# Robustly Recovering a Signal, Under a Group Action

Ankur Moitra (MIT)

Based on joint work with Alex Wein (NYU)

# APPLICATIONS IN ENGINEERING

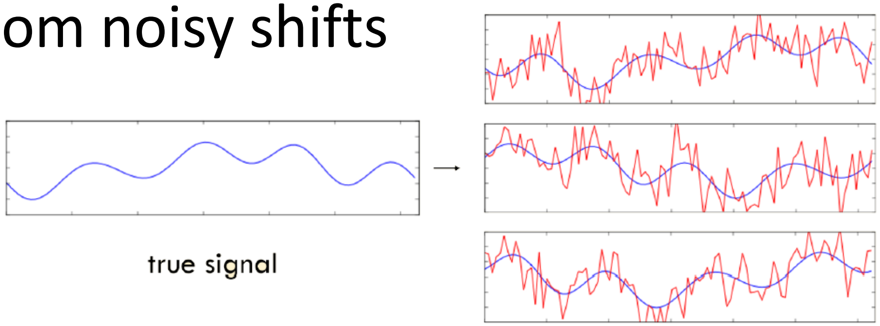
Many inverse problems where groups play a key role

# APPLICATIONS IN ENGINEERING

Many inverse problems where groups play a key role

**Multireference Alignment**

Recover a signal from random noisy shifts



true signal

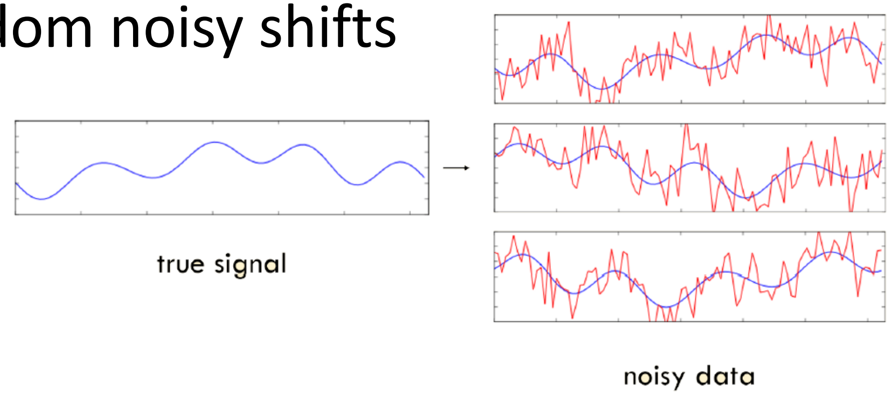
noisy data

# APPLICATIONS IN ENGINEERING

Many inverse problems where groups play a key role

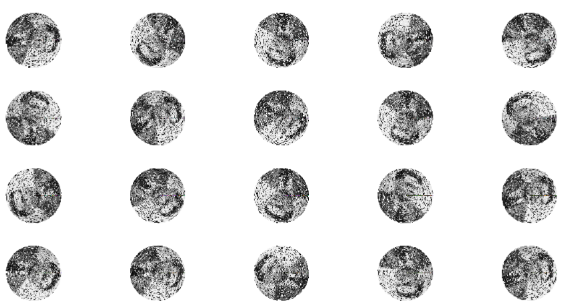
## Multireference Alignment

Recover a signal from random noisy shifts




## Global Registration

Estimate positions from rigid motions



# APPLICATIONS IN ENGINEERING

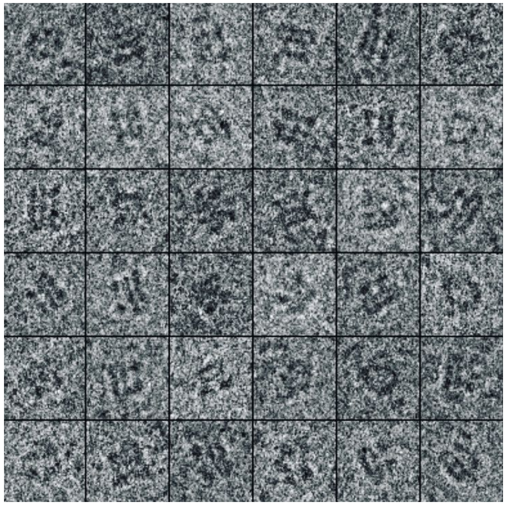
Many inverse problems where groups play a key role

# APPLICATIONS IN ENGINEERING

Many inverse problems where groups play a key role

**Cryo-electron microscopy**

Determine 3D structure from random noisy 2D projections



Revolutionary technique that was awarded the **2017 Nobel Prize in Chemisitry**

# General framework for capturing inverse problems under a group action

General framework for capturing inverse problems under a group action

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

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

General framework for capturing inverse problems under a group action

**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

General framework for capturing inverse problems under a group action

**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

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

$$\{\rho(g) \cdot x | g \in G\}$$

# BACK TO THE APPLICATIONS

How do popular inverse problems fit into this framework?

| Problem | Group | Notes |
|---|---|---|
| | | |
| | | |
| | | |
| | | |

# BACK TO THE APPLICATIONS

How do popular inverse problems fit into this framework?

| Problem | Group | Notes |
|---|---|---|
| **Discrete MRA** | $\mathbb{Z}/d\mathbb{Z}$ | |
| | | |
| | | |
| | | |

# BACK TO THE APPLICATIONS

How do popular inverse problems fit into this framework?

| Problem | Group | Notes |
|---|---|---|
| **Discrete MRA** | $\mathbb{Z}/d\mathbb{Z}$ | |
| **Continuous MRA** | $SO(2)$ | Assume x is band-limited |
| | | |
| | | |

# BACK TO THE APPLICATIONS

How do popular inverse problems fit into this framework?

| Problem | Group | Notes |
|---|---|---|
| **Discrete MRA** | $\mathbb{Z}/d\mathbb{Z}$ | |
| **Continuous MRA** | $SO(2)$ | Assume x is band-limited |
| **Image Registration** | $SO(2)$ | |
| | | |

# BACK TO THE APPLICATIONS

How do popular inverse problems fit into this framework?

| Problem | Group | Notes |
|---|---|---|
| **Discrete MRA** | $\mathbb{Z}/d\mathbb{Z}$ | |
| **Continuous MRA** | $SO(2)$ | Assume x is band-limited |
| **Image Registration** | $SO(2)$ | |
| **Cryo-EM** | $SO(3)$ | More general, apply a projection after rotation |

# ALGORITHMS?

In practice, often use iterative algorithms like EM or BP

# ALGORITHMS?

In practice, often use iterative algorithms like EM or BP

**Main Idea:** Attempt to align pairs of samples

$$y_i = \rho(g_i) \cdot x + \eta_i \quad \text{and} \quad y_j = \rho(g_j) \cdot x + \eta_j$$

to estimate the relative group action $\rho(g_i)\rho(g_j)^{-1}$

# ALGORITHMS?

In practice, often use iterative algorithms like EM or BP

**Main Idea:** Attempt to align pairs of samples

$$y_i = \rho(g_i) \cdot x + \eta_i \quad \text{and} \quad y_j = \rho(g_j) \cdot x + \eta_j$$

to estimate the relative group action $\rho(g_i)\rho(g_j)^{-1}$

**Serious Limitations:**

# ALGORITHMS?

In practice, often use iterative algorithms like EM or BP

**Main Idea:** Attempt to align pairs of samples

$$y_i = \rho(g_i) \cdot x + \eta_i \quad \text{and} \quad y_j = \rho(g_j) \cdot x + \eta_j$$

to estimate the relative group action $\rho(g_i)\rho(g_j)^{-1}$

**Serious Limitations:**

There are no provable guarantees

# ALGORITHMS?

In practice, often use iterative algorithms like EM or BP

**Main Idea:** Attempt to align pairs of samples

$$y_i = \rho(g_i) \cdot x + \eta_i \quad \text{and} \quad y_j = \rho(g_j) \cdot x + \eta_j$$

to estimate the relative group action $\rho(g_i)\rho(g_j)^{-1}$

**Serious Limitations:**

There are no provable guarantees

How do we know we recovered the actual 3D structure, or just got stuck in some **local minimum**?

