Singular Values of Random Subensembles of Frame Vectors

Mark Magsino

Ohio State University

Harmonic Analysis and PDE Seminar
CUNY Graduate Center
March 4, 2022

These are my collaborators.

Dustin Mixon
Ohio State University

Hans Parshall
Western Washington University

Table of Contents

Table of Contents

Frames are bases with more vectors.

Let's consider $\mathbb{F}^d$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

- A basis for $\mathbb{F}^d$ is a linearly independent spanning set of d vectors.
- A frame for $\mathbb{F}^d$ is **any** spanning set of $n \geq d$ vectors.
- We can think of them as bases with some extra vectors added.
- Extra vectors $\rightarrow$ no linear independence and redundant
- Linear independence $\rightarrow$ unique representations

Frames are bases with more vectors.

Let's consider $\mathbb{F}^d$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

- A basis for $\mathbb{F}^d$ is a linearly independent spanning set of d vectors.
- A frame for $\mathbb{F}^d$ is **any** spanning set of $n \geq d$ vectors.
- We can think of them as bases with some extra vectors added.
- Extra vectors $\rightarrow$ no linear independence and redundant
- Linear independence $\rightarrow$ unique representations

If frames are redundant and linear independence is nice, **why use frames?**

Redundancy is the benefit of frames.

Having redundant representations can be useful for a few things:

- Easing losses and noise from transmission
  - $(1, ?, 3)$

Redundancy is the benefit of frames.

Having redundant representations can be useful for a few things:

- Easing losses and noise from transmission
  - $(1, ?, 3) \to (1, 1, 1|2, ?, 2|3, 3, 3)$

Redundancy is the benefit of frames.

Having redundant representations can be useful for a few things:

- Easing losses and noise from transmission
  - $(1, ?, 3) \to (1, 1, 1|2, ?, 2|3, 3, 3)$
  - (3,1)-repetition code

Redundancy is the benefit of frames.

Having redundant representations can be useful for a few things:

- Easing losses and noise from transmission
  - $(1, ?, 3) \to (1, 1, 1|2, ?, 2|3, 3, 3)$
  - (3,1)-repetition code
- Obtaining sparse representations and data compression

Redundancy is the benefit of frames.

Having redundant representations can be useful for a few things:

- Easing losses and noise from transmission
  - $(1, ?, 3) \to (1, 1, 1|2, ?, 2|3, 3, 3)$
  - (3,1)-repetition code
- Obtaining sparse representations and data compression
  - Add v to standard basis in $\mathbb{R}^5$

Redundancy is the benefit of frames.

Having redundant representations can be useful for a few things:

- Easing losses and noise from transmission
  - $(1, ?, 3) \to (1, 1, 1|2, ?, 2|3, 3, 3)$
  - (3,1)-repetition code
- Obtaining sparse representations and data compression
  - Add v to standard basis in $\mathbb{R}^5$
  - $(0, 0, 0, 0, 0, 1)$

Redundancy is the benefit of frames.

Having redundant representations can be useful for a few things:

- Easing losses and noise from transmission
  - $(1, ?, 3) \to (1, 1, 1|2, ?, 2|3, 3, 3)$
  - (3,1)-repetition code
- Obtaining sparse representations and data compression
  - Add v to standard basis in $\mathbb{R}^5$
  - $(0, 0, 0, 0, 0, 1) \to (0, 5, 3)$

Redundancy is the benefit of frames.

Having redundant representations can be useful for a few things:

- Easing losses and noise from transmission
  - $(1, ?, 3) \to (1, 1, 1|2, ?, 2|3, 3, 3)$
  - (3,1)-repetition code
- Obtaining sparse representations and data compression
  - Add v to standard basis in $\mathbb{R}^5$
  - $(0, 0, 0, 0, 0, 1) \to (0, 5, 3)$
  - JPEG image compression

Redundancy is the benefit of frames.

Having redundant representations can be useful for a few things:

- Easing losses and noise from transmission
  - $(1, ?, 3) \to (1, 1, 1|2, ?, 2|3, 3, 3)$
  - (3,1)-repetition code
- Obtaining sparse representations and data compression
  - Add v to standard basis in $\mathbb{R}^5$
  - $(0, 0, 0, 0, 0, 1) \to (0, 5, 3)$
  - JPEG image compression

Redundant frames ease losses in transmission and can obtain sparse representations. How do we know which ones are good?

