Phase transitions in random matrices and the spiked tensor model

Mihai Nica - University of Toronto

October 20, 2019

Outline

- **Motivating real world scenario:**
  Detecting a signal in a noisy/spam filled world
- **Toy Model 1:**
  Detecting a spike in Spiked Random Matrices
- **Toy Model 2:**
  Detecting a spike in Random Tensors

in-Ear Lightning Earphones Earbuds Headphones-Mic and Volume Remote and Noise Isolation, MFI-Certified, Compatible with iPhone X/XS/XR/XS Max, 7/7P, 8/8P iPad, iPod

by DAIRLE

4 customer reviews

Price: $20.98

- ♫PREMIUM SOUND - Built-in DAC & Φ10 dynamic driver finely-tuned to produce detailed, crisp and clear sound with powerful bass earphone builts in hands-free microphone and remote control, allow you to take call, play, pause, volume up and down
- ♫PLUG N PLAY - NEW generation MFi-Certified lightning earphone flawlessly compatible with iPhone/iPad/iPod; LOW POWER
- ♫STYLISH & ERGONOMIC DESIGN - Lightweight aluminum body; comfortable in-ear design that

This is the best MFI certified lightning earbuds we can find. Ryan Blanche

May 29, 2019

Color: Black. | Verified Purchase

The quality of these headphones is very high! perfect! The binaural compatible earbuds are very good and will not tangled.Excellent sound quality. Still the best stereo in history, affordable and durable headphones. This is what I like

Great Apple MFI Certified Lightning earbuds Felicia Wells

May 29, 2019

Color: Black. | Verified Purchase

These earbuds work well during operation, they don't fall out, and you can change the volume and pause the music, which I really need.Excellent voice! Good sound quality. I am very satisfied with this

Great Apple MFI Lightning headphones Ethel Stout

June 3, 2019

Color: Black. | Verified Purchase

The earphones with remote and microphone is my absolute favorite ear pod set. It is light weight so I hardly know I'm wearing it. The ear bud part is just the right size so it doesn't hurt my ears.

https://reviewmeta.com/

Most Trusted Reviews

We couldn't find any quality reviews

Least Trusted Reviews

5/5 Great Apple MFI Certified Lightning earbuds

These earbuds work well during operation, they don't fall out, ... [Go to full review]

May 29, 2019

0% TRUST

- Verified Purchaser
- Contains repetitive phrases (show)

Reviewer: Felicia Wells

- One-Hit Wonder
- Overrepresented Participation (posted 1 review)
- Easy Grader (avg. rating: 5.0)

5/5 This is the best MFI certified lightning earbuds we can find.

The quality of these headphones is very high! perfect! The ... [Go to full review]

May 29, 2019

0% TRUST

- Verified Purchaser
- Overrepresented word count (41 words)
- Contains repetitive phrases (show)

Reviewer: Ryan Blanche

The model (for a distribution on spammy reviews):

The model (for a distribution on spammy reviews):

- Set of N **words**: $\mathcal{W} = \{Word_1, Word_2, \ldots, Word_N\}$

The model (for a distribution on spammy reviews):

- Set of N **words**: $\mathcal{W} = \{Word_1, Word_2, \ldots, Word_N\}$
- A **document** is a "subset", $D \subset \mathcal{W}$ (aka "bag-of-words")

The model (for a distribution on spammy reviews):

- Set of N **words**: $\mathcal{W} = \{Word_1, Word_2, \ldots, Word_N\}$
- A **document** is a "subset", $D \subset \mathcal{W}$ (aka "bag-of-words")
- We receive a list of documents $D_1, \ldots, D_{N_{Docs}}$

The model (for a distribution on spammy reviews):

- Set of N **words**: $\mathcal{W} = \{Word_1, Word_2, \ldots, Word_N\}$
- A **document** is a "subset", $D \subset \mathcal{W}$ (aka "bag-of-words")
- We receive a list of documents $D_1, \ldots, D_{N_{Docs}}$
- All documents are either **Spam** or **Ham** (unknown to us!)

The model (for a distribution on spammy reviews):

- Set of N **words**: $\mathcal{W} = \{Word_1, Word_2, \ldots, Word_N\}$
- A **document** is a "subset", $D \subset \mathcal{W}$ (aka "bag-of-words")
- We receive a list of documents $D_1, \ldots, D_{N_{Docs}}$
- All documents are either **Spam** or **Ham** (unknown to us!)
- Each document is **Ham** with probability $\mathbf{P}(Ham) \in [0, 1]$, and **Spam** with probability $\mathbf{P}(Spam) = 1 - \mathbf{P}(Ham)$. (Hopefully $\mathbf{P}(Ham) > \mathbf{P}(Spam)$)

The model (for a distribution on spammy reviews):

- Set of N **words**: $\mathcal{W} = \{Word_1, Word_2, \ldots, Word_N\}$
- A **document** is a "subset", $D \subset \mathcal{W}$ (aka "bag-of-words")
- We receive a list of documents $D_1, \ldots, D_{N_{Docs}}$
- All documents are either **Spam** or **Ham** (unknown to us!)
- Each document is **Ham** with probability $\mathbf{P}(Ham) \in [0, 1]$, and **Spam** with probability $\mathbf{P}(Spam) = 1 - \mathbf{P}(Ham)$. (Hopefully $\mathbf{P}(Ham) > \mathbf{P}(Spam)$)
- The **words** from **every ham document** are drawn independently and identically distributed according a probability vector:

  $$\vec{P}_{Ham} = (\mathbf{P}(Word_1 | Ham), \mathbf{P}(Word_2 | Ham), \ldots \mathbf{P}(Word_N | Ham))^T$$

- The **spam documents** are created by some unknown mechanism.

“Inverse” Problem

- The model tell us the distribution of documents given $\mathbf{P}(Ham)$ and $\vec{P}_{Ham}$.

"Inverse" Problem

- The model tell us the distribution of documents given $\mathbf{P}(Ham)$ and $\vec{P}_{Ham}$.
- The real world problem is the inverse:
  Can we recover information about $\vec{P}_{Ham}$ given a sample of the documents?