# ALGORITHMS?

In practice, often use iterative algorithms like EM or BP

**Main Idea:** Attempt to align pairs of samples

$$y_i = \rho(g_i) \cdot x + \eta_i \quad \text{and} \quad y_j = \rho(g_j) \cdot x + \eta_j$$

to estimate the relative group action $\rho(g_i)\rho(g_j)^{-1}$

**Serious Limitations:**

# ALGORITHMS?

In practice, often use iterative algorithms like EM or BP

**Main Idea:** Attempt to align pairs of samples

$$y_i = \rho(g_i) \cdot x + \eta_i \quad \text{and} \quad y_j = \rho(g_j) \cdot x + \eta_j$$

to estimate the relative group action $\rho(g_i)\rho(g_j)^{-1}$

**Serious Limitations:**

They work poorly when there is a lot of noise

# ALGORITHMS?

In practice, often use iterative algorithms like EM or BP

**Main Idea:** Attempt to align pairs of samples

$$y_i = \rho(g_i) \cdot x + \eta_i \quad \text{and} \quad y_j = \rho(g_j) \cdot x + \eta_j$$

to estimate the relative group action $\rho(g_i)\rho(g_j)^{-1}$

**Serious Limitations:**

They work poorly when there is a lot of noise

When the noise is large, it is often **impossible** to find the relative group action

# ALGORITHMS?

In practice, often use iterative algorithms like EM or BP

**Main Idea:** Attempt to align pairs of samples

$$y_i = \rho(g_i) \cdot x + \eta_i \quad \text{and} \quad y_j = \rho(g_j) \cdot x + \eta_j$$

to estimate the relative group action $\rho(g_i)\rho(g_j)^{-1}$

**Serious Limitations:**

# ALGORITHMS?

In practice, often use iterative algorithms like EM or BP

**Main Idea:** Attempt to align pairs of samples

$$y_i = \rho(g_i) \cdot x + \eta_i \quad \text{and} \quad y_j = \rho(g_j) \cdot x + \eta_j$$

to estimate the relative group action $\rho(g_i)\rho(g_j)^{-1}$

**Serious Limitations:**

They don't extend to other settings

# ALGORITHMS?

In practice, often use iterative algorithms like EM or BP

**Main Idea:** Attempt to align pairs of samples

$$y_i = \rho(g_i) \cdot x + \eta_i \quad \text{and} \quad y_j = \rho(g_j) \cdot x + \eta_j$$

to estimate the relative group action $\rho(g_i)\rho(g_j)^{-1}$

**Serious Limitations:**

They don't extend to other settings

For example, can we allow **heterogeneity**? This is natural when a molecule has multiple isomers

# TENSOR DECOMPOSITIONS

There are some algorithms, with provable guarantees, based on tensor decompositions

$$T = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$$

# TENSOR DECOMPOSITIONS

There are some algorithms, with provable guarantees, based on tensor decompositions

$$T = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$$

**Moments of the distribution**

# TENSOR DECOMPOSITIONS

There are some algorithms, with provable guarantees, based on tensor decompositions

$$T = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$$

**Moments of the distribution**

**Hope: The factors in the low-rank decomposition be used to find x?**

# TENSOR DECOMPOSITIONS

There are some algorithms, with provable guarantees, based on tensor decompositions

$$T = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$$

**Moments of the distribution**

**Hope: The factors in the low-rank decomposition be used to find x?**

**But, as we will see they only work for finite groups and ignore the group structure**

# OUTLINE

## Part I: Introduction

- Orbit Recovery Problems
- A More General Tensor Spectral Toolkit?

## Part II: Discrete vs. Continuous MRA

- Jennrich's Algorithm
- The Challenge of Infinite Groups
- Our Results

## Part III: Tensor Networks

- Spectral Methods from Tensor Networks
- Our Approach, and Discussion

# OUTLINE

## Part I: Introduction

- Orbit Recovery Problems
- **A More General Tensor Spectral Toolkit?**

## Part II: Discrete vs. Continuous MRA

- Jennrich's Algorithm
- The Challenge of Infinite Groups
- Our Results

## Part III: Tensor Networks

- Spectral Methods from Tensor Networks
- Our Approach, and Discussion

# RESEARCH GOALS

Can we develop a more general tensor spectral toolkit for solving orbit recovery that

# RESEARCH GOALS

Can we develop a more general tensor spectral toolkit for solving orbit recovery that

(1) works with **low signal-to-noise ratio**,

# RESEARCH GOALS

Can we develop a more general tensor spectral toolkit for solving orbit recovery that

(1) works with **low signal-to-noise ratio**,

(2) **actually uses the group structure**, e.g. for infinite and non-abelian groups and

# RESEARCH GOALS

Can we develop a more general tensor spectral toolkit for solving orbit recovery that

(1) works with **low signal-to-noise ratio**,

(2) **actually uses the group structure**, e.g. for infinite and non-abelian groups and

(3) allows for **heterogeneity**?

# RESEARCH GOALS

Can we develop a more general tensor spectral toolkit for solving orbit recovery that

(1) works with **low signal-to-noise ratio**,

(2) **actually uses the group structure**, e.g. for infinite and non-abelian groups and

(3) allows for **heterogeneity**?

As usual, we want algorithms with **provable guarantees**!

# RESEARCH GOALS

Can we develop a more general tensor spectral toolkit for solving orbit recovery that

(1) works with **low signal-to-noise ratio**,

(2) **actually uses the group structure**, e.g. for infinite and non-abelian groups and

(3) allows for **heterogeneity**?

As usual, we want algorithms with **provable guarantees**!

We will need to solve some challenging problems at the interface between combinatorics and representation theory along the way

# OUTLINE

**Part I: Introduction**

- Orbit Retrieval Problems
- A More General Tensor Spectral Toolkit?

**Part II: Discrete vs. Continuous MRA**

- Jennrich's Algorithm
- The Challenge of Infinite Groups
- Our Results

**Part III: Tensor Networks**

- Spectral Methods from Tensor Networks
- Our Approach, and Discussion

# OUTLINE

## Part I: Introduction

- Orbit Retrieval Problems
- A More General Tensor Spectral Toolkit?

## Part II: Discrete vs. Continuous MRA

- Jennrich's Algorithm
- The Challenge of Infinite Groups
- Our Results

## Part III: Tensor Networks

- Spectral Methods from Tensor Networks
- Our Approach, and Discussion

# WARM-UP: DISCRETE MRA

Recall, we get $y_i = \rho(g_i) \cdot x + \eta_i$ , where $\eta_i \sim \mathcal{N}(0, \sigma^2 I)$

**random cyclic shift**

# WARM-UP: DISCRETE MRA

Recall, we get $y_i = \rho(g_i) \cdot x + \eta_i$ , where $\eta_i \sim \mathcal{N}(0, \sigma^2 I)$

**random cyclic shift**

**Natural Idea:** Work with a polynomial that is **invariant under the group action**

# WARM-UP: DISCRETE MRA

Recall, we get $y_i = \rho(g_i) \cdot x + \eta_i$ , where $\eta_i \sim \mathcal{N}(0, \sigma^2 I)$

**random cyclic shift**

**Natural Idea:** Work with a polynomial that is **invariant under the group action**, e.g.

$$T = \sum_{g \in G} (\rho(g) \cdot x)^{\otimes 3}$$

# WARM-UP: DISCRETE MRA

Recall, we get $y_i = \rho(g_i) \cdot x + \eta_i$, where $\eta_i \sim \mathcal{N}(0, \sigma^2 I)$

**random cyclic shift**

**Natural Idea:** Work with a polynomial that is **invariant under the group action**, e.g.

$$T = \sum_{g \in G} (\rho(g) \cdot x)^{\otimes 3}$$

**Lemma [Perry, Weed, Bandeira, Rigollet, Singer]:** T can be estimated from samples through the relation

$$T = \mathbb{E}[(\rho(g) \cdot y)^{\otimes 3} - 3\sigma^2 \, \mathrm{sym}((\rho(g) \cdot y) \otimes I)]$$

where the expectation is over y and a random g

# OUTLINE

## Part I: Introduction

- Orbit Retrieval Problems
- A More General Tensor Spectral Toolkit?

## Part II: Discrete vs. Continuous MRA

- Jennrich's Algorithm
- The Challenge of Infinite Groups
- Our Results

## Part III: Tensor Networks

- Spectral Methods from Tensor Networks
- Our Approach, and Discussion

# OUTLINE

**Part I: Introduction**

- Orbit Retrieval Problems
- A More General Tensor Spectral Toolkit?