Frames have a reconstruction formula using frame vectors.

Theorem

Let $\{v_j\}_{j=1}^n$ *be a frame for* $\mathbb{F}^d$ *and let* $\mathbf{S}$ *be its frame operator. Then,*

$$x = \sum_{j=1}^{n} \langle \mathbf{S}^{-1} v_j, x \rangle v_j, \quad \forall x \in \mathbb{F}^d.$$

- Frame operators are linear mappings $\rightarrow$ it has a matrix form!
- $\mathbf{S}$ is always invertible but matrix inversion can be expensive.
- For orthonormal bases $\mathbf{S} = \mathbf{I}$.
- Other reconstruction coefficients possible!

Frames have a reconstruction formula using frame vectors.

Theorem

Let $\{v_j\}_{j=1}^n$ be a frame for $\mathbb{F}^d$ and let $\mathbf{S}$ be its frame operator. Then,

$$x = \sum_{j=1}^{n} \langle \mathbf{S}^{-1} v_j, x \rangle v_j, \quad \forall x \in \mathbb{F}^d.$$

- Frame operators are linear mappings $\rightarrow$ it has a matrix form!
- $\mathbf{S}$ is always invertible but matrix inversion can be expensive.
- For orthonormal bases $\mathbf{S} = \mathbf{I}$.
- Other reconstruction coefficients possible!

The inverse of frame operators are needed for the reconstruction formula, but it can be hard to compute. Can we make it easier?

Tight frames give us good frame operators.

- Let $\{v_j\}_{j=1}^n$ be a frame for $\mathbb{F}^d$. The frame operator, $\mathbf{S} : \mathbb{F}^d \to \mathbb{F}^d$, is the linear map
$$\mathbf{S}(x) = \sum_{j=1}^{n} \langle v_j, x \rangle v_j.$$
- $\mathbf{S}$ is self-adjoint and positive definite $\to$ $\mathbf{S}$ has real and positive eigenvalues.
- A frame is tight if all those eigenvalues are the same.
- In this case we have $\mathbf{S} = A\mathbf{I}$, where A is the repeated eigenvalue.

Tight frames give us good frame operators.

- Let $\{v_j\}_{j=1}^n$ be a frame for $\mathbb{F}^d$. The frame operator, $\mathbf{S} : \mathbb{F}^d \to \mathbb{F}^d$, is the linear map

$$\mathbf{S}(x) = \sum_{j=1}^{n} \langle v_j, x \rangle v_j.$$

- $\mathbf{S}$ is self-adjoint and positive definite $\to$ $\mathbf{S}$ has real and positive eigenvalues.
- A frame is tight if all those eigenvalues are the same.
- In this case we have $\mathbf{S} = A\mathbf{I}$, where A is the repeated eigenvalue.

Tight frames are good for computing. What are the difficulties in computing in practice?

Extremely high dimensions can still be difficult.

- Inverting $d \times d$ frame operators is $O(d^3)$.
- This seems fine unless d is **very large**.
  - Hyperspectral imaging
  - Video processing
- Inversion computations might still be too slow.
- **Idea:** Can we do more with less by taking well-conditioned subensembles?

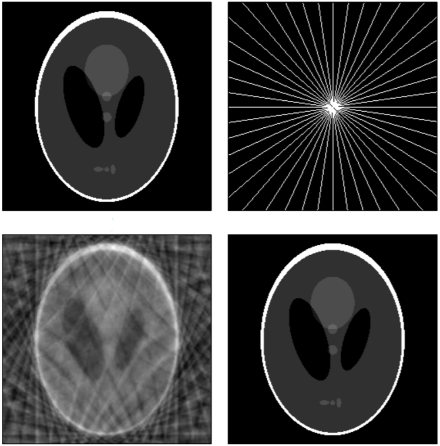
Extremely high dimensions can still be difficult.

- Inverting $d \times d$ frame operators is $O(d^3)$.
- This seems fine unless d is **very large**.
  - Hyperspectral imaging
  - Video processing
- Inversion computations might still be too slow.
- **Idea:** Can we do more with less by taking well-conditioned subensembles?

With very high dimensional data, computation could still be slow. We want to use well-conditioned subensembles but what does that mean?