A random matrix: Word Co-Occurrence Matrix

$$
\begin{aligned}
& \text{Co-occurance}\left(Word_i, Word_j\right) \\
= \quad & \frac{1}{N_{DOCS}} \sum_{d=1}^{N_{DOCS}} \frac{\#\left\{\text{Word pairs in } Doc_d \text{ that are } \left(Word_i, Word_j\right)\right\}}{\#\left\{\text{Word pairs in } Doc_d\right\}} \\
= \quad & \frac{1}{N_{DOCS}} \sum_{d=1}^{N_{DOCS}} \frac{\#_{Doc_d}\left(Word_i\right) \times \#_{Doc_d}\left(Word_j\right)}{\left|Doc_d\right|^2} \\
=: \quad & \hat{\mathsf{P}}\left(\text{Two uniform random words from a doc are } Word_i \text{ \& } Word_j\right)
\end{aligned}
$$

A random matrix: Word Co-Occurrence Matrix

$$
\begin{aligned}
& \text{Co-occurance}\left(Word_i, Word_j\right) \\
= \quad & \frac{1}{N_{DOCS}} \sum_{d=1}^{N_{DOCS}} \frac{\#\left\{\text{Word pairs in } Doc_d \text{ that are } \left(Word_i, Word_j\right)\right\}}{\#\left\{\text{Word pairs in } Doc_d\right\}} \\
= \quad & \frac{1}{N_{DOCS}} \sum_{d=1}^{N_{DOCS}} \frac{\#_{Doc_d}\left(Word_i\right) \times \#_{Doc_d}\left(Word_j\right)}{\left|Doc_d\right|^2} \\
=: \quad & \hat{\mathbf{P}}\left(\text{Two uniform random words from a doc are } Word_i \text{ \& } Word_j\right)
\end{aligned}
$$

Proposition

Let C be the $N \times N$ matrix with $C_{ij} = \text{Co-occurrence}(Word_i, Word_j)$.

$$
\begin{aligned}
(\dagger) \quad \mathbf{E}\left[C\right] &= \mathbf{P}\left(Ham\right) \vec{P}_{Ham} \vec{P}_{Ham}^T + \mathbf{P}\left(Spam\right) S_{Spam} \\
i.e. \quad C &= \mathbf{P}\left(Ham\right) \vec{P}_{Ham} \vec{P}_{Ham}^T + \mathbf{P}\left(Spam\right) S_{Spam} + \left(C - \mathbf{E}[C]\right)
\end{aligned}
$$

where S is a co-occurrence matrix of the spam.

A random matrix: Word Co-Occurrence Matrix

A random matrix: Word Co-Occurrence Matrix

Proposition

Let C be the $N \times N$ matrix with $C_{ij} = \textit{Co-occurrence}(Word_i, Word_j)$.

$$
\begin{array}{lrcl}
(\dagger) & \mathbf{E}[C] & = & \mathbf{P}(Ham)\,\vec{P}_{Ham}\vec{P}_{Ham}^{T} + \mathbf{P}(Spam)\,S_{Spam} \\
i.e. & C & = & \mathbf{P}(Ham)\,\vec{P}_{Ham}\vec{P}_{Ham}^{T} + \mathbf{P}(Spam)\,S_{Spam} + (C - \mathbf{E}[C])
\end{array}
$$

where S is a co-occurrence matrix of the spam.

A random matrix: Word Co-Occurrence Matrix

Proposition

Let C be the $N \times N$ matrix with $C_{ij} = \text{Co-occurrence}(Word_i, Word_j)$.

$$\begin{array}{lrcl} (\dagger) & \mathbf{E}[C] & = & \mathbf{P}(Ham)\, \vec{P}_{Ham} \vec{P}_{Ham}^T + \mathbf{P}(Spam)\, S_{Spam} \\ i.e. & C & = & \mathbf{P}(Ham)\, \vec{P}_{Ham} \vec{P}_{Ham}^T + \mathbf{P}(Spam)\, S_{Spam} + (C - \mathbf{E}[C]) \end{array}$$

where S is a co-occurrence matrix of the spam.

Problem:

Can we recover the $\vec{P}_{Ham}$ from observation of the random matrix C?

A random matrix: Word Co-Occurrence Matrix

Proposition

Let C be the $N \times N$ matrix with $C_{ij} = \textit{Co-occurrence}(\textit{Word}_i, \textit{Word}_j)$.

$$
\begin{aligned}
(\dagger) \quad \mathbf{E}[C] &= \mathbf{P}(\textit{Ham})\, \vec{P}_{\textit{Ham}} \vec{P}_{\textit{Ham}}^{T} + \mathbf{P}(\textit{Spam})\, S_{\textit{Spam}} \\
\textit{i.e.} \quad C &= \mathbf{P}(\textit{Ham})\, \vec{P}_{\textit{Ham}} \vec{P}_{\textit{Ham}}^{T} + \mathbf{P}(\textit{Spam})\, S_{\textit{Spam}} + (C - \mathbf{E}[C])
\end{aligned}
$$

where S is a co-occurrence matrix of the spam.

Problem:
Can we recover the $\vec{P}_{Ham}$ from observation of the random matrix C?
Solution Idea:
Find the closest rank 1 approximation to C, i.e. the top eigenvector of C ("Principle component analysis")

A random matrix: Word Co-Occurrence Matrix

Proposition

Let C be the $N \times N$ matrix with $C_{ij} = \text{Co-occurrence}(Word_i, Word_j)$.

$$
\begin{aligned}
(\dagger) \quad \mathbf{E}[C] &= \mathbf{P}(Ham)\,\vec{P}_{Ham}\vec{P}_{Ham}^{T} + \mathbf{P}(Spam)\,S_{Spam} \\
i.e. \quad C &= \mathbf{P}(Ham)\,\vec{P}_{Ham}\vec{P}_{Ham}^{T} + \mathbf{P}(Spam)\,S_{Spam} + (C - \mathbf{E}[C])
\end{aligned}
$$

where S is a co-occurrence matrix of the spam.

Problem:
Can we recover the $\vec{P}_{Ham}$ from observation of the random matrix C?
Solution Idea:
Find the closest rank 1 approximation to C, i.e. the top eigenvector of C ("Principle component analysis")
Issues:
For the random matrix C, is the largest eigenvector close to $\vec{P}_{Ham}$? Is the largest eigenvalue close $\mathbf{P}(Ham)$? What is the effect of the noise?

Spiked Random Matrix - Toy Model #1

Definition

For a signal parameter$\lambda > 0$, and a noise parameter $\sigma^2 > 0$ the spiked GOE (one spike) is the random matrix

$$(\dagger\dagger) \qquad X^{\lambda,\sigma} = \lambda \vec{v}\vec{v}^T + \frac{1}{\sqrt{N}} G^{\sigma^2}$$

where $\vec{v} \in \mathbb{R}^N$, $\|\vec{v}\| = 1$ and $G^{\sigma^2} \sim GOE_N(\sigma^2)$ is the $N \times N$ symmetric matrix whose entries are independent Gaussian $G_{ij}^{\sigma^2} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$.

Spiked Random Matrix - Toy Model #1

Definition

For a signal parameter$\lambda > 0$, and a noise parameter $\sigma^2 > 0$ the spiked GOE (one spike) is the random matrix

$$(\dagger\dagger) \qquad X^{\lambda,\sigma} = \lambda \vec{v}\vec{v}^T + \frac{1}{\sqrt{N}} G^{\sigma^2}$$

where $\vec{v} \in \mathbb{R}^N$, $\|\vec{v}\| = 1$ and $G^{\sigma^2} \sim GOE_N(\sigma^2)$ is the $N \times N$ symmetric matrix whose entries are independent Gaussian $G_{ij}^{\sigma^2} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$.

