Tensor Decompositions and Their Applications

Ankur Moitra (MIT)

Simons Institute Bootcamp Tutorial, Part 1
SPEARMAN’S HYPOTHESIS

Charles Spearman (1904): There are two types of intelligence, *eductive* and *reproductive*
SPEARMAN’S HYPOTHESIS

Charles Spearman (1904): There are two types of intelligence, *eductive* and *reproductive*

eductive (adj): the ability to make sense out of complexity
reproductive (adj): the ability to store and reproduce information
SPEARMAN’S HYPOTHESIS

Charles Spearman (1904): There are two types of intelligence, *eductive* and *reproductive*.

To test this theory, he invented **Factor Analysis:**

$$M \approx AB^T$$

- **eductive** (adj): the ability to make sense out of complexity
- **reproductive** (adj): the ability to store and reproduce information
SPEARMAN’S HYPOTHESIS

Charles Spearman (1904): There are two types of intelligence, *eductive* and *reproductive*.

To test this theory, he invented **Factor Analysis**:

```
M ≈ A B^T
```

students (1000)  inner-dimension (2)

tests (10)

eductive (adj): the ability to make sense out of complexity
reproductive (adj): the ability to store and reproduce information
Given:  \[ M = \sum a_i \otimes b_i \]

\[ = AB^\top \]

“correct” factors
Given: \[ M = \sum a_i \otimes b_i \]
\[ = AB^\top \]

“correct” factors

When can we find the factors \( \{a_i\} \) and \( \{b_i\} \) uniquely?
Given: \[ M = \sum a_i \otimes b_i \]

\[ = AB^\top = \left( AR \right) \left( R^{-1} B^\top \right) \]

“correct” factors

alternative factorization

When can we find the factors \( \{a_i\} \) and \( \{b_i\} \) uniquely?
Given:  \[ M = \sum a_i \otimes b_i \]

\[ = AB^\top = (AR) \left( R^{-1} B^\top \right) \]

"correct" factors                        alternative factorization

When can we find the factors \( \{a_i\} \) and \( \{b_i\} \) uniquely?

Claim: The factors \( \{a_i\} \) and \( \{b_i\} \) are not determined uniquely unless we impose additional conditions on them
Given: \[ M = \sum a_i \otimes b_i \]

\[ = AB^\top = \left( AR \right) \left( R^{-1} B^\top \right) \]

“correct” factors alternative factorization

When can we find the factors \( \{ a_i \} \) and \( \{ b_i \} \) uniquely?

Claim: The factors \( \{ a_i \} \) and \( \{ b_i \} \) are not determined uniquely unless we impose additional conditions on them.

e.g. if \( \{ a_i \} \) and \( \{ b_i \} \) are orthogonal, or \( \text{rank}(M) = 1 \)
Given: \[ M = \sum a_i \otimes b_i \]

\[ = AB^\top = \left( AR \right) \left( R^{-1} B^\top \right) \]

“correct” factors alternative factorization

When can we find the factors \( \{ a_i \} \) and \( \{ b_i \} \) uniquely?

Claim: The factors \( \{ a_i \} \) and \( \{ b_i \} \) are not determined uniquely unless we impose additional conditions on them

e.g. if \( \{ a_i \} \) and \( \{ b_i \} \) are orthogonal, or \( \text{rank}(M) = 1 \)

This is called the \textbf{rotation problem}, and is a major issue in factor analysis and motivates the study of \textbf{tensor methods}...
OUTLINE

**Part I: Introduction**
- The Rotation Problem
- Jennrich’s Algorithm

**Part II: Applications**
- Phylogenetic Reconstruction
- Mixtures of Gaussians
- Orbit Retrieval
Part I: Introduction
- The Rotation Problem
- Jennrich’s Algorithm

Part II: Applications
- Phylogenetic Reconstruction
- Mixtures of Gaussians
- Orbit Retrieval
\[ M = a_1 \otimes b_1 + a_2 \otimes b_2 + \cdots + a_R \otimes b_R \]
\[ M = a_1 \otimes b_1 + a_2 \otimes b_2 + \cdots + a_R \otimes b_R \]

\[ T = a_1 \otimes b_1 \otimes c_1 + \cdots + a_R \otimes b_R \otimes c_R \]

(i, j, k) entry of \( x \otimes y \otimes z \) is \( x(i) \times y(j) \times z(k) \)
When are tensor decompositions unique?
When are tensor decompositions unique?