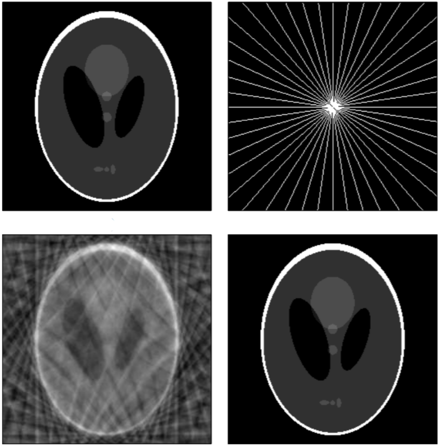
Table of Contents

We can cheat with low-dimensional signals. [1]



[1] Figure 1 in Candes, Romberg, Tao 2005

We can cheat with low-dimensional signals. [1]



If we sample images with lots of zeros correctly, we can rebuild them near perfectly, or sometimes perfectly! Why does this work?

[1] Figure 1 in Candes, Romberg, Tao 2005

We can formulate the problem using frames.

- We wish to measure an image with n pixels using $d \ll n$ linear measurements.
- How do we recover the image from our undersampled measurements?

Problem

Let $d \ll n$. Given an $d \times n$ matrix $\mathbf{A}$ and $y \in \mathbb{F}^d$, solve

$$y = \mathbf{A}x.$$

We can formulate the problem using frames.

- We wish to measure an image with n pixels using $d \ll n$ linear measurements.
- How do we recover the image from our undersampled measurements?

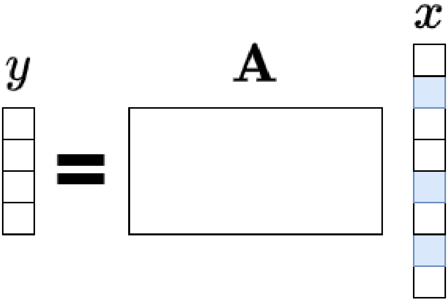
Problem

Let $d \ll n$. Given an $d \times n$ matrix $\mathbf{A}$ and $y \in \mathbb{F}^d$, solve

$$y = \mathbf{A}x.$$

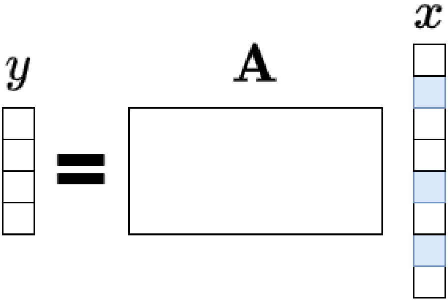
We are trying to solve a matrix-vector equation that should have many solutions. Why would having lots of zeros make that solution easier to find?

Solutions with many zeros have few variables.



- x is secretly a vector of 3 variables.
- 4 equations, 3 unknowns, could be solvable!
- We only know y...which x_i are the variables?

Solutions with many zeros have few variables.



- x is secretly a vector of 3 variables.
- 4 equations, 3 unknowns, could be solvable!
- We only know y...which x_i are the variables?

Our image secretly only has a few variables to solve for but which ones?
How do we design $\mathbf{A}$ to account for this?

RIP accounts for different variable combinations.

Definition

Let k be a positive integer and $\delta > 0$. An $d \times n$ matrix $\mathbf{A}$ satisfies the (k, δ)-restricted isometry property (RIP) if for every vector $x \in \mathbb{F}^n$ with at most k nonzero entries

$$(1-\delta)\|x\|^2 \leq \|\mathbf{A}x\|^2 \leq (1+\delta)\|x\|^2.$$

- We call x with this property k-sparse.
- Mapping $\mathbf{A}$ to k-dimensional subspaces is almost an isometry $\rightarrow$ almost information preserving.
- This requires $\binom{n}{k}$ checks. This is computationally infeasible...

RIP accounts for different variable combinations.

Definition

Let k be a positive integer and $\delta > 0$. An $d \times n$ matrix $\mathbf{A}$ satisfies the (k, δ)-restricted isometry property (RIP) if for every vector $x \in \mathbb{F}^n$ with at most k nonzero entries

$$(1-\delta)\|x\|^2 \leq \|\mathbf{A}x\|^2 \leq (1+\delta)\|x\|^2.$$

- We call x with this property k-sparse.
- Mapping $\mathbf{A}$ to k-dimensional subspaces is almost an isometry $\rightarrow$ almost information preserving.
- This requires $\binom{n}{k}$ checks. This is computationally infeasible...