**Part II: Discrete vs. Continuous MRA**

- **Jennrich's Algorithm**
- The Challenge of Infinite Groups
- Our Results

**Part III: Tensor Networks**

- Spectral Methods from Tensor Networks
- Our Approach, and Discussion

# JENNRICH'S ALGORITHM

**Theorem [Jennrich]:** There is a polynomial time algorithm to decompose a tensor of the form

$$T = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$$

provided that the factors $\{u_i\}_i$, $\{v_i\}_i$ and $\{w_i\}_i$ are all linearly independent

# JENNRICH'S ALGORITHM

**Theorem [Jennrich]:** There is a polynomial time algorithm to decompose a tensor of the form

$$T = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$$

provided that the factors $\{u_i\}_i, \{v_i\}_i$ and $\{w_i\}_i$ are all linearly independent

We will be able to use this algorithm to solve discrete MRA

Recall, we estimate the following tensor from samples

$$T = \sum_{g \in G} (\rho(g) \cdot x)^{\otimes 3}$$

Recall, we estimate the following tensor from samples

$$T = \sum_{g \in G} (\rho(g) \cdot x)^{\otimes 3}$$

To apply Jennrich's algorithm, we need to check that the factors

$$\{\rho(g) \cdot x\}_g$$

are linearly independent

Recall, we estimate the following tensor from samples

$$T = \sum_{g \in G} (\rho(g) \cdot x)^{\otimes 3}$$

To apply Jennrich's algorithm, we need to check that the factors

$$\{\rho(g) \cdot x\}_g$$

are linearly independent, which they are for generic x

Recall, we estimate the following tensor from samples

$$T = \sum_{g \in G} (\rho(g) \cdot x)^{\otimes 3}$$

To apply Jennrich's algorithm, we need to check that the factors

$$\{\rho(g) \cdot x\}_g$$

are linearly independent, which they are for generic x

**Theorem [Perry, Weed, Bandeira, Rigollet, Singer]:** There is an algorithm with polynomial running time and sample complexity to solve discrete MRA for smoothed x

Recall, we estimate the following tensor from samples

$$T = \sum_{g \in G} (\rho(g) \cdot x)^{\otimes 3}$$

To apply Jennrich's algorithm, we need to check that the factors

$$\{\rho(g) \cdot x\}_g$$

are linearly independent, which they are for generic x

**Theorem [Perry, Weed, Bandeira, Rigollet, Singer]:** There is an algorithm with polynomial running time and sample complexity to solve discrete MRA for smoothed x

The running time and sample complexity depend polynomially on inverse accuracy, and on the condition number of the factors

# OUTLINE

## Part I: Introduction

- Orbit Retrieval Problems
- A More General Tensor Spectral Toolkit?

## Part II: Discrete vs. Continuous MRA

- Jennrich's Algorithm
- The Challenge of Infinite Groups
- Our Results

## Part III: Tensor Networks

- Spectral Methods from Tensor Networks
- Our Approach, and Discussion

# OUTLINE

## Part I: Introduction

- Orbit Retrieval Problems
- A More General Tensor Spectral Toolkit?

## Part II: Discrete vs. Continuous MRA

- Jennrich's Algorithm
- **The Challenge of Infinite Groups**
- Our Results

## Part III: Tensor Networks

- Spectral Methods from Tensor Networks
- Our Approach, and Discussion

# GROUP STRUCTURE?

**Upon closer inspection, the algorithm does not actually depend on the group structure!**

# GROUP STRUCTURE?

**Upon closer inspection, the algorithm does not actually depend on the group structure!**

In fact, we can think of samples from discrete MRA

$$y_i = \rho(g_i) \cdot x + \eta_i \text{ , where } \eta_i \sim \mathcal{N}(0, \sigma^2 I)$$

as samples from a structured **mixture of gaussians**

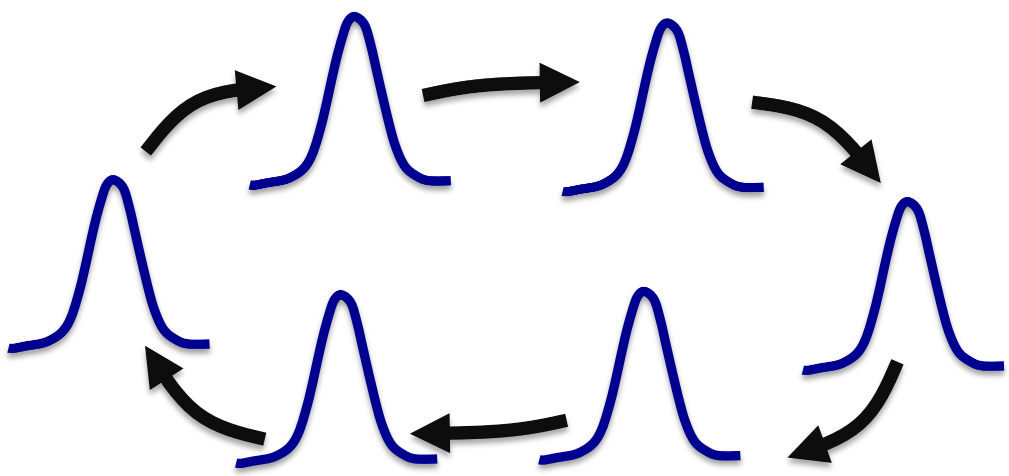
# GROUP STRUCTURE?

**Upon closer inspection, the algorithm does not actually depend on the group structure!**

In fact, we can think of samples from discrete MRA

$$y_i = \rho(g_i) \cdot x + \eta_i \text{ , where } \eta_i \sim \mathcal{N}(0, \sigma^2 I)$$

as samples from a structured **mixture of gaussians where the centers are all cyclic shifts of x**



# GROUP STRUCTURE?

**Upon closer inspection, the algorithm does not actually depend on the group structure!**

In fact, we can think of samples from discrete MRA

$$y_i = \rho(g_i) \cdot x + \eta_i \text{ , where } \eta_i \sim \mathcal{N}(0, \sigma^2 I)$$

as samples from a structured **mixture of gaussians where the centers are all cyclic shifts of x**

# GROUP STRUCTURE?

**Upon closer inspection, the algorithm does not actually depend on the group structure!**

In fact, we can think of samples from discrete MRA

$$y_i = \rho(g_i) \cdot x + \eta_i \text{ , where } \eta_i \sim \mathcal{N}(0, \sigma^2 I)$$

as samples from a structured **mixture of gaussians where the centers are all cyclic shifts of x**

$$\sum_{g \in G} \frac{1}{d} \mathcal{N}(\rho(g) \cdot x, \sigma^2 I)$$

# GROUP STRUCTURE?

**Upon closer inspection, the algorithm does not actually depend on the group structure!**

In fact, we can think of samples from discrete MRA

$$y_i = \rho(g_i) \cdot x + \eta_i \text{ , where } \eta_i \sim \mathcal{N}(0, \sigma^2 I)$$

as samples from a structured **mixture of gaussians where the centers are all cyclic shifts of x**

$$\sum_{g \in G} \frac{1}{d} \mathcal{N}(\rho(g) \cdot x, \sigma^2 I)$$

and we're using tensor decomposition to learn the model, following **[Hsu, Kakade]**, but ignoring the relationship between the centers

# BEYOND FINITE GROUPS

What if we have a continuous group?

# BEYOND FINITE GROUPS

What if we have a continuous group?

E.g. for continuous MRA, it is natural to work with the tensor

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

# BEYOND FINITE GROUPS

What if we have a continuous group?

E.g. for continuous MRA, it is natural to work with the tensor

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

**Now the tensor is no longer low-rank!**

# BEYOND FINITE GROUPS

What if we have a continuous group?

E.g. for continuous MRA, it is natural to work with the tensor

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

**Now the tensor is no longer low-rank!**

**We call this problem orbit tensor decomposition, and it seems to be the key to solving more general orbit recovery problems**

# OUTLINE

## Part I: Introduction

- Orbit Retrieval Problems
- A More General Tensor Spectral Toolkit?

## Part II: Discrete vs. Continuous MRA

- Jennrich's Algorithm
- The Challenge of Infinite Groups
- Our Results

## Part III: Tensor Networks

- Spectral Methods from Tensor Networks
- Our Approach, and Discussion

# OUTLINE

## Part I: Introduction

- Orbit Retrieval Problems
- A More General Tensor Spectral Toolkit?

## Part II: Discrete vs. Continuous MRA

- Jennrich's Algorithm
- The Challenge of Infinite Groups
- **Our Results**

## Part III: Tensor Networks

- Spectral Methods from Tensor Networks
- Our Approach, and Discussion

# MAIN RESULTS

We give the first polynomial time algorithm to solve orbit tensor decomposition over an infinite group

# MAIN RESULTS

We give the first polynomial time algorithm to solve orbit tensor decomposition over an infinite group

As a result we get:

**Theorem [Moitra, Wein]:** There is a polynomial time algorithm for list recovery for continuous MRA that works with additive noise and in a heterogenous setting where there are a polynomial number of random components

# MAIN RESULTS

We give the first polynomial time algorithm to solve orbit tensor decomposition over an infinite group

As a result we get:

**Theorem [Moitra, Wein]:** There is a polynomial time algorithm for list recovery for continuous MRA that works with additive noise and in a heterogenous setting where there are a polynomial number of random components

The algorithm is inspired by recent algorithms for rounding the **sum-of-squares hierarchy**, and an abstraction called **tensor networks**

# OUTLINE

## Part I: Introduction

- Orbit Retrieval Problems
- A More General Tensor Spectral Toolkit?

## Part II: Discrete vs. Continuous MRA

- Jennrich's Algorithm
- The Challenge of Infinite Groups
- Our Results

## Part III: Tensor Networks

- Spectral Methods from Tensor Networks
- Our Approach, and Discussion

# OUTLINE

## Part I: Introduction

- Orbit Retrieval Problems
- A More General Tensor Spectral Toolkit?