Theorem [Jennrich 1970]: Suppose \( \{a_i\} \) and \( \{b_i\} \) are linearly independent and no pair of vectors in \( \{c_i\} \) is a scalar multiple of each other...
When are tensor decompositions unique?

**Theorem [Jennrich 1970]:** Suppose \( \{a_i\} \) and \( \{b_i\} \) are linearly independent and no pair of vectors in \( \{c_i\} \) is a scalar multiple of each other. Then

\[
T = a_1 \otimes b_1 \otimes c_1 + \cdots + a_R \otimes b_R \otimes c_R
\]

is unique up to permuting the rank one terms and rescaling the factors.
When are tensor decompositions unique?

**Theorem [Jennrich 1970]:** Suppose \(\{a_i\}\) and \(\{b_i\}\) are linearly independent and no pair of vectors in \(\{c_i\}\) is a scalar multiple of each other. Then

\[
T = a_1 \otimes b_1 \otimes c_1 + \cdots + a_R \otimes b_R \otimes c_R
\]

is unique up to permuting the rank one terms and rescaling the factors.

Equivalently, the rank one factors are **unique**
When are tensor decompositions unique?

**Theorem [Jennrich 1970]:** Suppose \( \{a_i\} \) and \( \{b_i\} \) are linearly independent and no pair of vectors in \( \{c_i\} \) is a scalar multiple of each other. Then

\[
T = a_1 \otimes b_1 \otimes c_1 + \cdots + a_R \otimes b_R \otimes c_R
\]

is unique up to permuting the rank one terms and rescaling the factors.

Equivalently, the rank one factors are **unique**

There is a simple algorithm to compute the factors too!
JENNRICH’S ALGORITHM

Compute $T(\cdot, \cdot, x)$

i.e. add up matrix slices

$$\sum_i x_i T_i$$
JENNRICH’S ALGORITHM

Compute $T(\cdot, \cdot, x)$

i.e. add up matrix slices

$$\sum_i x_i T_i$$

If $T = a \otimes b \otimes c$ then $T(\cdot, \cdot, x) = \langle c, x \rangle a \otimes b$
JENNIRICH’S ALGORITHM

Compute \( T(\cdot, \cdot, x) = \sum \langle c_i, x \rangle a_i \otimes b_i \)

i.e. add up matrix slices

\[ \sum_i x_i T_i \]
JENNRIICH’S ALGORITHM

Compute \( T(\cdot, \cdot, x) = \sum \langle c_i, x \rangle a_i \otimes b_i \)

i.e. add up matrix slices

\[ \sum_i x_i T_i \]

\( (x \text{ is chosen uniformly at random from } \mathbb{S}^{n-1}) \)
JENNRICH’S ALGORITHM

Compute $T(\cdot, \cdot, x) = AD_x B^\top$

Diag($\{\langle c_i, x \rangle \}_i$)

i.e. add up matrix slices

\[
\sum_i x_i T_i
\]

(x is chosen uniformly at random from $\mathbb{S}^{n-1}$)
Jennrich’s Algorithm

Compute \( T(\cdot, \cdot, x) = AD_x B^\top \)
JENNHRICH’S ALGORITHM

Compute $T(\cdot, \cdot, x) = AD_x B^\top$

Compute $T(\cdot, \cdot, y) = AD_y B^\top$
JENNHRICH’S ALGORITHM

Compute \( T(\cdot, \cdot, x) = AD_x B^\top \)

Compute \( T(\cdot, \cdot, y) = AD_y B^\top \)

Diagonalize \( T(\cdot, \cdot, x) \left( T(\cdot, \cdot, y) \right)^{-1} \)
JENNRRICH’S ALGORITHM

1. Compute $T(\cdot, \cdot, x) = AD_xB^\top$
2. Compute $T(\cdot, \cdot, y) = AD_yB^\top$
3. Diagonalize $T(\cdot, \cdot, x) \left( T(\cdot, \cdot, y) \right)^{-1}$

\[ AD_xB^\top (B^\top)^{-1} D_y^{-1} A^{-1} \]
JENNRRICH’S ALGORITHM

Compute \( T(\cdot, \cdot, x) = AD_x B^\top \)

Compute \( T(\cdot, \cdot, y) = AD_y B^\top \)