RIP matrices scan k-subspaces nicely, but they're hard to find. Is it worth it?

RIP quantifies how many zeros is enough.

Theorem (Candes '08)

Let k be a positive integer and let $\delta < \sqrt{2} - 1$. Suppose x' is a k-sparse vector we wish to recover. If $\mathbf{A}$ is $(2k, \delta)$-RIP, then the solution to

$$\operatorname{argmin} \quad \|x\|_1 \quad \textit{subject to} \quad y = \mathbf{A}x$$

is precisely x'. In particular, the recovery of x' is exact.

RIP quantifies how many zeros is enough.

Theorem (Candes '08)

Let k be a positive integer and let $\delta < \sqrt{2} - 1$. Suppose x' is a k-sparse vector we wish to recover. If $\mathbf{A}$ is $(2k, \delta)$-RIP, then the solution to

$$\operatorname{argmin} \quad \|x\|_1 \quad \textit{subject to} \quad y = \mathbf{A}x$$

is precisely x'. In particular, the recovery of x' is exact.

RIP matrices are good measurement matrices. If they're hard to check how do people find them?

Dealing with RIP is a battle between average and worst case performance.

How do people make RIP matrices?

1. Random matrices: Works with high probability but you can't be sure...
2. Explicit constructions: Can make some guarantees based on dot products between columns but worse performance than random.
3. Random subensembles: Check random subensembles to verify RIP with high probability.

Dealing with RIP is a battle between average and worst case performance.

How do people make RIP matrices?

1. Random matrices: Works with high probability but you can't be sure...
2. Explicit constructions: Can make some guarantees based on dot products between columns but worse performance than random.
3. Random subensembles: Check random subensembles to verify RIP with high probability.

Dealing with RIP is a battle between average and worst case performance.

How do people make RIP matrices?

1. Random matrices: Works with high probability but you can't be sure...
2. Explicit constructions: Can make some guarantees based on dot products between columns but worse performance than random.
3. Random subensembles: Check random subensembles to verify RIP with high probability.

Random ensembles balance explicit constructions and good average vs worst case performance.

Table of Contents

We can analyze the spectra of random subensembles.

- Given an $n \times n$ self-adjoint matrix Z with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$, the empirical spectral distribution, μ_Z, is the uniform probability measure over the spectrum of Z.
- Our Z will be the frame operator or the Gram matrix of our random subensembles.
- We want to compare to other known distributions associated with random matrices.

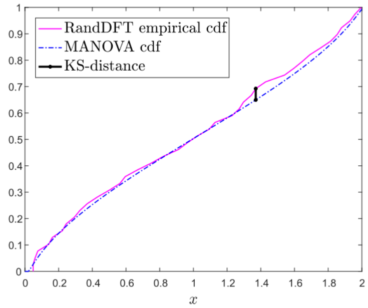
We can analyze the spectra of random subensembles.

- Given an $n \times n$ self-adjoint matrix Z with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$, the empirical spectral distribution, μ_Z, is the uniform probability measure over the spectrum of Z.
- Our Z will be the frame operator or the Gram matrix of our random subensembles.
- We want to compare to other known distributions associated with random matrices.

The empirical spectral distribution allows us to analyze spectra of random subensembles. How do we compare them?

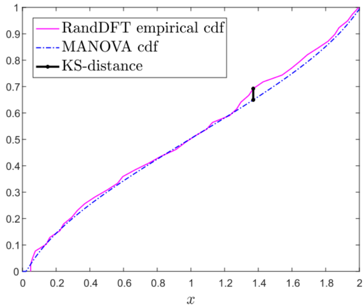
We can compare using Kolmogorov-Smirnov distance.

- KS distance = largest distance between the two CDFs.
- $X \in \mathbb{F}^{d \times n}$ is our frame.
- X_K in $\mathbb{F}^{d \times k}$ is our submatrix of X.
- Compare KS distance of the CDFs of:
  - ESD of $X_K X_K^*$
  - hypothesized limiting distribution



We can compare using Kolmogorov-Smirnov distance.