## Part II: Discrete vs. Continuous MRA

- Jennrich's Algorithm
- The Challenge of Infinite Groups
- Our Results

## Part III: Tensor Networks

- Spectral Methods from Tensor Networks
- Our Approach, and Discussion

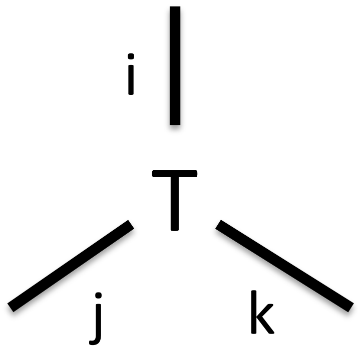
# TENSOR NETWORKS

Tensor networks are a graphical representation for tensors and operations on them, e.g.

# TENSOR NETWORKS

Tensor networks are a graphical representation for tensors and operations on them, e.g.

**third order tensors have three legs**

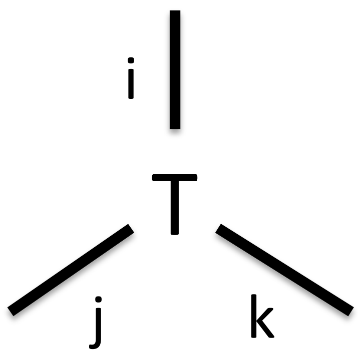



$$T = (T_{i,j,k})$$

# TENSOR NETWORKS

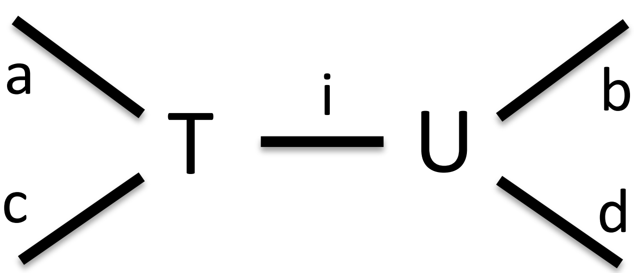
Tensor networks are a graphical representation for tensors and operations on them, e.g.

**third order tensors have three legs**




$$T = (T_{i,j,k})$$

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




$$B_{a,b,c,d} = \sum_i T_{a,c,i} U_{b,d,i}$$

# OUTLINE

## Part I: Introduction

- Orbit Retrieval Problems
- A More General Tensor Spectral Toolkit?

## Part II: Discrete vs. Continuous MRA

- Jennrich's Algorithm
- The Challenge of Infinite Groups
- Our Results

## Part III: Tensor Networks

- Spectral Methods from Tensor Networks
- Our Approach, and Discussion

# OUTLINE

## Part I: Introduction

- Orbit Retrieval Problems
- A More General Tensor Spectral Toolkit?

## Part II: Discrete vs. Continuous MRA

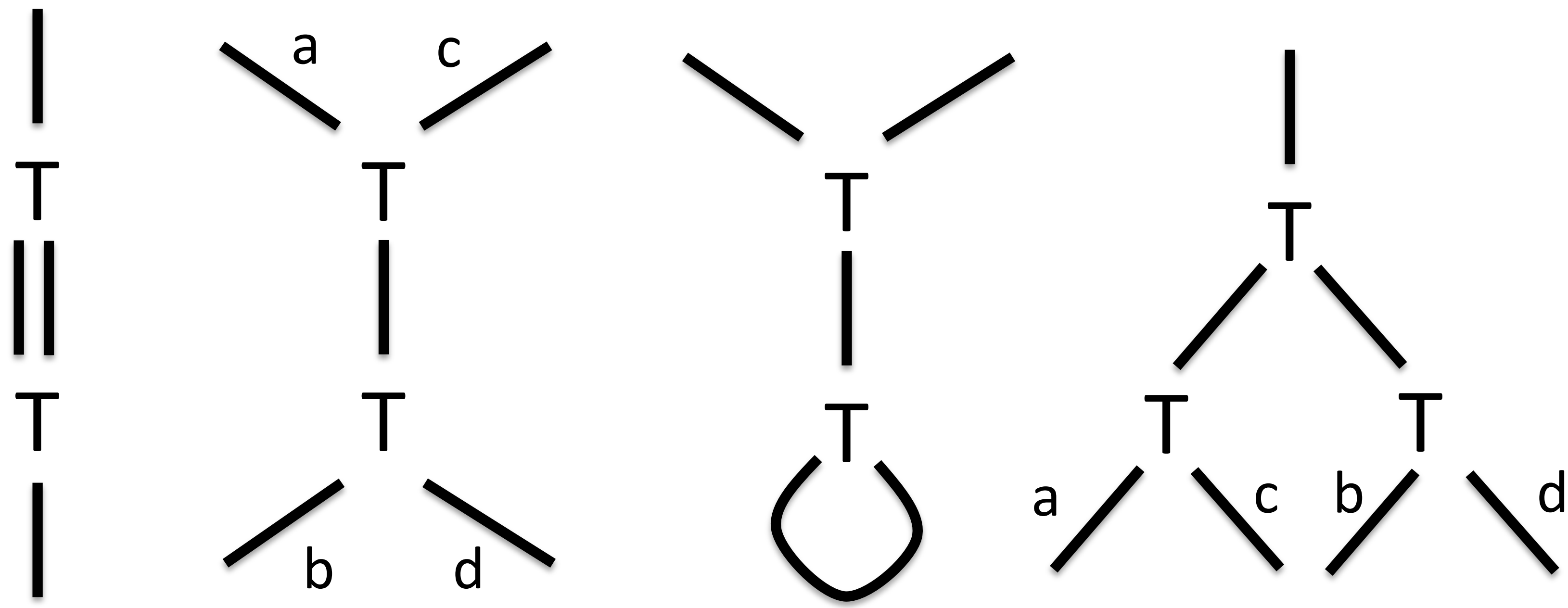
- Jennrich's Algorithm
- The Challenge of Infinite Groups
- Our Results

## Part III: Tensor Networks

- **Spectral Methods from Tensor Networks**
- Our Approach, and Discussion

# 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

# SPECTRAL METHODS FROM TENSOR NETS

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

**We use the trace method to show the top eigenvector is close to the orbit of x**

# SPECTRAL METHODS FROM TENSOR NETS

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

**We use the trace method to show the top eigenvector is close to the orbit of x**

**For tensor networks, this turns into a labelled counting problem**

# OUTLINE

## Part I: Introduction

- Orbit Retrieval Problems
- A More General Tensor Spectral Toolkit?

## Part II: Discrete vs. Continuous MRA

- Jennrich's Algorithm
- The Challenge of Infinite Groups
- Our Results

## Part III: Tensor Networks

- Spectral Methods from Tensor Networks
- Our Approach, and Discussion

# OUTLINE

## Part I: Introduction

- Orbit Retrieval Problems
- A More General Tensor Spectral Toolkit?

## Part II: Discrete vs. Continuous MRA

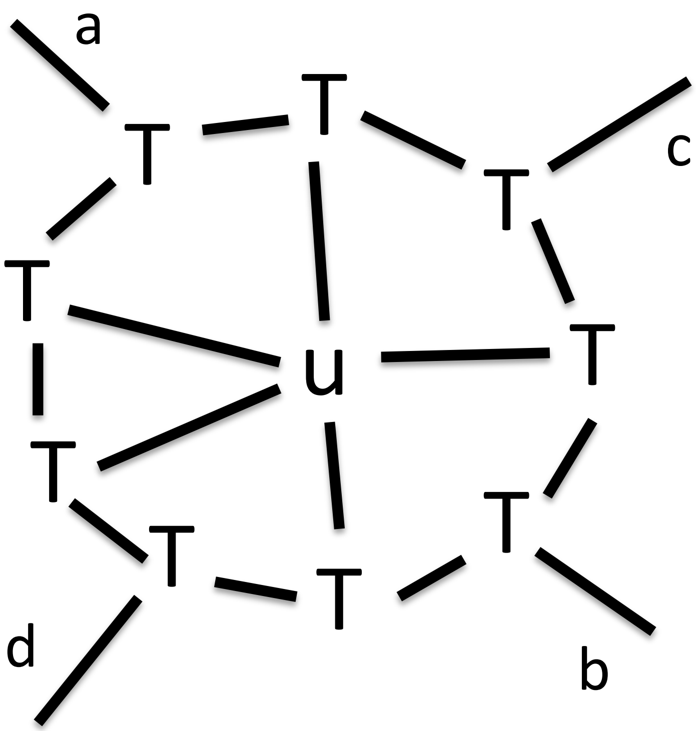
- Jennrich's Algorithm
- The Challenge of Infinite Groups
- Our Results