This is akin to:

$$(\dagger) \qquad C = \mathbf{P}\,(Ham)\,\vec{P}_{Ham}\vec{P}_{Ham}^T + \mathbf{P}\,(Spam)\,S + \mathcal{N}$$

Spiked Random Matrix

Definition

Signal parameter$\lambda > 0$, Noise parameter $\sigma^2 > 0$, $\vec{v} \in \mathbb{R}^N$, $\|\vec{v}\| = 1$ and $G^{\sigma^2} \sim GOE_N(\sigma^2)$:

$$X^{\lambda,\sigma} = \lambda \vec{v}\vec{v}^T + \frac{1}{\sqrt{N}} G^{\sigma^2} \qquad (\dagger\dagger)$$

Problem: Given the spiked random matrix $X^{\lambda,\sigma}$ can you recover the vector $\vec{v}$ and the size of the spike λ?

Spiked Random Matrix

Definition

Signal parameter$\lambda > 0$, Noise parameter $\sigma^2 > 0$, $\vec{v} \in \mathbb{R}^N$, $\|\vec{v}\| = 1$ and $G^{\sigma^2} \sim GOE_N(\sigma^2)$:

$$(\dagger\dagger) \qquad X^{\lambda,\sigma} = \lambda \vec{v}\vec{v}^T + \frac{1}{\sqrt{N}} G^{\sigma^2}$$

Problem: Given the spiked random matrix $X^{\lambda,\sigma}$ can you recover the vector $\vec{v}$ and the size of the spike λ?

Proposition (Féral-Péché '06)

As $N \to \infty$, the top eigenvalue behaves like:

Spiked Random Matrix

Definition

Signal parameter$\lambda > 0$, Noise parameter $\sigma^2 > 0$, $\vec{v} \in \mathbb{R}^N$, $\|\vec{v}\| = 1$ and $G^{\sigma^2} \sim GOE_N(\sigma^2)$:

$$(\dagger\dagger) \qquad X^{\lambda,\sigma} = \lambda \vec{v}\vec{v}^T + \frac{1}{\sqrt{N}} G^{\sigma^2}$$

Problem: Given the spiked random matrix $X^{\lambda,\sigma}$ can you recover the vector $\vec{v}$ and the size of the spike λ?

Proposition (Féral-Péché '06)

As $N \to \infty$, the top eigenvalue behaves like:

Case I - High signal-to-noise: $\lambda > \sigma$:

Top eigenvalue of $X^{\lambda,\sigma}$ can recover λ with $\mathbf{E}[\lambda_{TOP}] = \lambda + \frac{1}{\lambda}\sigma^2$

Spiked Random Matrix

Definition

Signal parameter$\lambda > 0$, Noise parameter $\sigma^2 > 0$, $\vec{v} \in \mathbb{R}^N$, $\|\vec{v}\| = 1$ and $G^{\sigma^2} \sim GOE_N(\sigma^2)$:

$$(\dagger\dagger) \qquad X^{\lambda,\sigma} = \lambda \vec{v}\vec{v}^T + \frac{1}{\sqrt{N}} G^{\sigma^2}$$

Problem: Given the spiked random matrix $X^{\lambda,\sigma}$ can you recover the vector $\vec{v}$ and the size of the spike λ?

Proposition (Féral-Péché '06)

As $N \to \infty$, the top eigenvalue behaves like:

Case I - High signal-to-noise: $\lambda > \sigma$:

Top eigenvalue of $X^{\lambda,\sigma}$ can recover λ with $\mathbf{E}[\lambda_{TOP}] = \lambda + \frac{1}{\lambda}\sigma^2$

Case II - Low-signal-to-noise: $\lambda < \sigma$:

Top eigenvalue of $X^{\lambda,\sigma}$ indistinguishable from pure noise case $\lambda = 0$.

Picture Proof (Hermitian Matrix Case)

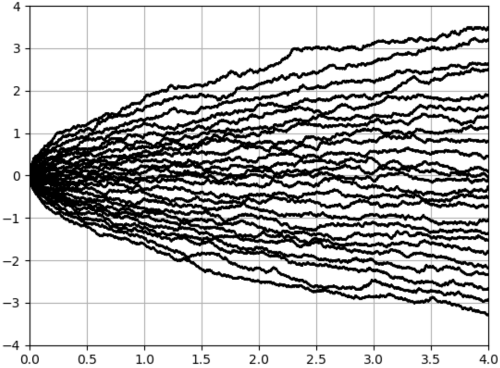
Create a *dynamic* matrix, $G(t)$, where each entry is a Brownian motion

Picture Proof (Hermitian Matrix Case)

Create a *dynamic* matrix, $G(t)$, where each entry is a Brownian motion The eigenvalues of $\frac{1}{\sqrt{N}}G(t)$ evolve like a *Dyson Brownian Motion* a.k.a. Non-intersecting Brownian Motions.

Picture Proof (Hermitian Matrix Case)

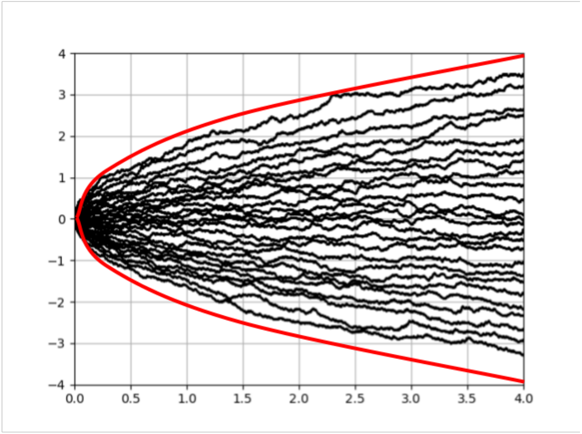
Create a *dynamic* matrix, $G(t)$, where each entry is a Brownian motion The eigenvalues of $\frac{1}{\sqrt{N}}G(t)$ evolve like a *Dyson Brownian Motion* a.k.a. Non-intersecting Brownian Motions.



At time t, distribution is $\approx$semi-circle distribution on $[-2\sqrt{t}, 2\sqrt{t}]$.
We look at $t = \sigma^2$.

Picture Proof (Hermitian Matrix Case)

Create a *dynamic* matrix, $G(t)$, where each entry is a Brownian motion. The eigenvalues of $\frac{1}{\sqrt{N}}G(t)$ evolve like a *Dyson Brownian Motion* a.k.a. Non-intersecting Brownian Motions.