Diagonalize \( T(\cdot, \cdot, x) \left( T(\cdot, \cdot, y) \right)^{-1} \)

\[
AD_x D_y^{-1} A^{-1}
\]
JENNRICH’S ALGORITHM

Compute \( T(\cdot, \cdot, x) = AD_x B^\top \)
Compute \( T(\cdot, \cdot, y) = AD_y B^\top \)
Diagonalize \( T(\cdot, \cdot, x) \left( T(\cdot, \cdot, y) \right)^{-1} \)

\[ AD_x D_y^{-1} A^{-1} \]

Claim: whp (over \( x,y \)) the eigenvalues are distinct, so the Eigendecomposition is unique and recovers \( \alpha_i \)
JENNRICH’S ALGORITHM

Compute $T(\cdot, \cdot, x) = AD_x B^\top$

Compute $T(\cdot, \cdot, y) = AD_y B^\top$

Diagonalize $T(\cdot, \cdot, x) \left( T(\cdot, \cdot, y) \right)^{-1}$
JENNRICH’S ALGORITHM

Compute $T(\cdot, \cdot, x) = AD_x B^\top$

Compute $T(\cdot, \cdot, y) = AD_y B^\top$

Diagonalize $T(\cdot, \cdot, x) \left( T(\cdot, \cdot, y) \right)^{-1}$

Diagonalize $T(\cdot, \cdot, y) \left( T(\cdot, \cdot, x) \right)^{-1}$
JENNIRICH’S ALGORITHM

Compute \( T(\cdot, \cdot, x) = AD_x B^\top \)

Compute \( T(\cdot, \cdot, y) = AD_y B^\top \)

Diagonalize \( T(\cdot, \cdot, x) \left( T(\cdot, \cdot, y) \right)^{-1} \)

Diagonalize \( T(\cdot, \cdot, y) \left( T(\cdot, \cdot, x) \right)^{-1} \)

Match up the factors (their eigenvalues are reciprocals) and find \( \{c_i\}_i \) by solving a linear syst.
Given: \[ M = \sum a_i \otimes b_i \]

When can we find the factors \( \{a_i\} \) and \( \{b_i\} \) uniquely?

Only possible if \( \{a_i\} \) and \( \{b_i\} \) are orthogonal, or \( \text{rank}(M) = 1 \)
Given: \[ M = \sum a_i \otimes b_i \]

When can we find the factors \( \{a_i\} \) and \( \{b_i\} \) uniquely?

Only possible if \( \{a_i\} \) and \( \{b_i\} \) are orthogonal, or \( \text{rank}(M) = 1 \)

Given: \[ T = \sum a_i \otimes b_i \otimes c_i \]

When can we find the factors \( \{a_i\} \), \( \{b_i\} \) and \( \{c_i\} \) uniquely?
Given: \( M = \sum a_i \otimes b_i \)

When can we find the factors \( \{a_i\} \) and \( \{b_i\} \) uniquely?

Only possible if \( \{a_i\} \) and \( \{b_i\} \) are orthogonal, or \( \text{rank}(M) = 1 \)

Given: \( T = \sum a_i \otimes b_i \otimes c_i \)

When can we find the factors \( \{a_i\} \), \( \{b_i\} \) and \( \{c_i\} \) uniquely?

**Jennrich:** If \( \{a_i\} \) and \( \{b_i\} \) are full rank and no pair in \( \{c_i\} \) are scalar multiples of each other
OUTLINE

Part I: Introduction

• The Rotation Problem
• Jennrich’s Algorithm

Part II: Applications

• Phylogenetic Reconstruction
• Mixtures of Gaussians
• Orbit Retrieval
OUTLINE

Part I: Introduction

• The Rotation Problem
• Jennrich’s Algorithm

Part II: Applications

• Phylogenetic Reconstruction
• Mixtures of Gaussians
• Orbit Retrieval
PHYLOGENETIC RECONSTRUCTION

“Tree of Life”

- Green circles represent extinct species.
- Red circles represent extant species.
PHYLOGENETIC RECONSTRUCTION

![Phylogenetic tree diagram](image)

- **x**
- **a**
- **b**
- **z**
- **c**
- **d**

- ○ = extinct
- ■ = extant
PHYLOGENETIC RECONSTRUCTION

root: $\pi : \sum \rightarrow \mathbb{R}^+$

“initial distribution”