- KS distance = largest distance between the two CDFs.
- $X \in \mathbb{F}^{d \times n}$ is our frame.
- X_K in $\mathbb{F}^{d \times k}$ is our submatrix of X.
- Compare KS distance of the CDFs of:
  - ESD of $X_K X_K^*$
  - hypothesized limiting distribution



We can compare distributions with KS distance. What should the limiting distribution be?

The limiting distribution seems to be Wachter's MANOVA.

Wachter: As $k/d \to \beta$ and $d/n \to \gamma$, the ESD of MANOVA$(n, d, k, \mathbb{R})$ ensembles converges to the MANOVA(β, γ) distribution with density

$$f_{\beta,\gamma}^{MANOVA}(x) = \frac{\sqrt{(x - r_-)(r_+ - x)}}{2\beta\pi x(1 - \gamma x)} \cdot I_{(r_-, r_+)}(x) + \left(1 + \frac{1}{\beta} - \frac{1}{\beta\gamma}\right)^+ \cdot \delta(x - \frac{1}{\gamma})$$

- $(x)^+ = \max(0, x)$
- $r_\pm = (\sqrt{\beta(1-\gamma)^2} \pm \sqrt{(1 - \beta\gamma)})^2$
- Supported on $[r_-, r_+]$.

The limiting distribution seems to be Wachter's MANOVA.

Wachter: As $k/d \to \beta$ and $d/n \to \gamma$, the ESD of MANOVA$(n, d, k, \mathbb{R})$ ensembles converges to the MANOVA(β, γ) distribution with density

$$f_{\beta,\gamma}^{MANOVA}(x) = \frac{\sqrt{(x-r_-)(r_+-x)}}{2\beta\pi x(1-\gamma x)} \cdot I_{(r_-,r_+)}(x) + \left(1 + \frac{1}{\beta} - \frac{1}{\beta\gamma}\right)^+ \cdot \delta(x - \frac{1}{\gamma})$$

- $(x)^+ = \max(0, x)$
- $r_\pm = (\sqrt{\beta(1-\gamma)^2} \pm \sqrt{(1-\beta\gamma)})^2$
- Supported on $[r_-, r_+]$.

MANOVA ensembles seem related and their ESD converges to Wachter's distribution. Is there evidence they are related to frame subensembles?

Evidence shows it really is Wachter's MANOVA.

- $\Delta_{KS}(X_K) = \|F(X_K) - F^{MANOVA}_{\beta,\gamma}\|_{KS}$
- $d/n \to \gamma = 0.5$
- $k/d \to \beta = 0.8$
- Graph: $-\frac{1}{2}\ln \mathbb{E}_K(\Delta_{KS}(X_K))$ vs. $\ln(n)$

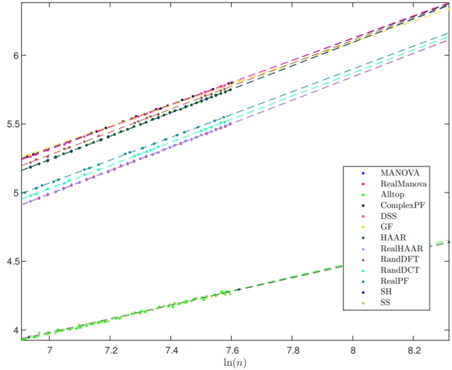


Figure: Haikin, Gavish, Zamir computer simulation (2017)

Evidence shows it really is Wachter's MANOVA.

- $\Delta_{KS}(X_K) = \|F(X_K) - F^{MANOVA}_{\beta,\gamma}\|_{KS}$
- $d/n \to \gamma = 0.5$
- $k/d \to \beta = 0.8$
- Graph: $-\frac{1}{2}\ln \mathbb{E}_K(\Delta_{KS}(X_K))$ vs. $\ln(n)$

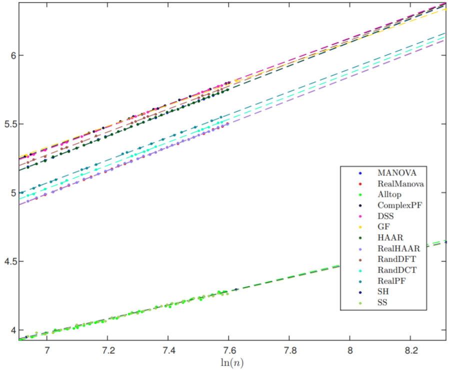


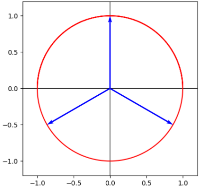
Figure: Haikin, Gavish, Zamir computer simulation (2017)

We have numerical evidence of convergence to MANOVA(β, γ) for tight frames. Can we prove this for any of them?