At time t, distribution is $\approx$semi-circle distribution on $[-2\sqrt{t}, 2\sqrt{t}]$.
We look at $t = \sigma^2$

The eigenvalues of:

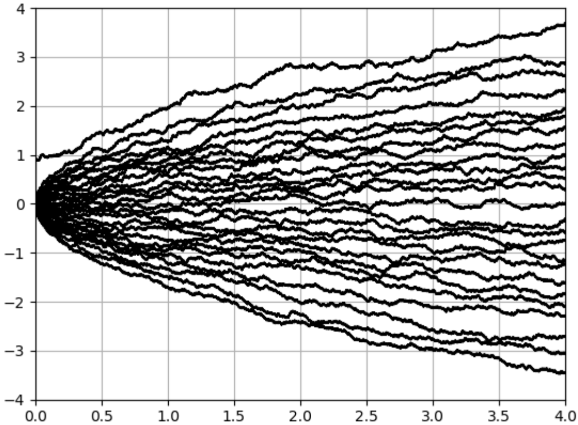
$$X^\lambda(t) = \lambda \vec{v}\vec{v}^T + \frac{1}{\sqrt{N}} G(t)$$

have the same evolution, but different initial condition!

The eigenvalues of:

$$X^{\lambda}(t) = \lambda \vec{v}\vec{v}^{T} + \frac{1}{\sqrt{N}} G(t)$$

have the same evolution, but different initial condition! e.g. $\lambda = 1, N = 25$.



The eigenvalues of:

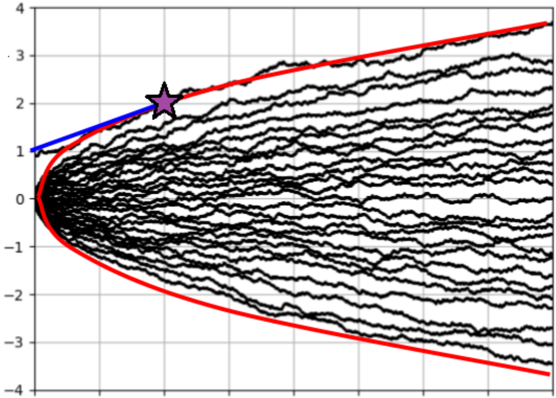
$$X^{\lambda}(t) = \lambda \vec{v}\vec{v}^{T} + \frac{1}{\sqrt{N}} G(t)$$

have the same evolution, but different initial condition!

The eigenvalues of:

$$X^{\lambda}(t) = \lambda \vec{v}\vec{v}^{T} + \frac{1}{\sqrt{N}} G(t)$$

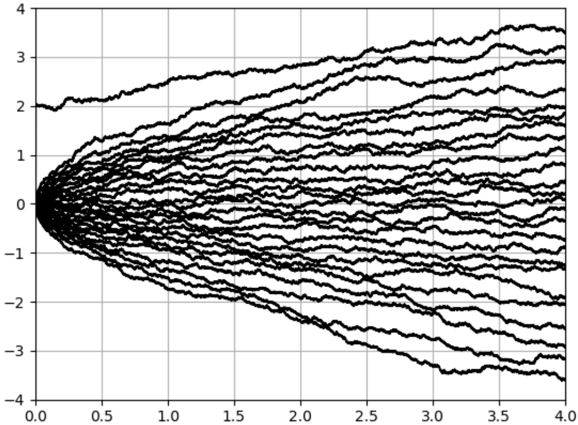
have the same evolution, but different initial condition! e.g. $\lambda = 1, N = 25$. Low-signal-to-noise: Indistinguishable from pure noise at $t = \sigma^2 = 4$!



The eigenvalues of:

$$X^{\lambda}(t) = \lambda \vec{v}\vec{v}^{T} + \frac{1}{\sqrt{N}} G(t)$$

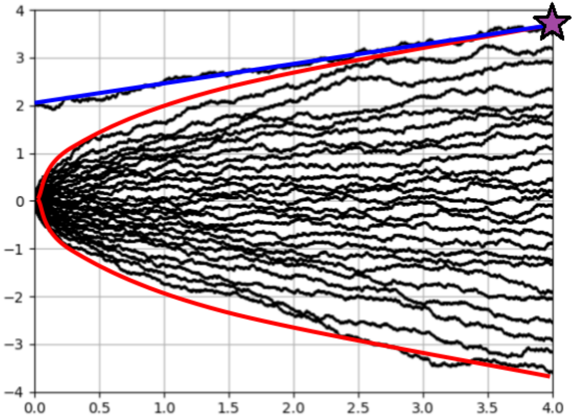
have the same evolution, but different initial condition! e.g. $\lambda = 2, N = 25$.



The eigenvalues of:

$$X^{\lambda}(t) = \lambda \vec{v}\vec{v}^{T} + \frac{1}{\sqrt{N}} G(t)$$

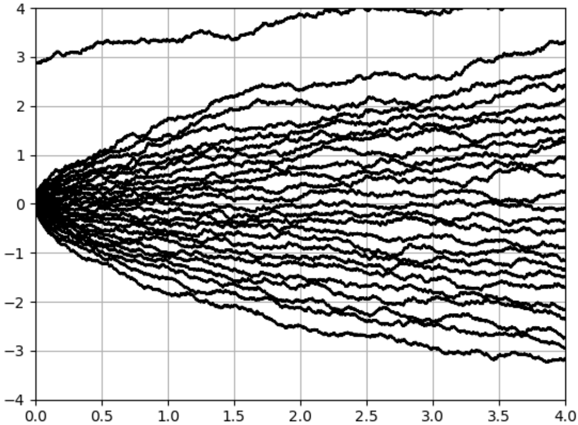
have the same evolution, but different initial condition! e.g. $\lambda = 2, N = 25$. Critical value of λ for $t = \sigma^2 = 4$.



The eigenvalues of:

$$X^\lambda(t) = \lambda \vec{v}\vec{v}^T + \frac{1}{\sqrt{N}} G(t)$$

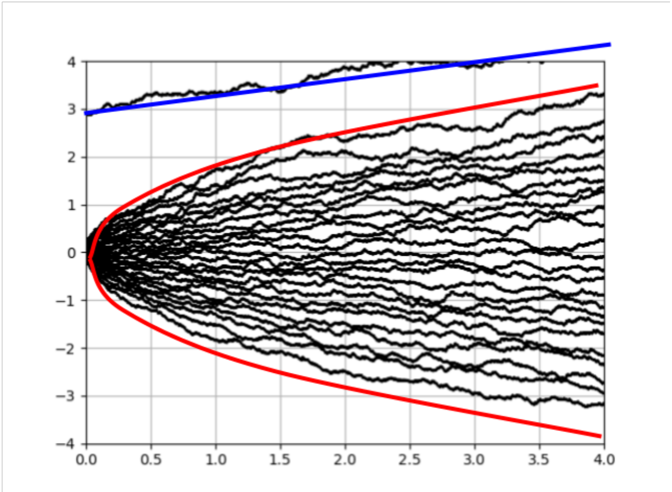
have the same evolution, but different initial condition! e.g. $\lambda = 3, N = 25$.



The eigenvalues of:

$$X^{\lambda}(t) = \lambda \vec{v}\vec{v}^{T} + \frac{1}{\sqrt{N}} G(t)$$

have the same evolution, but different initial condition! e.g. $\lambda = 3, N = 25$. High-signal-to-noise at $t = \sigma^2 = 4$.