- $\circ$ = extinct
- $\bullet$ = extant

$\sum$ = alphabet
PHYLOGENETIC RECONSTRUCTION

root: $\pi : \sum \rightarrow \mathbb{R}^+$

“initial distribution”

“conditional distribution”

$R_{z,b}$

$\sum = \text{alphabet}$

$\text{Extinct}$

$\text{Extant}$
In each sample, we observe a symbol ($\Sigma$) at each extant ($\bigcirc$) node where we sample from $\pi$ for the root, and propagate it using $R_{x,y}$, etc.
HIDDEN MARKOV MODELS

\[ x \rightarrow y \rightarrow z \]

\textcolor{green}{\textbullet} = \text{hidden}

\textcolor{red}{\textbullet} = \text{observed}

\( a \rightarrow b \rightarrow c \)
HIDDEN MARKOV MODELS

\[ \pi : \sum \rightarrow \mathbb{R}^+ \]
"
"initial distribution"

\[ x \]
\[ y \]
\[ z \]
\[ a \]
\[ b \]
\[ c \]

\[ \text{○} = \text{hidden} \]
\[ \text{●} = \text{observed} \]
HIDDEN MARKOV MODELS

\( \pi : \sum_S \rightarrow \mathbb{R}^+ \)

“initial distribution”

\( \pi(x) \)

“obs. matrices”

\( O_{x,a} \)

R_{x,y} “transition matrices”

x y z ...

\( a \quad b \quad c \)

○ = hidden

○ = observed
In each sample, we observe a symbol (\( \sum_0 \)) at each obs. (\( \bigcirc \)) node where we sample from \( \pi \) for the start, and propagate it using \( R_{x,y} \), etc (\( \sum_S \))
Can we reconstruct just the topology from random samples?
Can we reconstruct just the topology from random samples?

Usually, we assume $T_{x,y}$ etc are full rank so that we can re-root the tree arbitrarily.
Can we reconstruct just the topology from random samples?

Usually, we assume $T_{x,y}$, etc are full rank so that we can re-root the tree arbitrarily.

[Steel, 1994]: The following is a distance function on the edges

$$d_{x,y} = -\ln |\text{det}(P_{x,y})| + \frac{1}{2} \prod_{\sigma \text{ in } \Sigma} \pi_{x,\sigma} - \frac{1}{2} \prod_{\sigma \text{ in } \Sigma} \pi_{y,\sigma}$$

where $P_{x,y}$ is the joint distribution.
Can we reconstruct just the topology from random samples?

Usually, we assume $T_{x,y}$, etc are full rank so that we can re-root the tree arbitrarily

**[Steel, 1994]:** The following is a distance function on the edges

\[
    d_{x,y} = -\ln |\det(P_{x,y})| + \frac{1}{2} \prod_{\sigma \in \Sigma} \pi_{x,\sigma} - \frac{1}{2} \prod_{\sigma \in \Sigma} \pi_{y,\sigma}
\]

where $P_{x,y}$ is the joint distribution, and the distance between leaves is the sum of distances on the path in the tree.
Can we reconstruct just the topology from random samples?

Usually, we assume $T_{x,y}$, etc are full rank so that we can re-root the tree arbitrarily.

**[Steel, 1994]:** The following is a distance function on the edges

\[
d_{x,y} = -\ln |\det(P_{x,y})| + \frac{1}{2} \prod_{\sigma \text{ in } \Sigma} \pi_{x,\sigma} - \frac{1}{2} \prod_{\sigma \text{ in } \Sigma} \pi_{y,\sigma}
\]

where $P_{x,y}$ is the joint distribution, and the distance between leaves is the sum of distances on the path in the tree.

(It’s not even obvious it’s nonnegative!)
Can we reconstruct just the topology from random samples?

Usually, we assume $T_{x,y}$, etc are full rank so that we can re-root the tree arbitrarily
Can we reconstruct just the topology from random samples?

Usually, we assume $T_{x,y}$, etc are full rank so that we can re-root the tree arbitrarily.

[Erdoes, Steel, Szekely, Warnow, 1997]: Used Steel’s distance function and quartet tests

\[
\begin{array}{cc}
\text{OR} & \text{OR} \\
\text{OR} & \text{...}
\end{array}
\]

to reconstruction the topology
Can we reconstruct just the topology from random samples?

Usually, we assume $T_{x,y}$, etc are full rank so that we can re-root the tree arbitrarily.

