cdot (Y_j - Y_k))}\right) = H_{jk}.$$

From concentration inequalities we know that the spectrum of A^*A is close to the spectrum of its expectation.

T is the integral operator on $L^2(\mathbb{R}^d, P)$ given by

$$Tf(x) = \int \kappa(m(x-y))f(y)dP(y),$$

with $\kappa = \mathcal{F}Q$.
Singular values of A^*A, H and T are close with large probability.

Theorem [BK]. For $\xi > 0$, we have

$$\|\lambda(A^*A) - \lambda(T)\|_{\ell^2} \leq \frac{\sqrt{2}(\xi+1)}{\sqrt{n}}, \; \|\lambda(H) - \lambda(T)\|_{\ell^2} \leq \frac{(\xi+1)}{\sqrt{n}},$$

and

$$\|\lambda(A^*A) - \lambda(H)\|_{\ell^2} \leq \frac{(\xi+1)}{\sqrt{n}},$$

each of them with probability at least $1 - e^{-\xi^2}$.

Use of McDiarmid's concentration inequality

Concentration inequalities.

McDiarmid's inequality (1989) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function satisfying the bounded difference condition*

$$|f(z_1, \cdots, z_i, \cdots, z_n) - f(z_1, \cdots, z_i', \cdots, z_n)| \leq R$$

If $Z_1, \cdots Z_n$ are independent random variables, then we have

$$\mathbb{P}\left(|f(Z_1, \cdots, Z_n) - \mathbb{E}f(Z_1, \cdots, Z_n)| > \xi\right) \leq 2\exp\left(-\frac{2\xi^2}{nR^2}\right).$$

One needs also estimates for the expectation.

Uniform laws on symmetric intervals.

For a moment we assume that P and Q are uniform laws on the interval $(-1/2, +1/2)$. The κ is the sinc kernel given by

$$\kappa(x) = \frac{\sin(\pi x)}{\pi x},$$

and

$$\kappa(m(y-z)) = \frac{\sin(\pi m(y-z))}{\pi m(y-z))}.$$

The operator T, which we note $\mathcal{Q}_m$, has been largely studied from the sixties.

Simulations for uniform laws .

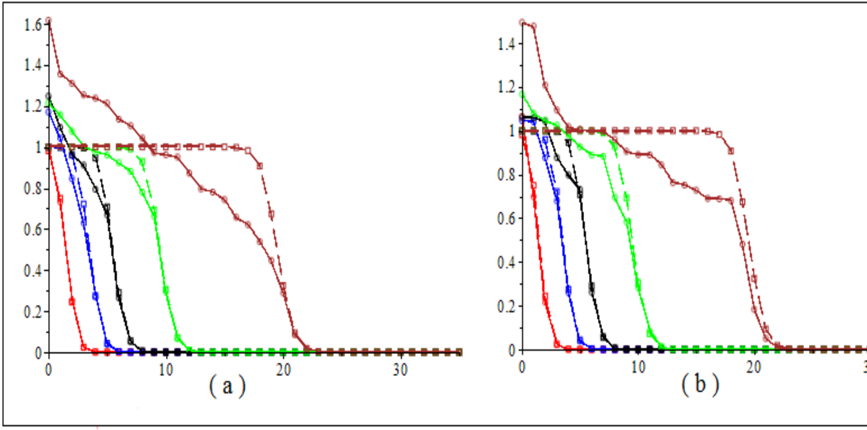


FIGURE – (a) Graphs of $\lambda(A^*A)$ versus $\lambda(\mathcal{Q}_m)$ with $n = 300$ and for the various values of $m = 2, 4, 6, 10, 20$, (from the left to the right), (b) same as (a) with $\lambda(E_Z(A^*A))$ instead of $\lambda(A^*A)$.

Degrees of freedom for the sinc operator.

Let $\varepsilon < 1$.

(1) For the integral operator
Define the number of degrees of freedom as the number of λ_s's such that $\lambda_s > \varepsilon$:

$$\deg_\infty(\mathcal{Q}_m, \varepsilon) := \min\{s; \lambda_s \leq \varepsilon\}.$$

With this definition

$$\deg_\infty(\mathcal{Q}_m, \varepsilon) = m + O\left(\frac{log(m)}{\varepsilon(1-\varepsilon)}\right).$$

(2) for the random matrix : the number of degree of freedom at level ε and confidence level α is defined as

$$\deg_\infty(M, \varepsilon, \alpha) = \min\{s; \lambda_s(T) \leq \varepsilon \quad \text{with probability } \geq \alpha\}.$$

We have

$$\deg_\infty(A^*A, \varepsilon, 1 - e^{-\xi^2}) = m + O_\varepsilon(\log m + \frac{\sqrt{2}(\xi^2+1)m^2}{n}).$$

« Hommage d'artistes à Notre-Dame », Mohamed Lekleti

What can be said for general laws P and Q ?

Work in progress with Zaineb Aloui.

Assume that P and Q have bounded density φ and ψ on $\mathbb{R}^d$. Then

$$\|T\|_{HS}^2 = \int \mathcal{F}\psi(m(x-y))^2\varphi(x)\varphi(y)dxdy = \frac{\|\varphi\|_2^2\|\psi\|_2^2}{m^d} + o(m^{-d}).$$

Corollary. The theoretical errors found between ℓ^2 norms of the spectra of T, H and A^*A are small when m^d/n is small.

$$\|T\| \leq \frac{\|\varphi\|_\infty\|\psi\|_\infty}{m^d}.$$

As a consequence,

$$\deg_\infty(T, \frac{\varepsilon}{m^d}) \approx m^d.$$

Particular cases.

Assume that P and Q are uniform laws, respectively on E and F. Then

$$\|\varphi\|_2^2 = \|\varphi\|_\infty = |E|^{-1/2}.$$

As a consequence, for $\alpha^{-1} = |E||F|$,

$$\sum \lambda_j(T)(\frac{\alpha}{m^d} - \lambda_j(T)) = o(m^{-d}).$$

Roughly speaking, $|E||F|m^d$ of the λ_j's are approximately $\frac{\alpha}{m^d}$ while the other ones are 0 as in the case of intervals.

Euclidean Random matrices.

Physicists are interested by the matrix with coefficients

$$\frac{\sin(m\pi|Y_j - Y_k|)}{nm\pi|Y_j - Y_k|)}$$

in 3 dimensions, which appears in wave propagation. Studied by Skipetrov and Goetschy.

They say that in first approximation, $c^{-1}m^2$ eigenvalues are non zero and equal to cm^{-2}.

Here Q is the normalized euclidean measure of the unit sphere. By approaching it by the uniform measure on the shell

$$\{(1 - \frac{c}{m}) \leq |x| \leq 1\},$$

one gets a proof of this behavior for the associated integral operator.

Also valid for the the random matrix with large probability as long as the corresponding error $\sqrt{n}^{-1}$ is small compared to m^2.