## Part III: Tensor Networks

- Spectral Methods from Tensor Networks
- **Our Approach, and Discussion**

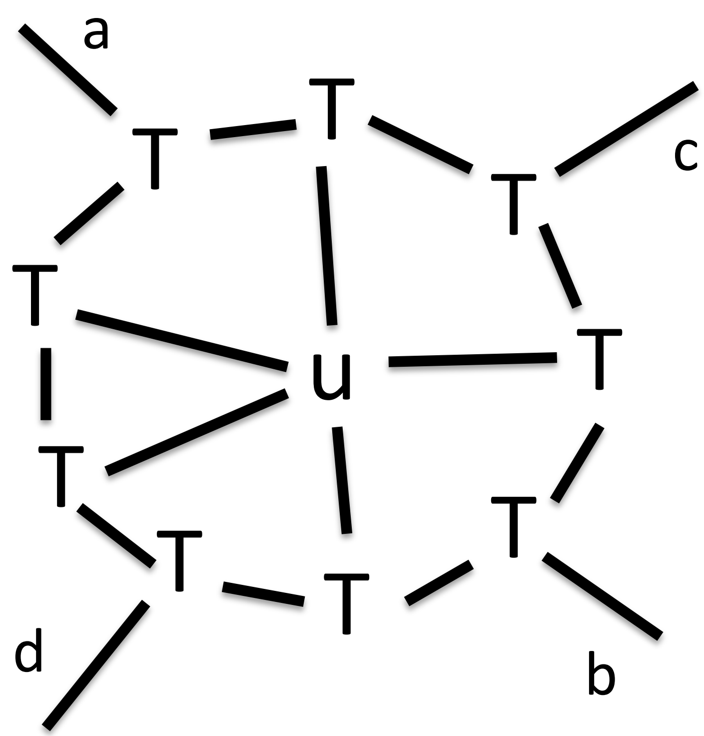
# THE BLUEPRINT

We give a spectral method based on the following tensor network



# THE BLUEPRINT

We give a spectral method based on the following tensor network




**Smaller tensor networks fail for this problem**

# THE TRACE METHOD

Main step in the analysis is to bound the largest eigenvalue of some matrix build from a tensor network (**after projecting out signal**)

# THE TRACE METHOD

Main step in the analysis is to bound the largest eigenvalue of some matrix build from a tensor network (**after projecting out signal**)

We do this through the **trace method**:

$$\mathrm{Tr}(M^{2k}) = \sum_i \lambda_i^{2k} \geq \lambda_{\max}^{2k}$$

# THE TRACE METHOD

Main step in the analysis is to bound the largest eigenvalue of some matrix build from a tensor network (**after projecting out signal**)

We do this through the **trace method**:

$$\mathrm{Tr}(M^{2k}) = \sum_i \lambda_i^{2k} \geq \lambda_{\max}^{2k}$$

Applying Markov's inequality we get the bound

$$\mathbb{P}[\lambda_{\max} \geq t] = \mathbb{P}[\lambda_{\max}^{2k} \geq t^{2k}] \leq \frac{\mathbb{E}[\mathrm{Tr}(M^{2k})]}{t^{2k}}$$

# THE TRACE METHOD

Main step in the analysis is to bound the largest eigenvalue of some matrix build from a tensor network (**after projecting out signal**)

We do this through the **trace method**:

$$\mathrm{Tr}(M^{2k}) = \sum_i \lambda_i^{2k} \geq \lambda_{\max}^{2k}$$

Applying Markov's inequality we get the bound

$$\mathbb{P}[\lambda_{\max} \geq t] = \mathbb{P}[\lambda_{\max}^{2k} \geq t^{2k}] \leq \frac{\mathbb{E}[\mathrm{Tr}(M^{2k})]}{t^{2k}}$$

With tensor networks, the trace method turns into a counting problem

# THE TRACE METHOD

Main step in the analysis is to bound the largest eigenvalue of some matrix build from a tensor network (**after projecting out signal**)

We do this through the **trace method**:

$$\mathrm{Tr}(M^{2k}) = \sum_i \lambda_i^{2k} \geq \lambda_{\max}^{2k}$$

Applying Markov's inequality we get the bound

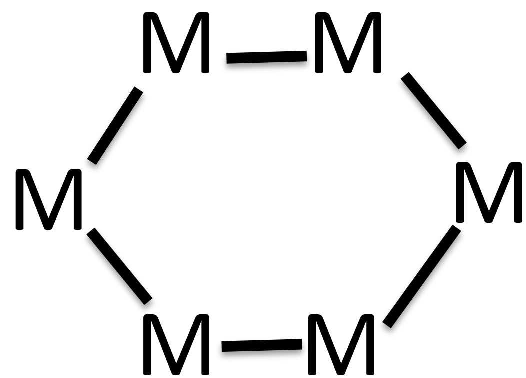
$$\mathbb{P}[\lambda_{\max} \geq t] = \mathbb{P}[\lambda_{\max}^{2k} \geq t^{2k}] \leq \frac{\mathbb{E}[\mathrm{Tr}(M^{2k})]}{t^{2k}}$$

With tensor networks, the trace method turns into a counting problem, **let's see some examples…**

Suppose M is an n x n symmetric matrix with iid Rademacher entries and zeros along the diagonal

Suppose M is an n x n symmetric matrix with iid Rademacher entries and zeros along the diagonal

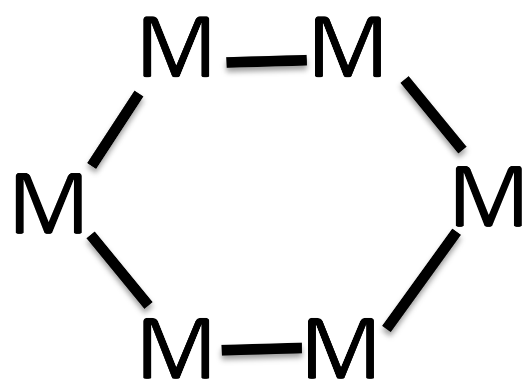
**Lemma:** $\mathbb{E}[\mathrm{Tr}(M^6)]$ is the number of ways of labeling the edges of




with labels from [n] so that **any pair of labels (i,j) is adjacent to an even number of M's**

Suppose M is an n x n symmetric matrix with iid Rademacher entries and zeros along the diagonal

**Lemma:** $\mathbb{E}[\mathrm{Tr}(M^6)]$ is the number of ways of labeling the edges of

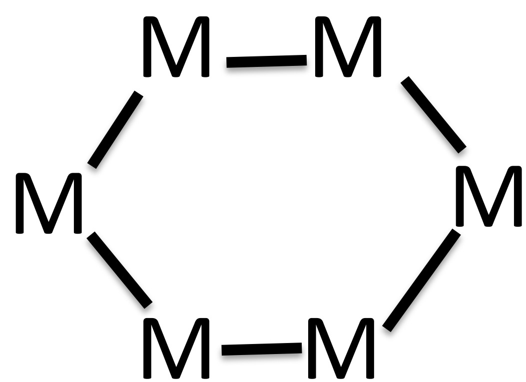



with labels from [n] so that **any pair of labels (i,j) is adjacent to an even number of M's**

**Proof:** First, $\mathrm{Tr}(M^6)$ is a sum over length six walks.

Suppose M is an n x n symmetric matrix with iid Rademacher entries and zeros along the diagonal

**Lemma:** $\mathbb{E}[\mathrm{Tr}(M^6)]$ is the number of ways of labeling the edges of




with labels from [n] so that **any pair of labels (i,j) is adjacent to an even number of M's**

**Proof:** First, $\mathrm{Tr}(M^6)$ is a sum over length six walks. Then observe that a term has expectation zero **unless each edge is traversed an even number of times**. $\blacksquare$

Suppose M is an n x n symmetric matrix with iid Rademacher entries and zeros along the diagonal

Suppose M is an n x n symmetric matrix with iid Rademacher entries and zeros along the diagonal

More generally:

**Lemma:** $\mathbb{E}[\mathrm{Tr}(M^{2k})]$ is the number of ways of labeling the edges of of a length 2k cycle so that any pair (i,j) is adjacent to an even number of M's

Suppose M is an n x n symmetric matrix with iid Rademacher entries and zeros along the diagonal

More generally:

**Lemma:** $\mathbb{E}[\mathrm{Tr}(M^{2k})]$ is the number of ways of labeling the edges of of a length 2k cycle so that any pair (i,j) is adjacent to an even number of M's

**The natural way to double cover edges with a walk is to take the depth first search of a tree**

Suppose M is an n x n symmetric matrix with iid Rademacher entries and zeros along the diagonal

More generally:

**Lemma:** $\mathbb{E}[\mathrm{Tr}(M^{2k})]$ is the number of ways of labeling the edges of of a length 2k cycle so that any pair (i,j) is adjacent to an even number of M's

**The natural way to double cover edges with a walk is to take the depth first search of a tree**

It turns out this is the dominant contribution:

**Theorem [Furedi, Komlos]:** $\mathbb{E}[\mathrm{Tr}(M^{2k})] \lesssim n^k$

Suppose M is an n x n symmetric matrix with iid Rademacher entries and zeros along the diagonal

More generally:

**Lemma:** $\mathbb{E}[\mathrm{Tr}(M^{2k})]$ is the number of ways of labeling the edges of of a length 2k cycle so that any pair (i,j) is adjacent to an even number of M's