[Erdoes, Steel, Szekely, Warnow, 1997]: Used Steel’s distance function and quartet tests

to reconstruction the topology, from polynomially many samples.
Can we reconstruct just the topology from random samples?

Usually, we assume $T_{x,y}$ etc are full rank so that we can re-root the tree arbitrarily.

**[Erdos, Steel, Szekely, Warnow, 1997]**: Used Steel’s distance function and quartet tests

![Diagram of tree topologies](image)

... to reconstruction the topology, from polynomially many samples

For many problems (e.g. HMMs) finding the transition matrices is the main issue...
[Chang, 1996]: The model is identifiable (if R’s are full rank)
[Chang, 1996]: The model is identifiable (if R’s are full rank)
[Chang, 1996]: The model is identifiable (if R’s are full rank)
[Chang, 1996]: The model is identifiable (if R’s are full rank)
Joint distribution over $(a, b, c)$:

$$
\sum_{\sigma} \mathbb{P}[z = \sigma] \mathbb{P}[a | z = \sigma] \otimes \mathbb{P}[b | z = \sigma] \otimes \mathbb{P}[c | z = \sigma]
$$

[Chang, 1996]: The model is identifiable (if R’s are full rank)
[Chang, 1996]: The model is identifiable (if R’s are full rank)

Joint distribution over \((a, b, c)\):

\[
\sum_{\sigma} \mathbb{P}(z = \sigma) \mathbb{P}(a \mid z = \sigma) \otimes \mathbb{P}(b \mid z = \sigma) \otimes \mathbb{P}(c \mid z = \sigma)
\]

columns of \(R_{z,b}\)
[Mossel, Roch, 2006]: There is an algorithm to PAC learn a phylogenetic tree or an HMM (if its transition/output matrices are full rank) from polynomially many samples.
[Mossel, Roch, 2006]: There is an algorithm to PAC learn a phylogenetic tree or an HMM (if its transition/output matrices are full rank) from polynomially many samples

Is the full-rank assumption necessary?
[Mossel, Roch, 2006]: There is an algorithm to PAC learn a phylogenetic tree or an HMM (if its transition/output matrices are full rank) from polynomially many samples.

Is the full-rank assumption necessary?

[Mossel, Roch, 2006]: It is as hard as noisy-parity to learn the parameters of a general HMM.
[Mossel, Roch, 2006]: There is an algorithm to PAC learn a phylogenetic tree or an HMM (if its transition/output matrices are full rank) from polynomially many samples.

Is the full-rank assumption necessary?

[Mossel, Roch, 2006]: It is as hard as noisy-parity to learn the parameters of a general HMM.

Noisy-parity is an infamous problem in learning, where $O(n)$ samples suffice but the best algorithms run in time $2^{n/log(n)}$.

Due to [Blum, Kalai, Wasserman, 2003]
There is an algorithm to PAC learn a phylogenetic tree or an HMM (if its transition/output matrices are full rank) from polynomially many samples.

Is the full-rank assumption necessary?

It is as hard as noisy-parity to learn the parameters of a general HMM.

Noisy-parity is an infamous problem in learning, where $O(n)$ samples suffice but the best algorithms run in time $2^{n/\log(n)}$.

Due to [Blum, Kalai, Wasserman, 2003]

(It’s now used as a hard problem to build cryptosystems!)
THE POWER OF CONDITIONAL INDEPENDENCE

[Phylogenetic Trees/HMMS]: (joint distribution on leaves a, b, c)

\[
\sum_{\sigma} \mathbb{P}[z = \sigma] \mathbb{P}[a | z = \sigma] \otimes \mathbb{P}[b | z = \sigma] \otimes \mathbb{P}[c | z = \sigma]
\]

following [Mossel, Roch, 2006]
OUTLINE

Part I: Introduction
  • The Rotation Problem
  • Jennrich’s Algorithm

Part II: Applications
  • Phylogenetic Reconstruction
  • Mixtures of Gaussians
  • Orbit Retrieval
OUTLINE

Part I: Introduction

• The Rotation Problem
• Jennrich’s Algorithm

Part II: Applications

• Phylogenetic Reconstruction
• Mixtures of Gaussians
• Orbit Retrieval
MIXTURES OF SPHERICAL GAUSSIANS