Equiangluar tight frames spread information evenly.

Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Let $X \subset \mathbb{F}^d$ be a tight frame of $n \geq d$ unit vectors. X is an equiangular tight frame if

$$|\langle x_i, x_j \rangle| = \sqrt{\frac{n-d}{d(n-1)}}, \quad i \neq j.$$

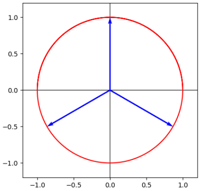


- ETFs are the next closest thing to orthonormal bases.
- If we think of ETFs as signal channels, they spread information and prevent interference.

Equiangluar tight frames spread information evenly.

Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Let $X \subset \mathbb{F}^d$ be a tight frame of $n \geq d$ unit vectors. X is an equiangular tight frame if

$$|\langle x_i, x_j \rangle| = \sqrt{\frac{n-d}{d(n-1)}}, \quad i \neq j.$$



- ETFs are the next closest thing to orthonormal bases.
- If we think of ETFs as signal channels, they spread information and prevent interference.

ETFs spread information well. How should we analyze the spectra of their subensembles?

We can study frames through its Gram matrix.

Given a frame $F \in \mathbb{F}^{d \times n}$, its Gram matrix, G is an $n \times n$ matrix defined by

$$G_{ij} = \langle x_i, x_j \rangle, \quad i, j \in \{1, \cdots, n\}.$$

In the case of real ETFs we can write

$$G = I + \left(\frac{n-d}{d(n-1)} \right)^{\frac{1}{2}} S,$$

where the off-diagonal entries of S consists of ± 1 and the main diagonal consists of zeros.

We can study frames through its Gram matrix.

Given a frame $F \in \mathbb{F}^{d\times n}$, its Gram matrix, G is an $n \times n$ matrix defined by

$$G_{ij} = \langle x_i, x_j \rangle, \quad i, j \in \{1, \cdots, n\}.$$

In the case of real ETFs we can write

$$G = I + \left(\frac{n-d}{d(n-1)}\right)^{\frac{1}{2}} S,$$

where the off-diagonal entries of S consists of ± 1 and the main diagonal consists of zeros.

The Gram matrix records information about inner products, but why is S special?

ETFs are closely tied to symmetric conference matrices.

An $n \times n$ matrix, S, is a conference matrix if

1. $S_{ii} = 0$, for every $i \in [n]$.
2. $S_{ij} \in \{\pm 1\}$, for every $i, j \in [n]$ and $i \neq j$.
3. $S^T S = (n-1)I$.

$$\left[\begin{array}{c|ccccc} 0 & + & + & + & + & + \\ \hline + & 0 & + & - & - & + \\ + & + & 0 & + & - & - \\ + & - & + & 0 & + & - \\ + & - & - & + & 0 & + \\ + & + & - & - & + & 0 \end{array}\right]$$

Relation with ETFs: If S is a symmetric conference matrix, then

$$I + \frac{1}{\sqrt{n-1}} S$$

is the Gram matrix of an ETF of n vectors in $\mathbb{R}^{n/2}$.

ETFs are closely tied to symmetric conference matrices.

An $n \times n$ matrix, S, is a conference matrix if

1. $S_{ii} = 0$, for every $i \in [n]$.
2. $S_{ij} \in \{\pm 1\}$, for every $i, j \in [n]$ and $i \neq j$.
3. $S^T S = (n-1)I$.

$$\left[\begin{array}{c|ccccc} 0 & + & + & + & + & + \\ \hline + & 0 & + & - & - & + \\ + & + & 0 & + & - & - \\ + & - & + & 0 & + & - \\ + & - & - & + & 0 & + \\ + & + & - & - & + & 0 \end{array}\right]$$