**The natural way to double cover edges with a walk is to take the depth first search of a tree**

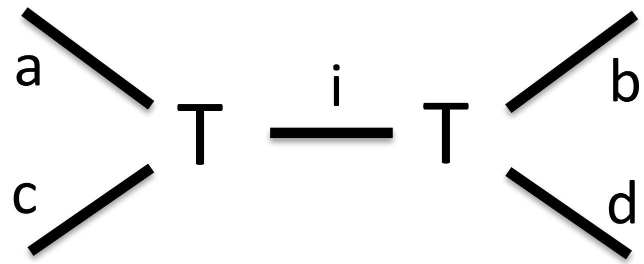
It turns out this is the dominant contribution:

**Theorem [Furedi, Komlos]:** $\mathbb{E}[\mathrm{Tr}(M^{2k})] \lesssim n^k$

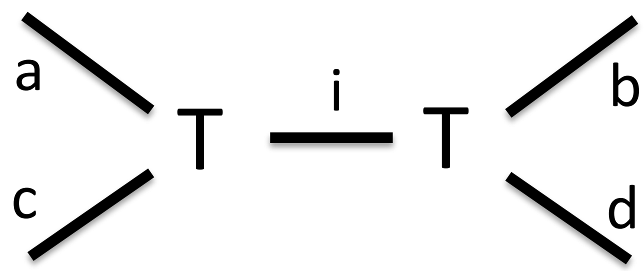
This gives sharp bounds on $||M||$ via the trace method

**More challenging example:** Suppose T is a symmetric tensor with iid Rademacher entries

**More challenging example:** Suppose T is a symmetric tensor with iid Rademacher entries and we plug it into the tensor network

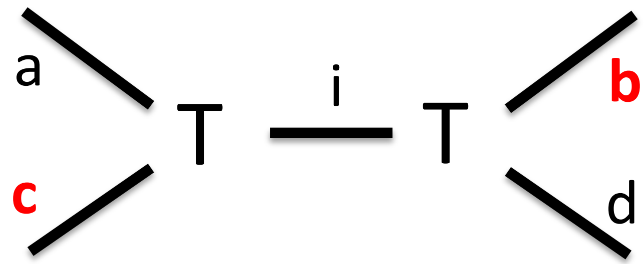


**More challenging example:** Suppose T is a symmetric tensor with iid Rademacher entries and we plug it into the tensor network




Now let M be the ({a, b}, {c, d})-flattenening

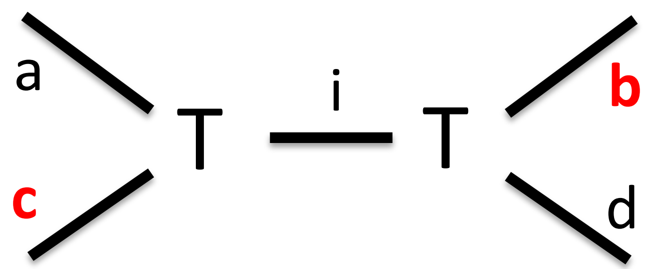
**More challenging example:** Suppose T is a symmetric tensor with iid Rademacher entries and we plug it into the tensor network




Now let M be the ({a, b}, {c, d})-flattenening

Note that the pair of indices {a,b} that index rows of M come from from different copies of T, and this is important

**More challenging example:** Suppose T is a symmetric tensor with iid Rademacher entries and we plug it into the tensor network

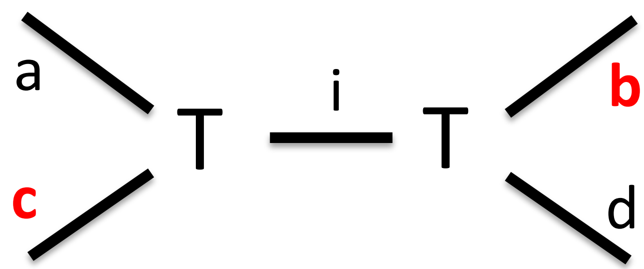



Now let M be the ({a, b}, {c, d})-flattenening

**Natural Goal:** Understand $||M||$ via the trace method

Note that the pair of indices {a,b} that index rows of M come from from different copies of T, and this is important

**More challenging example:** Suppose T is a symmetric tensor with iid Rademacher entries and we plug it into the tensor network




Now let M be the ({a, b}, {c, d})-flattenening

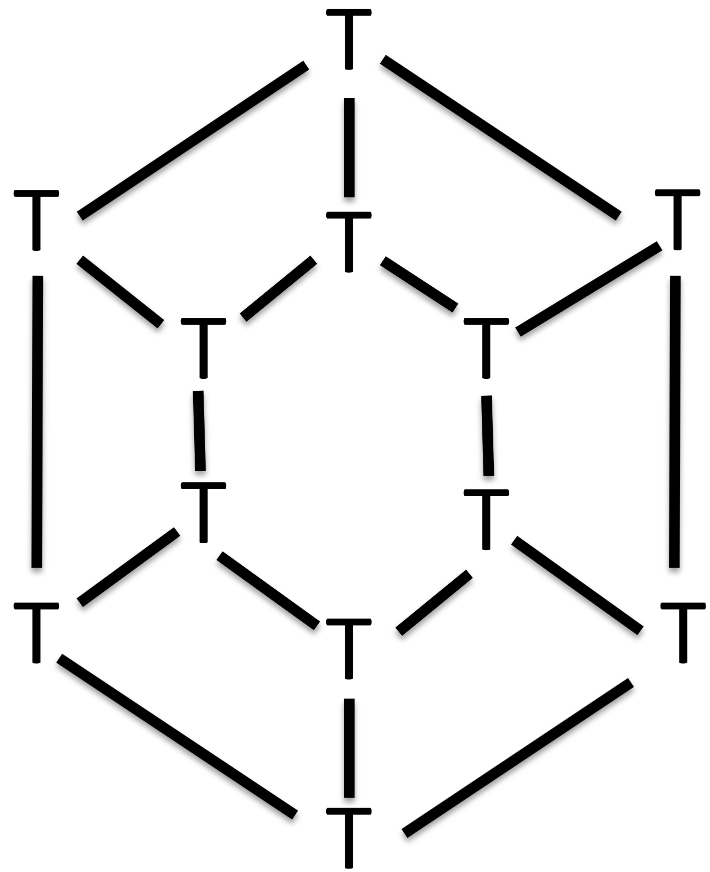
**Natural Goal:** Understand $||M||$ via the trace method

For example, if we want to compute $\mathbb{E}[\mathrm{Tr}(M^6)]$ we can plug the tensor network into the six cycle, and we get…

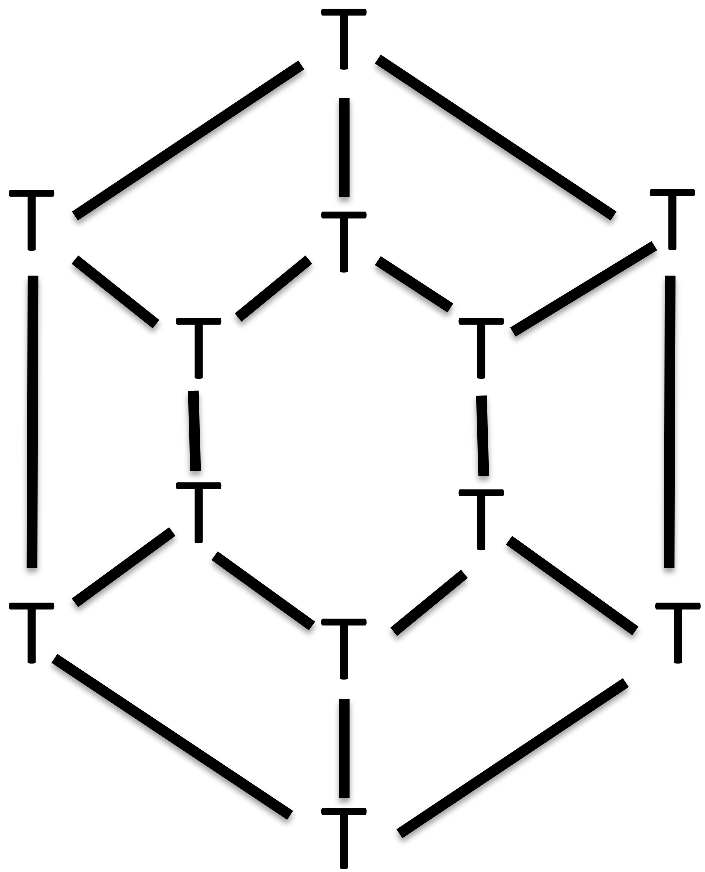
---

Note that the pair of indices {a,b} that index rows of M come from from different copies of T, and this is important

...we get:



…we get:




And $\mathbb{E}[\mathrm{Tr}(M^6)]$ is the number of ways to label the edges of the diagram so that **each triple {i, j, k} appears incident to an even number of T's**.