Let’s see another powerful application of tensor methods to learning mixtures of spherical Gaussians

\[
\sum_{i=1}^{k} \omega_i \mathcal{N}(\mu_i, \sigma^2 I, x)
\]
MIXTURES OF SPHERICAL GAUSSIANS

Let’s see another powerful application of tensor methods to learning mixtures of spherical Gaussians

\[ \sum_{i=1}^{k} w_i N(\mu_i, \sigma^2 I, x) \]

Can we reconstruct the parameters in polynomial time?
**MIXTURES OF SPHERICAL GAUSSIANS**

Let’s see another powerful application of tensor methods to learning mixtures of spherical Gaussians

\[
\sum_{i=1}^{k} w_i \mathcal{N}(\mu_i, \sigma^2 I, x)
\]

Can we reconstruct the parameters in polynomial time?

**Theorem [Hsu, Kakade, 2013]:** There is an algorithm that has polynomial run time/sample complexity that works when the \( \mu_i \)'s have full rank

Running time and sample complexity depend on \( 1/\sigma_{\text{min}} \)
**Main Lemma:** If $\sigma^2$ is known then the tensor

$$T = \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i$$

can be expressed through the empirical moments of the mixture
Main Lemma: If $\sigma^2$ is known then the tensor

$$T = \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i$$

can be expressed through the empirical moments of the mixture.

Again, there is a low rank tensor that can be computed from samples whose tensor decomposition reveals the parameters we want to learn.
Main Lemma: If $\sigma^2$ is known then the tensor

$$T = \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i$$

can be expressed through the empirical moments of the mixture.
Main Lemma: If $\sigma^2$ is known then the tensor

$$T = \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i$$

can be expressed through the empirical moments of the mixture

Proof: Consider the a, b, c entry of the third moment tensor
Main Lemma: If $\sigma^2$ is known then the tensor

$$T = \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i$$

can be expressed through the empirical moments of the mixture.

Proof: Consider the a, b, c entry of the third moment tensor.

Case #1: If a, b, c are distinct then we have

$$\mathbb{E}[x_a x_b x_c] = \left( \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i \right)_{a,b,c}$$
Main Lemma: If $\sigma^2$ is known then the tensor

$$T = \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i$$

can be expressed through the empirical moments of the mixture

Proof: Consider the a, b, c entry of the third moment tensor
Main Lemma: If $\sigma^2$ is known then the tensor

$$T = \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i$$

can be expressed through the empirical moments of the mixture.

Proof: Consider the $a$, $b$, $c$ entry of the third moment tensor.

Case #2: If $a = b \neq c$ then we have

$$\mathbb{E}[x_a x_b x_c] = \left( \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i \right)_{a,b,c} + \sigma^2 \left( \sum_{i=1}^{k} w_i \mu_i \right)_c$$
**Main Lemma:** If $\sigma^2$ is known then the tensor

$$T = \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i$$

can be expressed through the empirical moments of the mixture

**Proof:** Consider the $a$, $b$, $c$ entry of the third moment tensor

**Case #2:** If $a = b \neq c$ then we have

$$\mathbb{E}[x_a x_b x_c] = \left( \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i \right)_{a,b,c} + \sigma^2 \left( \sum_{i=1}^{k} w_i \mu_i \right)_{c}$$

*first moment*
Main Lemma: If $\sigma^2$ is known then the tensor

$$T = \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i$$

can be expressed through the empirical moments of the mixture.

Proof: Consider the a, b, c entry of the third moment tensor.
**Main Lemma:** If $\sigma^2$ is known then the tensor

$$T = \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i$$

can be expressed through the empirical moments of the mixture.

**Proof:** Consider the $a$, $b$, $c$ entry of the third moment tensor.

**Case #3:** If $a = b = c$ then we have

$$\mathbb{E}[x_a x_b x_c] = \left( \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i \right)_{a,b,c} - 3\sigma^2 \left( \sum_{i=1}^{k} w_i \mu_i \right)_c$$
Main Lemma: If $\sigma^2$ is known then the tensor

$$T = \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i$$

can be expressed through the empirical moments of the mixture

Proof: Consider the $a, b, c$ entry of the third moment tensor

Case #3: If $a = b = c$ then we have

$$\mathbb{E}[x_ax_bx_c] = \left( \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i \right)_{a,b,c} - 3\sigma^2 \left( \sum_{i=1}^{k} w_i \mu_i \right)_c$$
Main Lemma: If $\sigma^2$ is known then the tensor

$$T = \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i$$

can be expressed through the empirical moments of the mixture.
Main Lemma: If $\sigma^2$ is known then the tensor

$$T = \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i$$

can be expressed through the empirical moments of the mixture.