Relation with ETFs: If S is a symmetric conference matrix, then

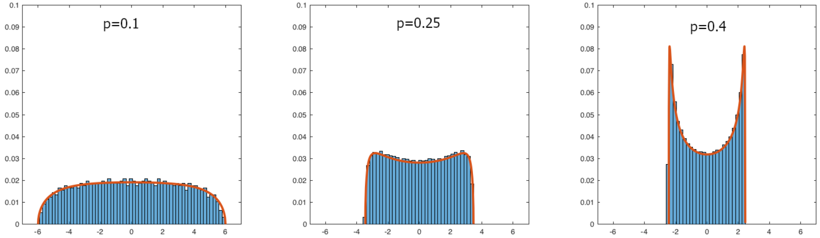
$$I + \frac{1}{\sqrt{n-1}} S$$

is the Gram matrix of an ETF of n vectors in $\mathbb{R}^{n/2}$.

Real ETFs of redundancy 2 are closely linked to symmetric conference matrices. What do “typical” ESDs of them look like?

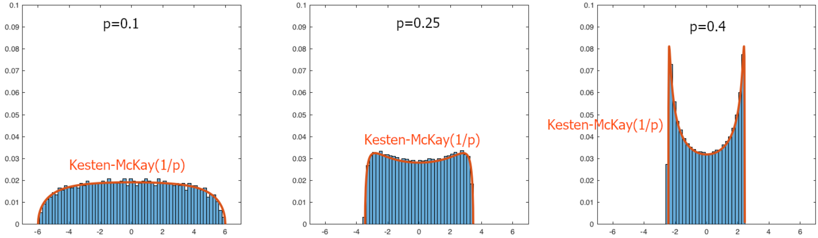
The spectrum of subensembles of Paley conference matrices look like a familiar distribution.

- Paley conference matrix of order $n = 10010$
- For each p: $I \subset [n]$ where $i \in I$ with probability p, independently
- ESD of principal submatrix using I



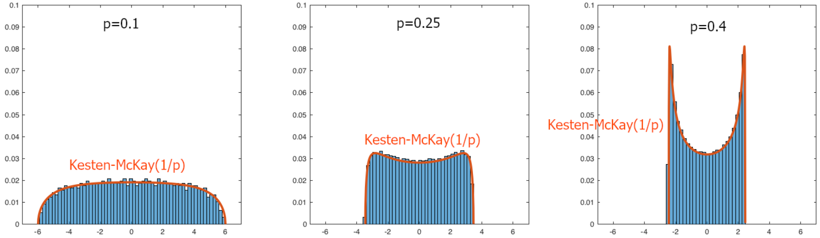
The spectrum of subensembles of Paley conference matrices look like a familiar distribution.

- Paley conference matrix of order $n = 10010$
- For each p: $I \subset [n]$ where $i \in I$ with probability p, independently
- ESD of principal submatrix using I



The spectrum of subensembles of Paley conference matrices look like a familiar distribution.

- Paley conference matrix of order $n = 10010$
- For each p: $I \subset [n]$ where $i \in I$ with probability p, independently
- ESD of principal submatrix using I



It seems when we subsample Paley conference matrices, the ESD converges to the Kesten-McKay distribution with parameter $\nu = 1/p$. Is this true?

The spectrum of subensembles of symmetric conferences matrices converge to the Kesten-McKay distribution.

A sequence $\{n_i\}_{i\in\mathbb{N}}$ is a lacunary sequence if there exists $\lambda > 1$ so that for each i, $n_{i+1} \geq \lambda n_i$.

Theorem (M., Mixon, Parshall (2019))

Fix $p \in (0, \frac{1}{2})$ and take any lacunary sequence L for which there exists a sequence $\{S_n\}_{n\in L}$ of symmetric conference matrices of increasing order n. Let X_n be the corresponding random principal submatrix of S_n where indices are included independently with probability p. Then, the empirical spectral distribution of $\frac{1}{p\sqrt{n}}X_n$ converges almost surely to the Kesten-McKay distribution with parameter $\nu = 1/p$.

The spectrum of subensembles of symmetric conferences matrices converge to the Kesten-McKay distribution.

A sequence $\{n_i\}_{i\in\mathbb{N}}$ is a lacunary sequence if there exists $\lambda > 1$ so that for each i, $n_{i+1} \geq \lambda n_i$.