Part 3: Functions on the Sphere: Random Tensors

Co-Occurrence Tensor, $k \geq 3$

Empirical Co-Occurrence for k-tuples of words:

$$\begin{aligned}&\text{Empirical Co-occurance}\left(Word_{i_1}, Word_{i_2}, \ldots Word_{i_k}\right) \\ = \;&\hat{\mathsf{P}}\left(k \text{ uniform words from a doc are } Word_{i_1}, Word_{i_2}, \ldots, Word_{i_k}\right)\end{aligned}$$

Co-Occurrence Tensor, $k \geq 3$

Empirical Co-Occurrence for k-tuples of words:

$$\begin{aligned} & \text{Empirical Co-occurance}\left(Word_{i_1}, Word_{i_2}, \ldots Word_{i_k}\right) \\ = \ & \hat{\mathbf{P}}\left(k \text{ uniform words from a doc are } Word_{i_1}, Word_{i_2}, \ldots, Word_{i_k}\right) \end{aligned}$$

Proposition

Let $C^{(k)}$ be the co-Occurrence k-tensor. Recall the topic probabilities $\mathbf{P}\left(Topic_j\right)$ and the probability vectors $\vec{P}_{Topic_j}$. Then:

$$C^{(k)} = \mathbf{P}\left(Ham\right) \vec{P}_{Ham}^{\otimes k} + \mathbf{P}\left(Spam\right) S_{Spam} + \mathcal{N}$$

Can we recover $\vec{P}_{Ham}$ from the noisy observation $C^{(k)}$?

Random Tensor Model

Definition

The spiked Gaussian k-*tensor* is:

$$(\dagger\dagger\dagger) \qquad Y^{(k)} = \lambda \vec{v}^{\otimes k} + \frac{1}{\sqrt{N}} G^{(k)}$$

where $G^{(k)}$ is the symmetric k tensor whose entries are $G^{(k)}_{i_1 \dots i_k} \overset{iid}{\sim} \mathcal{N}(0,1)$

Random Tensor Model

Definition

The spiked Gaussian k-<u>*tensor*</u> is:

$$(\dagger\dagger\dagger) \qquad Y^{(k)} = \lambda \vec{v}^{\otimes k} + \frac{1}{\sqrt{N}} G^{(k)}$$

where $G^{(k)}$ is the <u>symmetric</u> k tensor whose entries are $G^{(k)}_{i_1 \dots i_k} \overset{iid}{\sim} \mathcal{N}(0,1)$

Definition

The "energy landscape" of $Y^{(k)}$ is the function $f : \mathbb{R}^N \to \mathbb{R}$:

$$f(\vec{u}) \;=\; \left\langle Y^{(k)}, \vec{u}^{\otimes k} \right\rangle = \lambda \left\langle \vec{v}, \vec{u} \right\rangle^k + H_k(\vec{u})$$

where $H_k(\vec{u})$ is a random degree k polynomial:

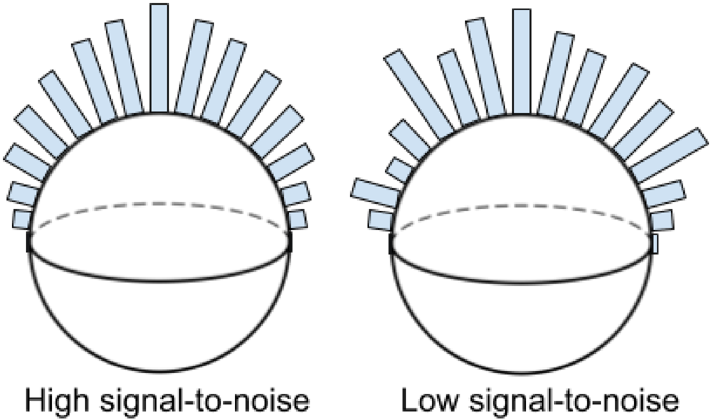
$$H_k(u) = \frac{1}{\sqrt{N}} \sum_{i_1,\dots,i_k=1}^{N} G^{(k)}_{i_1 \dots i_k} u_{i_1} \dots u_{i_k}$$

Problem: (Recall $f(\vec{u}) = \left\langle Y^{(k)}, \vec{u}^{\otimes k} \right\rangle = \lambda \left\langle \vec{v}, \vec{u} \right\rangle^k + H_k(\vec{u})$) "Tensor principle component analysis" is the maximization problem:

$$\begin{array}{rc} \text{maximize} & f(\vec{u}) \\ \text{subject to} & \|\vec{u}\| = 1 \end{array}$$

Problem: (Recall $f(\vec{u}) = \langle Y^{(k)}, \vec{u}^{\otimes k} \rangle = \lambda \langle \vec{v}, \vec{u} \rangle^k + H_k(\vec{u})$) "Tensor principle component analysis" is the maximization problem:

$$\begin{aligned} \text{maximize} \quad & f(\vec{u}) \\ \text{subject to} \quad & \|\vec{u}\| = 1 \end{aligned}$$



Problem: (Recall $f(\vec{u}) = \left\langle Y^{(k)}, \vec{u}^{\otimes k} \right\rangle = \lambda \left\langle \vec{v}, \vec{u} \right\rangle^k + H_k(\vec{u})$) "Tensor PCA" is the maximization problem:

$$\begin{array}{rc} \text{maximize} & f(\vec{u}) \\ \text{subject to} & \|\vec{u}\| = 1 \end{array}$$

Problem: (Recall $f(\vec{u}) = \left\langle Y^{(k)}, \vec{u}^{\otimes k} \right\rangle = \lambda \left\langle \vec{v}, \vec{u} \right\rangle^k + H_k(\vec{u})$) "Tensor PCA" is the maximization problem:

$$\begin{array}{ll} \text{maximize} & f(\vec{u}) \\ \text{subject to} & \|\vec{u}\| = 1 \end{array}$$

More general things we can look at: The set of all critical points. The set of all local maxima.

Problem: (Recall $f(\vec{u}) = \left\langle Y^{(k)}, \vec{u}^{\otimes k} \right\rangle = \lambda \left\langle \vec{v}, \vec{u} \right\rangle^k + H_k(\vec{u})$) "Tensor PCA" is the maximization problem:

$$\begin{aligned} \text{maximize} \quad & f(\vec{u}) \\ \text{subject to} \quad & \|\vec{u}\| = 1 \end{aligned}$$

More general things we can look at: The set of all critical points. The set of all local maxima.

Questions we can ask: Where are the critical points located? (i.e. how far are they from $\vec{v}$) What is the energy value of the critical points?

Problem: (Recall $f(\vec{u}) = \left\langle Y^{(k)}, \vec{u}^{\otimes k} \right\rangle = \lambda \langle \vec{v}, \vec{u} \rangle^k + H_k(\vec{u})$) "Tensor PCA" is the maximization problem:

$$\begin{aligned} \text{maximize} \quad & f(\vec{u}) \\ \text{subject to} \quad & \|\vec{u}\| = 1 \end{aligned}$$

More general things we can look at: The set of all critical points. The set of all local maxima.