It can be written compactly as

$$T = \mathbb{E}[x \otimes x \otimes x] - \sigma^2 \sum_{j=1}^{d} M_j \quad \text{with}$$

$$M_j = \left( \mathbb{E}[x] \otimes e_j \otimes e_j + e_j \otimes \mathbb{E}[x] \otimes e_j + e_j \otimes e_j \otimes \mathbb{E}[x] \right)$$
Main Lemma: If $\sigma^2$ is known then the tensor

$$T = \sum_{i=1}^{k} w_i \mu_i \otimes \mu_i \otimes \mu_i$$

can be expressed through the empirical moments of the mixture.

It can be written compactly as

$$T = \mathbb{E}[x \otimes x \otimes x] - \sigma^2 \sum_{j=1}^{d} M_j \quad \text{with}$$

$$M_j = \left( \mathbb{E}[x] \otimes e_j \otimes e_j + e_j \otimes \mathbb{E}[x] \otimes e_j + e_j \otimes e_j \otimes \mathbb{E}[x] \right)$$

Now use Jennrich’s Algorithm
THE POWER OF CONDITIONAL INDEPENDENCE

[Phylogenetic Trees/HMMS]: (joint distribution on leaves a, b, c)

$$\sum_{\sigma} P[z = \sigma] P[a|z = \sigma] \otimes P[b|z = \sigma] \otimes P[c|z = \sigma]$$

following [Mossel, Roch, 2006]
THE POWER OF CONDITIONAL INDEPENDENCE

[Phylogenetic Trees/HMMS]: (joint distribution on leaves a, b, c)

\[ \sum_{\sigma} \mathbb{P}[z = \sigma] \mathbb{P}[a | z = \sigma] \otimes \mathbb{P}[b | z = \sigma] \otimes \mathbb{P}[c | z = \sigma] \]

following [Mossel, Roch, 2006]

[Mixtures of Spherical Gaussians]: (corrections of third moment)

\[ \mathbb{E}[x \otimes x \otimes x] - \sigma^2 \sum_{j=1}^{d} M_j \]

following [Hsu, Kakade, 2013]
THE POWER OF CONDITIONAL INDEPENDENCE

[Pure Topic Models/LDA]: (joint distribution on first three words)

\[
\sum_j \mathbb{P}[\text{topic} = j] A_j \otimes A_j \otimes A_j
\]

following [Anandkumar, Hsu, Kakade, 2012]
THE POWER OF CONDITIONAL INDEPENDENCE

[Pure Topic Models/LDA]: (joint distribution on first three words)

$$\sum_j \mathbb{P}[\text{topic} = j] A_j \otimes A_j \otimes A_j$$

following [Anandkumar, Hsu, Kakade, 2012]

[Community Detection]: (counting stars)

$$\sum_j \mathbb{P}[C_x = j] \left( C_A \Pi \right)_j \otimes \left( C_B \Pi \right)_j \otimes \left( C_C \Pi \right)_j$$

following [Anandkumar, Ge, Hsu, Kakade, 2014]
OUTLINE

Part I: Introduction
- The Rotation Problem
- Jennrich’s Algorithm

Part II: Applications
- Phylogenetic Reconstruction
- Mixtures of Gaussians
- Orbit Retrieval
OUTLINE

Part I: Introduction
  • The Rotation Problem
  • Jennrich’s Algorithm

Part II: Applications
  • Phylogenetic Reconstruction
  • Mixtures of Gaussians
  • Orbit Retrieval
What if we want to learn the parameters of generative model with a continuous latent variable?
ORBIT RETRIEVAL

What if we want to learn the parameters of generative model with a continuous latent variable?

Multireference Alignment

Recover a signal from random noisy shifts
What if we want to learn the parameters of generative model with a continuous latent variable?
ORBIT RETRIEVAL

What if we want to learn the parameters of generative model with a continuous latent variable?

Global Registration

Estimate positions from rigid motions
What if we want to learn the parameters of generative model with a continuous latent variable?
What if we want to learn the parameters of generative model with a continuous latent variable?