Theorem (M., Mixon, Parshall (2019))

Fix $p \in (0, \frac{1}{2})$ and take any lacunary sequence L for which there exists a sequence $\{S_n\}_{n\in L}$ of symmetric conference matrices of increasing order n. Let X_n be the corresponding random principal submatrix of S_n where indices are included independently with probability p. Then, the empirical spectral distribution of $\frac{1}{p\sqrt{n}}X_n$ converges almost surely to the Kesten-McKay distribution with parameter $\nu = 1/p$.

The spectrum of random subensembles of symmetric conference matrices does indeed converge to the Kesten-McKay distribution. How does the proof work?

The proof involves probability and combinatorics.

Proposition

Let $\{\zeta_i\}_{i=1}^{\infty}$ *be a sequence of uniformly subgaussian random probability measures, and let* μ *be a non-random subgaussian probability measure. Suppose that for every* $k \in \mathbb{N}$ *it holds that*

1. $\mathbb{E} \int_{\mathbb{R}} x^k d\zeta_i(x) \to \int_{\mathbb{R}} x^k d\mu(x)$, *and*
2. $\sum_{i=1}^{\infty} \mathrm{Var}\left(\int_{\mathbb{R}} x^k d\zeta_i(x)\right) < \infty.$

Then, ζ_i *converges almost surely to* μ*.*

The proof involves probability and combinatorics.

Proposition

Let $\{\zeta_i\}_{i=1}^{\infty}$ *be a sequence of uniformly subgaussian random probability measures, and let* μ *be a non-random subgaussian probability measure. Suppose that for every* $k \in \mathbb{N}$ *it holds that*

1. $\mathbb{E} \int_{\mathbb{R}} x^k d\zeta_i(x) \to \int_{\mathbb{R}} x^k d\mu(x)$, *and*
2. $\sum_{i=1}^{\infty} \mathrm{Var}\left(\int_{\mathbb{R}} x^k d\zeta_i(x)\right) < \infty.$

Then, ζ_i *converges almost surely to* μ.

This theorem uses a lot of ideas from both probability theory and graph theory. How did we end up in the realm of combinatorics?

Kesten-McKay moments involve Catalan's triangle.

Catalan's triangle is given by

$$C(n,k) := \frac{(n+k)!(n-k+1)}{k!(n+1)!},$$

and the n-th Catalan number is given by $C(n,n)$. The Kesten-McKay moments are given by

$$\int_{\mathbb{R}} x^k d\mu_\nu(x) = \begin{cases} \sum_{j=1}^{k/2} C(k/2-1, k/2-j)\nu^j(\nu-1)^{k/2-j} & \text{if } k \text{ is even,} \\ 0 & \text{if } k \text{ is odd.} \end{cases}$$

Kesten-McKay moments involve Catalan's triangle.

Catalan's triangle is given by

$$C(n,k) := \frac{(n+k)!(n-k+1)}{k!(n+1)!},$$

and the n-th Catalan number is given by $C(n,n)$. The Kesten-McKay moments are given by

$$\int_{\mathbb{R}} x^k d\mu_\nu(x) = \begin{cases} \sum_{j=1}^{k/2} C(k/2-1, k/2-j)\nu^j(\nu-1)^{k/2-j} & \text{if } k \text{ is even,} \\ 0 & \text{if } k \text{ is odd.} \end{cases}$$

The proof involves Catalan numbers, because it naturally arises from the moments of the Kesten-McKay distribution. How far can these ideas go?

There's more work to be done!

Our result applies to a very specific case: real equiangular tight frames of redundancy 2.

- Can this be extended to other redundancies?
- What about non-equiangular tight frames?

Haikin, Gavish, and Zamir other conjectures:

- Power law for rate of convergence
- Universality for ETF convergence rate

There's more work to be done!

Our result applies to a very specific case: real equiangular tight frames of redundancy 2.

- Can this be extended to other redundancies?
- What about non-equiangular tight frames?

Haikin, Gavish, and Zamir other conjectures:

- Power law for rate of convergence
- Universality for ETF convergence rate

Our work is only tip of the iceberg. There are many directions this can still go.

These are some good references.

- Haikin, Zamir, Gavish. "Random subsets of structured deterministic frames have MANOVA spectra."
- M., Mixon, Parshall. "Kesten-McKay law for random subensembles of Paley equiangular tight frames."
- Haikin, Gavish, Mixon, Zamir. "Asymptotic Frame Theory for Analog Coding."

Thanks for listening! Any questions?