Questions we can ask: Where are the critical points located? (i.e. how far are they from $\vec{v}$) What is the energy value of the critical points?

Theorem 1 (Ben Arous, Mei, Montenari, N. 2017)

Let $M \subset (-1, 1)$ and let $E \subset \mathbb{R}$. Let $Crt_{N,*}(M, E)$ be the number of critical points of the function $\vec{f}(\cdot)$ that have $\langle \vec{u}, \vec{v} \rangle \in M$ and $\vec{f}(\vec{u}) \in E$. Then:

$$\mathbf{E}\left[Crt_{N,*}(M, E)\right] \approx \exp\left(N \sup_{m \in M, e \in E} S_*(m, e) \right)$$

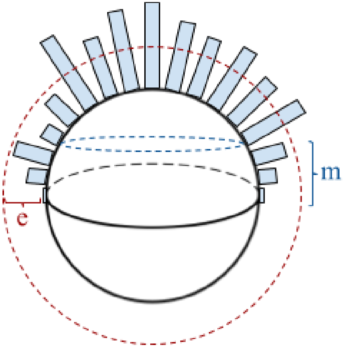
where S_* is a nasty but explicit function.

Theorem 1 (Ben Arous, Mei, Montenari, N. 2017)

Let $M \subset (-1, 1)$ and let $E \subset \mathbb{R}$. Let $Crt_{N,*}(M, E)$ be the number of critical points of the function $\vec{f}(\cdot)$ that have $\langle \vec{u}, \vec{v} \rangle \in M$ and $\vec{f}(\vec{u}) \in E$. Then:

$$\mathbf{E}\left[Crt_{N,*}(M, E)\right] \approx \exp\left(N \sup_{m \in M, e \in E} S_*\left(m, e\right)\right)$$

where S_* is a nasty but explicit function.

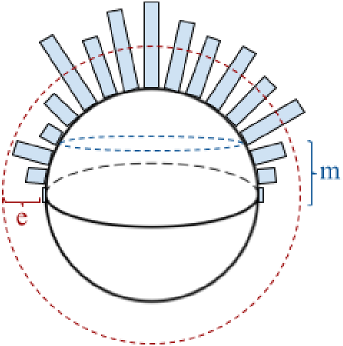


Theorem 2 (Ben Arous, Mei, Montenari, N. 2017)

Let $M \subset (-1,1)$ and let $E \subset \mathbb{R}$. Let $Crt_{N,0}(M,E)$ be the number of local maxima of the function $\vec{f}(\cdot)$ that have $\langle \vec{u}, \vec{v} \rangle \in M$ and $\vec{f}(\vec{u}) \in E$. Then:

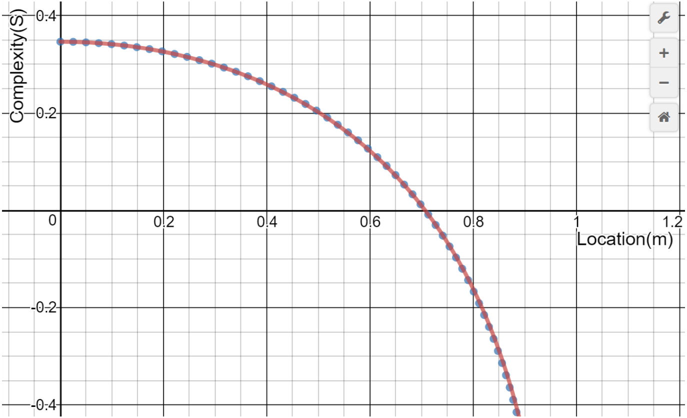
$$\mathbf{E}\left[Crt_{N,0}(M,E)\right] \approx \exp\left(N \sup_{m \in M, e \in E} S_0\left(m, e\right)\right)$$

where S_0 is a nasty but explicit function.



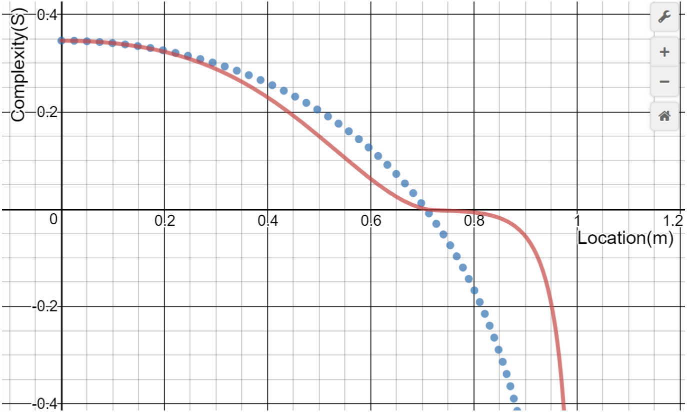
Location of Critical Points

$S_*(m) = \max_{e\in\mathbb{R}} S_*(e,x)$ is the exponential growth rate of # of critical points at $\langle \vec{u}, \vec{v}\rangle = m$.
(`https://www.desmos.com/calculator/lez8qptvu1`) Signal-to-noise $\lambda = 0.1$



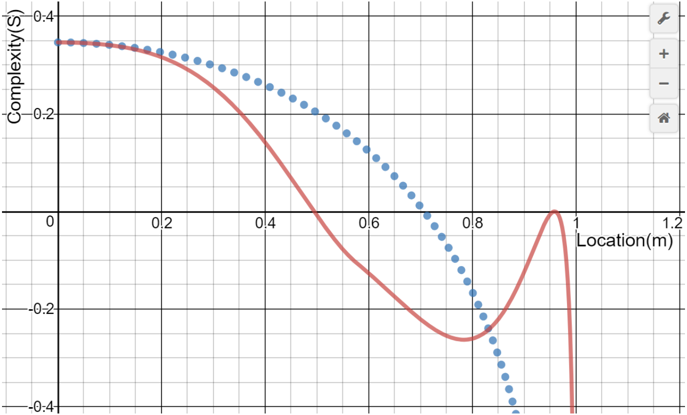
Location of Critical Points

$S_*(m) = \max_{e \in \mathbb{R}} S_*(e, x)$ is the exponential growth rate of # of critical points at $\langle \vec{u}, \vec{v} \rangle = m$.
(`https://www.desmos.com/calculator/lez8qptvu1`) S/N $\lambda = 0.75$