# SIDE REMARK

The tensor network formalism gives a visual way to understand some subtleties

# SIDE REMARK

The tensor network formalism gives a visual way to understand some subtleties

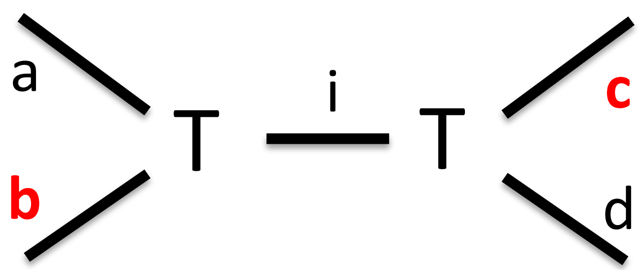
What if we flattened the tensor network differently?

# SIDE REMARK

The tensor network formalism gives a visual way to understand some subtleties

What if we flattened the tensor network differently?

For example, if M is the ({a,b}, {c,d})-flattening of

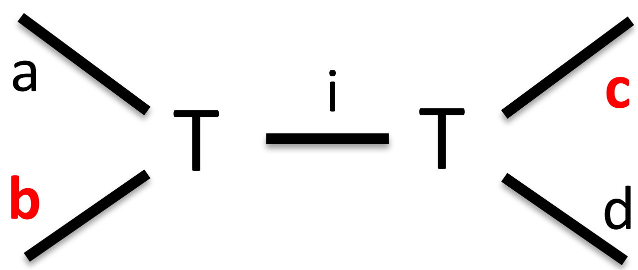


# SIDE REMARK

The tensor network formalism gives a visual way to understand some subtleties

What if we flattened the tensor network differently?

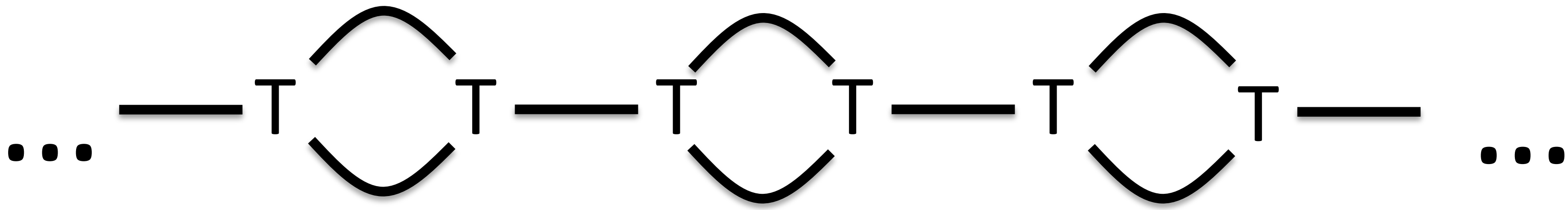
For example, if M is the ({a,b}, {c,d})-flattening of




then plugging it into the six cycle we would get **something different**

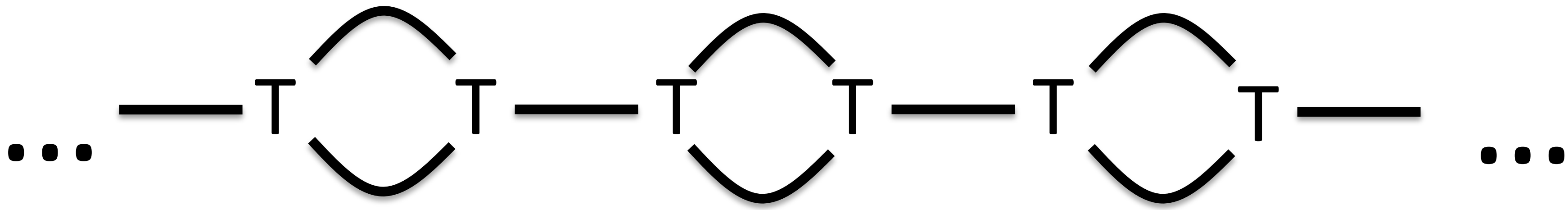
# SIDE REMARK

…we get:



# SIDE REMARK

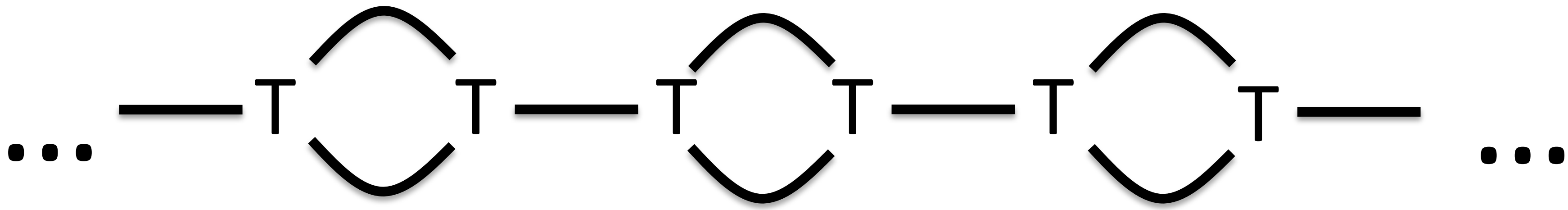
...we get:




**Informal Claim:** There are now many more labellings where each triple is incident to an even number of T's

# SIDE REMARK

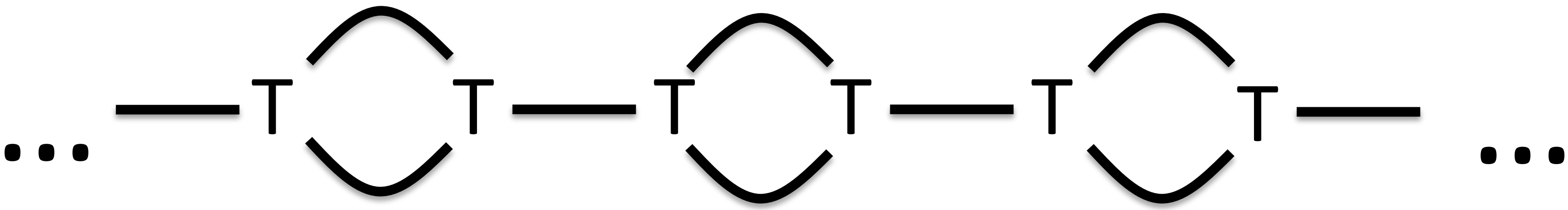
...we get:




**Informal Claim:** There are now many more labellings where each triple is incident to an even number of T's, **because the graph is only 1-connected**

# SIDE REMARK

...we get:

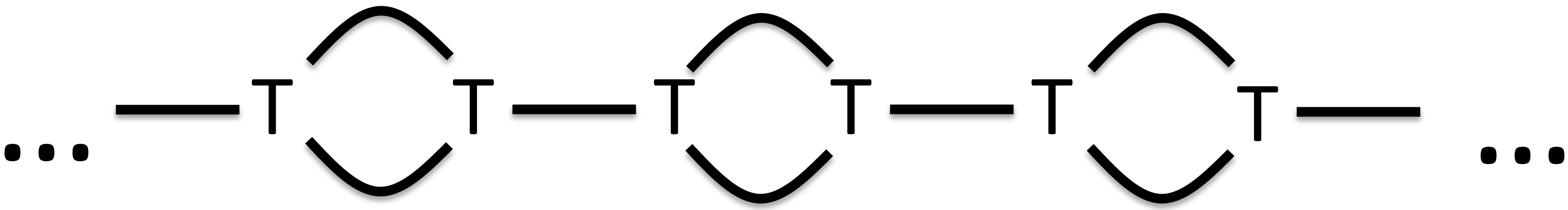



**Informal Claim:** There are now many more labellings where each triple is incident to an even number of T's, **because the graph is only 1-connected**

This trick first appears in **[Coja-Oghlan, Goerdt, Lanka]** to refute random 3-SAT with $n^{3/2}$ clauses

# SIDE REMARK

...we get:




**Informal Claim:** There are now many more labellings where each triple is incident to an even number of T's, **because the graph is only 1-connected**

This trick first appears in **[Coja-Oghlan, Goerdt, Lanka]** to refute random 3-SAT with $n^{3/2}$ clauses

**Tensor networks are a convenient way to think about this trick, and others that appear in the sum-of-squares literature**

# THE TRACE METHOD, WITH GROUPS

The particular labeling problem depends on the group structure

# THE TRACE METHOD, WITH GROUPS

The particular labeling problem depends on the group structure

**Key Observation:** The tensor

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

is sparse in the fourier domain

# THE TRACE METHOD, WITH GROUPS

The particular labeling problem depends on the group structure

**Key Observation:** The tensor

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

is sparse in the fourier domain, because integrating over the group action projects onto the span of degree three invariant polynomials

# THE TRACE METHOD, WITH GROUPS

The particular labeling problem depends on the group structure

**Key Observation:** The tensor

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

is sparse in the fourier domain, because integrating over the group action projects onto the span of degree three invariant polynomials

In the fourier domain, T is supported on indices (i, j, k) where

$$i + j + k = 0$$

and this becomes a constraint in our labeling problem

# THE TRACE METHOD, WITH GROUPS

More generally, for any orbit recovery problem, its group determines the combinatorics of the labeling problem

# THE TRACE METHOD, WITH GROUPS

More generally, for any orbit recovery problem, its group determines the combinatorics of the labeling problem

For more complex groups, can we find good tensor networks?