**Cryo-electron microscopy**

Determine 3D structure from random noisy 2D projections
Definition: An orbit retrieval problem is specified by a group $G$ and a linear homomorphism

$$\rho : G \to GL(\mathbb{R}^d)$$

We get noisy observations under the group action

$$\rho(g) \cdot x + \eta$$

where $g$ is chosen from the Haar measure on $G$ and $\eta$ is Gaussian noise.
**ORBIT RETRIEVAL**

**Definition:** An orbit retrieval problem is specified by a group $G$ and a linear homomorphism

$$\rho : G \rightarrow GL(\mathbb{R}^d)$$

We get noisy observations under the group action

$$\rho(g) \cdot x + \eta$$

where $g$ is chosen from the Haar measure on $G$ and $\eta$ is Gaussian noise

**Goal:** Recover some $\hat{x}$ that is close to the orbit

$$\{\rho(g) \cdot x \mid g \in G\}$$
In many settings we can estimate

$$T = \int_{g \in G} (\rho(g) \cdot x)^{\otimes 3} dg$$
In many settings we can estimate

\[ T = \int_{g \in G} \left( \rho(g) \cdot x \right) \otimes^3 dg \]

Can we recover \( x \) up to its orbit?
ORBIT TENSOR DECOMPOSITION

In many settings we can estimate

\[ T = \int_{g \in G} (\rho(g) \cdot x) \otimes^3 dg \]

Can we recover \( x \) up to its orbit?

**Theorem [Moitra, Wein, 2019]:** There is a polynomial time algorithm that works for \( \text{SO}(2) \) when \( x \) is random.
In many settings we can estimate

\[ T = \int_{g \in G} (\rho(g) \cdot x) \otimes^3 dg \]

Can we recover \( x \) up to its orbit?

**Theorem [Moitra, Wein, 2019]:** There is a polynomial time algorithm that works for \( \text{SO}(2) \) when \( x \) is random.

What about for non-abelian groups?
Tensor networks are a graphical representation for tensors and operations on them, e.g.
Tensor networks are a graphical representation for tensors and operations on them, e.g.

**third order tensors have three legs**

\[
T = (T_{i,j,k})
\]
Tensor networks are a graphical representation for tensors and operations on them, e.g.

**third order tensors have three legs**

![Diagram of third order tensor](image1)

\[ T = (T_{i,j,k}) \]

**tensors can be attached by summing over connected indices**

![Diagram of tensor attachment](image2)

\[ B_{a,b,c,d} = \sum_{i} T_{a,c,i} U_{b,d,i} \]
REVISITING PRIOR WORK

Prior work implicitly uses this framework

See [Richard, Montanari], [Barak, Moitra], [Hopkins, Shi, Steurer], [Hopkins et al.], [Hopkins, Shi, Steurer] for applications to tensor principal component analysis, tensor completion, decomposing random overcomplete third order tensors, etc.
Given input tensor T

- **Step #1:** Build a new tensor $B$ by connecting copies of $T$ according to the tensor network
- **Step #2:** Flatten $B$ to form a symmetric matrix $M$
- **Step #3:** Compute the leading eigenvector of $M$
THE BLUEPRINT

We give a spectral method based on the following tensor network

\[
\begin{array}{c}
\text{T} \\
\text{T} \\
\text{T} \\
\text{T} \\
\text{T} \\
\text{T} \\
\text{T} \\
\text{T} \\
\text{T} \\
\text{T} \\
\text{T} \\
\end{array}
\]
THE BLUEPRINT

We give a spectral method based on the following tensor network

Smaller tensor networks fail for this problem
TUTORIAL OUTLINE

Part I: Tensor Decompositions and Their Applications

Part II: Robust and Computationally Efficient Parameter Estimation

Part III: Noise Models in Supervised Learning and Connections to Fairness

Part IV: Provable Algorithms for Inverse Problems in the Sciences?
Summary:

• Tensor decompositions are unique under more general conditions than matrix decompositions

• Jennrich’s Algorithm

• Applications to Phylogenetic Reconstruction, HMMs, Mixtures of Gaussians, Topic Models, ...

• Are there tensor methods that work with group structure?
Summary:

• Tensor decompositions are unique under more general conditions than matrix decompositions
• Jennrich’s Algorithm
• Applications to Phylogenetic Reconstruction, HMMs, Mixtures of Gaussians, Topic Models, ...
• Are there tensor methods that work with group structure?

Thanks! Any Questions?