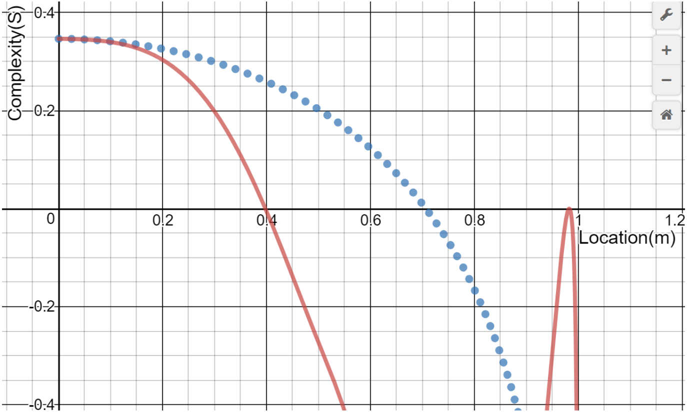
Location of Critical Points

$S_*(m) = \max_{e \in \mathbb{R}} S_*(e, x)$ is the exponential growth rate of # of critical points at $\langle \vec{u}, \vec{v} \rangle = m$.
(`https://www.desmos.com/calculator/lez8qptvu1`) S/N $\lambda = 1.5$



Location of Critical Points

$S_*(m) = \max_{e \in \mathbb{R}} S_*(e, x)$ is the exponential growth rate of # of critical points at $\langle \vec{u}, \vec{v} \rangle = m$.
(`https://www.desmos.com/calculator/lez8qptvu1`) S/N $\lambda = 2.25$



Location of Critical Points - Results

Corollary

Let:

$$\lambda_c := \sqrt{\frac{1}{2k}\frac{(k-1)^{(k-1)}}{(k-2)^{(k-2)}}}$$

If $\lambda < \lambda_c$ then there are no "good" critical points. If $\lambda > \lambda_c$ then $S_*(m) = 0$ at the point where:

$$m^{2k-4}\left(1-m^2\right) = \frac{1}{2k\lambda^2}$$

Proof Ideas - Kac-Rice Formula

Main Idea, the Kac-Rice formula $f : [a, b] \to \mathbb{R}$

$$\# \text{ zeros of } f \text{ on } [a, b] = \lim_{\epsilon \to 0} \int_a^b |f'(x)| \cdot \delta_\epsilon (f(x)) \, \mathrm{d}x$$

where $\delta_\epsilon(x)$ is an approximate dirac-delta, $\lim_{\epsilon \to 0} \int \delta_\epsilon(x) g(x) = g(0)$.

Proof Ideas

We count zeros of ∇f, so Kac-Rice formula gives:

$$\mathbf{E}\left[Crt_*(M,E)\right] = \int_{\{\vec{u}:\langle\vec{u},\vec{v}\rangle\in M\}} \mathbf{E}\left[\left|\det\left(Hess\ f(\vec{u})\right)\right| \cdot 1\left\{f(\vec{u})\in E\right\} \,\middle|\, \nabla\vec{f}(\vec{u}) = 0\right]$$

The Hessian $Hess(f(\vec{u}))$ is a spiked random matrix!

$$Hess\left(f(\vec{u})\right) \sim k(k-1)\lambda m^{k-2}(1-m^2)\vec{e}_1\vec{e}_1^T + GOE_{N-1}$$

Use "large deviations" for spiked random matrix to evaluate $\mathbf{E}\left[Crt_*(M,E)\right]$

Large Deviation Results Used

The spectrum of the GOE concentrates around the semi-circle law:

$$\mathbf{P}\left(L_n \notin B(\sigma_{SC}, \epsilon)\right) \approx \exp\left(-N^2 C(\epsilon)\right)$$

For the GOE:

$$\begin{aligned}
\mathbf{P}\left(\lambda_{max}^{GOE} \geq t\right) &\approx \exp\left(-N I_1(t)\right) \\
I_1(t) &= \begin{cases} \int_2^t \sqrt{\left(\frac{y}{2}\right)^2 - 1}\,\mathrm{d}y & t \geq 2 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

For the spiked GOE $X = \theta e_1 e_1^T + GOE_N$ we have [Maida '07]:

$$\begin{aligned}
\mathbf{P}\left(\lambda_{max}^{\theta-GOE} \leq t\right) &\approx \exp\left(-N L(\theta, t)\right) \\
L(\theta, t) &= \begin{cases} \int_{\theta+\frac{1}{\theta}}^t \sqrt{(\frac{1}{2}y)^2 - 1}\,\mathrm{d}y - \frac{\theta}{2}\left[t - \left(\theta + \frac{1}{\theta}\right)\right] \\ +\frac{1}{8}\left[t^2 - \left(\theta + \frac{1}{\theta}\right)^2\right] & 2 \leq t < \\ \infty & t < 2 \\ 0 & \text{otherwi} \end{cases}
\end{aligned}$$

Formulas for S_* and S_0

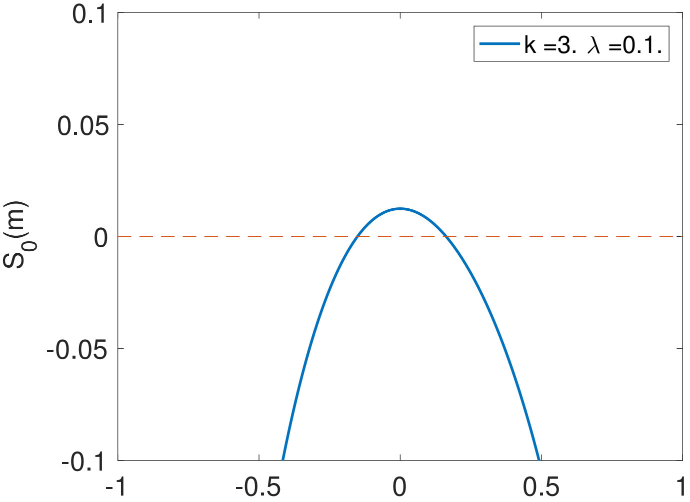
Recall $\mathbf{P}\left(\lambda_{max}^{GOE} \geq t\right) \approx \exp\left(-NI_1(t)\right)$ and
$\mathbf{P}\left(\lambda_{max}^{\theta-GOE} \leq t\right) \approx \exp\left(-NL(\theta, t)\right)$

$$
\begin{aligned}
S_*(m,x) &= \frac{1}{2}\log(k-1) + \frac{1}{2}\log(1-m^2) \\
&\quad -k\lambda^2 m^{2k-2}(1-m^2) - (x-\lambda m^k)^2 \\
&\quad +\frac{k}{2(k-1)}x^2 + I_1\left(\left|\sqrt{\frac{2k}{k-1}}t\right|\right) \\
S_0(m,x) &= S_*(m,x) - L\left(\sqrt{2k(k-1)}\lambda m^{k-2}(1-m^2), \sqrt{\frac{2k}{k-1}}x\right)
\end{aligned}
$$

Location of Local Maxima

Definition

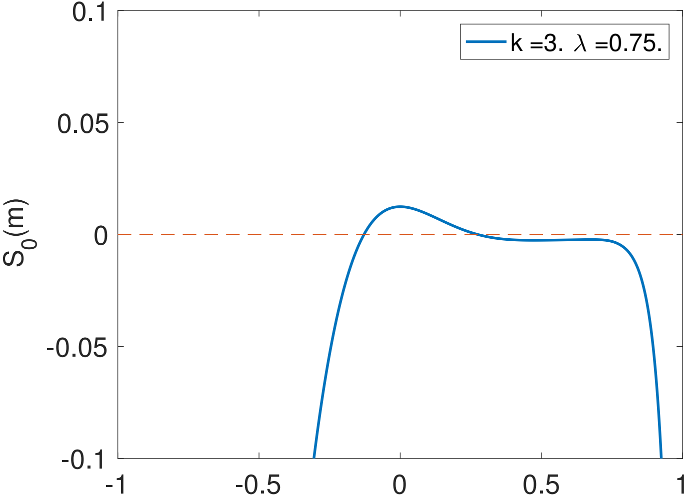
Define $S_0(m) = \max_{e \in \mathbb{R}} S_0(m, e)$ to be the exponential growth rate of the number of local maxima with $\langle \vec{u}, \vec{v} \rangle = m$.