# THE TRACE METHOD, WITH GROUPS

More generally, for any orbit recovery problem, its group determines the combinatorics of the labeling problem

For more complex groups, can we find good tensor networks?

**We understand how invariant theory governs the statistical complexity of orbit recovery, but not how to get good algorithms!**

# OUTLINE

**Part IV: Invariant Theory**

- Generating the Ring, Generically
- List Recovery for Orbit Retrieval

# OUTLINE

**Part IV: Invariant Theory**

- Generating the Ring, Generically
- List Recovery for Orbit Retrieval

# THE RING OF INVARIANT POLYNOMIALS

**Definition:** The ring of invariant polynomials consists of all polynomials q that satisfy

$$q(\rho(g) \cdot x) = q(x), \forall g \in G$$

# THE RING OF INVARIANT POLYNOMIALS

**Definition:** The ring of invariant polynomials consists of all polynomials q that satisfy

$$q(\rho(g) \cdot x) = q(x), \forall g \in G$$

**Claim:** For discrete MRA, the ring of invariant polynomials is generated by

$$\prod_{j=1}^{s} \widehat{x}_{i_j} \text{ , where } i_1 + i_2 + \cdots + i_s = 0$$

and $\widehat{x}$ is the fourier transform of x

# THE RING OF INVARIANT POLYNOMIALS

**Definition:** The ring of invariant polynomials consists of all polynomials q that satisfy

$$q(\rho(g) \cdot x) = q(x), \forall g \in G$$

**Claim:** For discrete MRA, the ring of invariant polynomials is generated by

$$\prod_{j=1}^{s} \widehat{x}_{i_j} \text{ , where } i_1 + i_2 + \cdots + i_s = 0$$

and $\widehat{x}$ is the fourier transform of x

When s = 3, this is called the **bispectrum** and was introduced by Tukey in a statistical context

# OUTLINE

## Part IV: Invariant Theory

- Generating the Ring, Generically
- List Recovery for Orbit Retrieval

# OUTLINE

**Part IV: Invariant Theory**

- **Generating the Ring, Generically**
- List Recovery for Orbit Retrieval

# INVARIANT THEORY

Invariant polynomials are the key to orbit recovery:

# INVARIANT THEORY

Invariant polynomials are the key to orbit recovery:

**Theorem:** If G is compact and acts continuously then for every x, the invariant ring determines x up to its orbit

# INVARIANT THEORY

Invariant polynomials are the key to orbit recovery:

**Theorem:** If G is compact and acts continuously then for every x, the invariant ring determines x up to its orbit

The **method of moments** asks how many moments are needed to uniquely determine the parameters (e.g. mixtures of gaussians)

# INVARIANT THEORY

Invariant polynomials are the key to orbit recovery:

**Theorem:** If G is compact and acts continuously then for every x, the invariant ring determines x up to its orbit

The **method of moments** asks how many moments are needed to uniquely determine the parameters (e.g. mixtures of gaussians)

It often leads to algorithms with optimal sample complexity

# INVARIANT THEORY

Invariant polynomials are the key to orbit recovery:

**Theorem:** If G is compact and acts continuously then for every x, the invariant ring determines x up to its orbit

The **method of moments** asks how many moments are needed to uniquely determine the parameters (e.g. mixtures of gaussians)

It often leads to algorithms with optimal sample complexity

For orbit recovery, does invariant theory lead to optimal sample complexity?

# INVARIANT THEORY

Invariant polynomials are the key to orbit recovery:

**Theorem:** If G is compact and acts continuously then for every x, the invariant ring determines x up to its orbit

The **method of moments** asks how many moments are needed to uniquely determine the parameters (e.g. mixtures of gaussians)

It often leads to algorithms with optimal sample complexity

For orbit recovery, does invariant theory lead to optimal sample complexity?

There is a subtlety already for discrete MRA…

# GENERATING THE RING

**Fact:** In discrete MRA in d dimensions, the invariant polynomials of degree at most d generate the invariant ring. Moreover lower degree does not!

# GENERATING THE RING

**Fact:** In discrete MRA in d dimensions, the invariant polynomials of degree at most d generate the invariant ring. Moreover lower degree does not!

The problem is with worst-case signals x, and we should be asking the same question for **generic** x

# GENERATING THE RING

**Fact:** In discrete MRA in d dimensions, the invariant polynomials of degree at most d generate the invariant ring. Moreover lower degree does not!

The problem is with worst-case signals x, and we should be asking the same question for **generic** x

**Theorem [Kakarala]:** Provided that

$$\widehat{x}_i \neq 0 \text{ for all } i \in [d]$$

the bispectrum uniquely determines x up to cyclic shift

# GENERATING THE RING

**Fact:** In discrete MRA in d dimensions, the invariant polynomials of degree at most d generate the invariant ring. Moreover lower degree does not!

The problem is with worst-case signals x, and we should be asking the same question for **generic** x

**Theorem [Kakarala]:** Provided that

$$\widehat{x}_i \neq 0 \text{ for all } i \in [d]$$

the bispectrum uniquely determines x up to cyclic shift

More generally, for what degree d* do the polynomials of degree at most d* generate the invariant ring for generic x?

# OUTLINE

**Part IV: Invariant Theory**

- Generating the Ring, Generically
- List Recovery for Orbit Retrieval

# OUTLINE

**Part IV: Invariant Theory**

- Generating the Ring, Generically
- **List Recovery for Orbit Retrieval**

# INVARIANT THEORY

**Theorem [Blum-Smith, Bandeira, Kileel, Perry, Wein, Weed]:**
In orbit recovery on a compact group G

$$y_i = \rho(g_i) \cdot x + \eta_i \text{ , where } \eta_i \sim \mathcal{N}(0, \sigma^2 I)$$

the number of samples needed for **list recovery** generically is

$$\Theta(\sigma^{2d^*})$$

# INVARIANT THEORY

**Theorem [Blum-Smith, Bandeira, Kileel, Perry, Wein, Weed]:**
In orbit recovery on a compact group G

$$y_i = \rho(g_i) \cdot x + \eta_i \text{ , where } \eta_i \sim \mathcal{N}(0, \sigma^2 I)$$

the number of samples needed for **list recovery** generically is

$$\Theta(\sigma^{2d^*})$$

Moreover, they gave computational methods to determine d* that can be run for small enough dimension d

# INVARIANT THEORY

**Theorem [Blum-Smith, Bandeira, Kileel, Perry, Wein, Weed]:**
In orbit recovery on a compact group G

$$y_i = \rho(g_i) \cdot x + \eta_i \text{, where } \eta_i \sim \mathcal{N}(0, \sigma^2 I)$$

the number of samples needed for **list recovery** generically is

$$\Theta(\sigma^{2d^*})$$

Moreover, they gave computational methods to determine d* that can be run for small enough dimension d

**But they do not get algorithms (because they would need to solve systems of polynomial equations to find a signal that is consistent with the system of invariant polynomials)**

# INVARIANT THEORY

**Theorem [Blum-Smith, Bandeira, Kileel, Perry, Wein, Weed]:**
In orbit recovery on a compact group G

$$y_i = \rho(g_i) \cdot x + \eta_i \text{ , where } \eta_i \sim \mathcal{N}(0, \sigma^2 I)$$

the number of samples needed for **list recovery** generically is

$$\Theta(\sigma^{2d^*})$$

Moreover, they gave computational methods to determine d* that can be run for small enough dimension d

# INVARIANT THEORY

**Theorem [Blum-Smith, Bandeira, Kileel, Perry, Wein, Weed]:**
In orbit recovery on a compact group G

$$y_i = \rho(g_i) \cdot x + \eta_i \text{ , where } \eta_i \sim \mathcal{N}(0, \sigma^2 I)$$

the number of samples needed for **list recovery** generically is

$$\Theta(\sigma^{2d^*})$$

Moreover, they gave computational methods to determine d* that can be run for small enough dimension d

**Our tensor network approach can be thought of as a first step towards making this algorithmic! (We get optimal sample complexity for continuous MRA)**

## Summary:

- Orbit recovery is a challenging family of problems over a group action, with many applications
- We gave the first algorithms that work with infinite groups, along with a general framework to design tensor spectral algorithms from tensor networks
- **Is invariant theory the new method of moments?**

**Summary:**

- Orbit recovery is a challenging family of problems over a group action, with many applications
- We gave the first algorithms that work with infinite groups, along with a general framework to design tensor spectral algorithms from tensor networks
- **Is invariant theory the new method of moments?**

# Thanks! Any Questions?