Location of Local Maxima

Definition

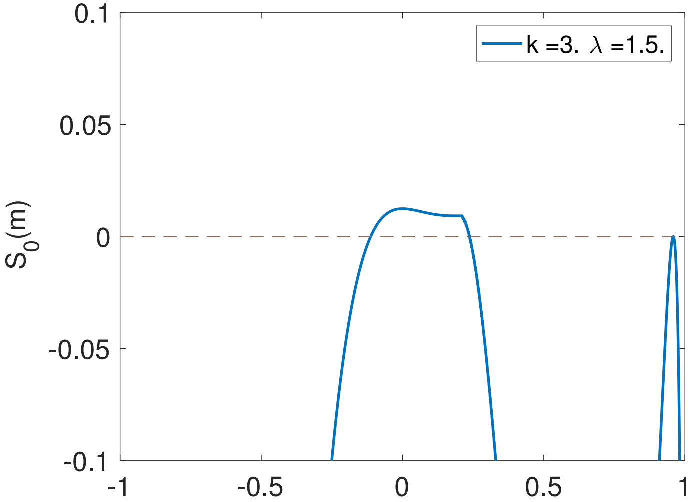
Define $S_0(m) = \max_{x \in \mathbb{R}} S_0(m, x)$ to be the exponential growth rate of the number of local maxima with $\langle \vec{u}, \vec{v} \rangle = m$.



Location of Local Maxima

Definition

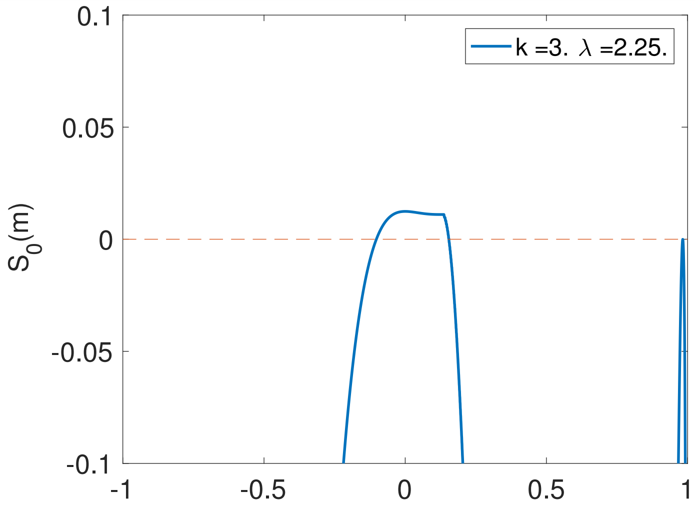
Define $S_0(m) = \max_{x \in \mathbb{R}} S_0(m, x)$ to be the exponential growth rate of the number of local maxima with $\langle \vec{u}, \vec{v} \rangle = m$.



Location of Local Maxima

Definition

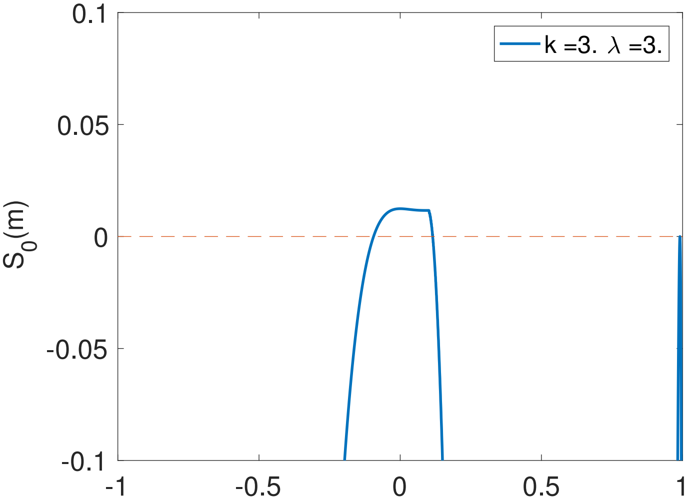
Define $S_0(m) = \max_{x \in \mathbb{R}} S_0(m, x)$ to be the exponential growth rate of the number of local maxima with $\langle \vec{u}, \vec{v} \rangle = m$.



Location of Local Maxima

Definition

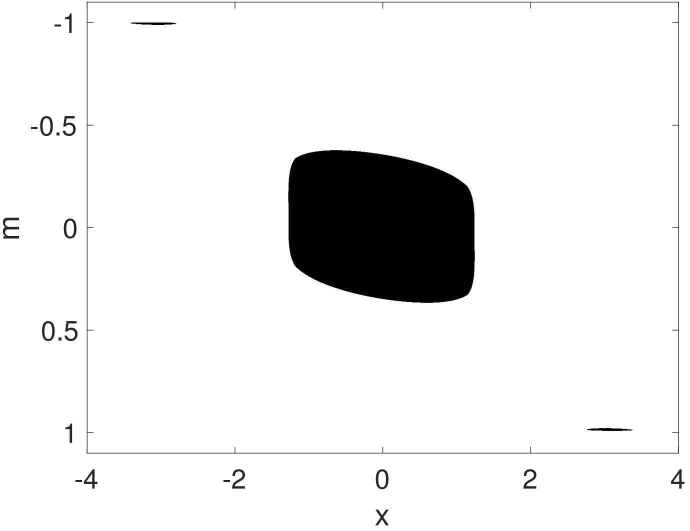
Define $S_0(m) = \max_{x \in \mathbb{R}} S_0(m, x)$ to be the exponential growth rate of the number of local maxima with $\langle \vec{u}, \vec{v} \rangle = m$.



Location and Energy

$S(m,e) \subset [-1,1] \times \mathbb{R}$ where there are exponential numerous critical points

$S_\star(m,x) > 0$ $(k=3, \lambda=3)$



Location and Energy

$S(m,e) \subset [-1,1] \times \mathbb{R}$ where there are exponential numerous local maxima: $S_0(m,x) > 0$ $(k=3, \lambda=3)$