# Multi-Representation Manifold Learning on Fibre Bundles

Tingran Gao

Committee on Computational and Applied Mathematics
Department of Statistics
University of Chicago

Applied Math & Analysis Seminar
Duke University, Durham NC

September 25, 2019

# Outline

Motivation

- Class Averaging and Phase Synchronization

Multi-Frequency Phase Synchronization

- A Multi-Frequency Formulation
- A Proof by Picture

Multi-Frequency Class Averaging

- Some Representation Theoretic Patterns

From a Fibre Bundle Point of View

Joint work with **Yifeng Fan (UIUC)** & **Zhizhen Zhao (UIUC)**

# Outline

**Motivation**

- **Class Averaging and Phase Synchronization**

Multi-Frequency Phase Synchronization

- A Multi-Frequency Formulation
- A Proof by Picture

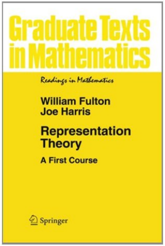
Multi-Frequency Class Averaging

- Some Representation Theoretic Patterns

From a Fibre Bundle Point of View

Joint work with **Yifeng Fan (UIUC)** & **Zhizhen Zhao (UIUC)**

# Representation?

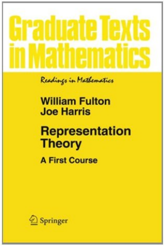



GTM129

Robert Delaunay (1885–1941)

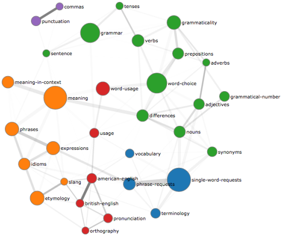

Word2Vec

# Representation?



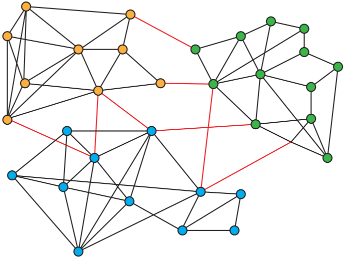

GTM129

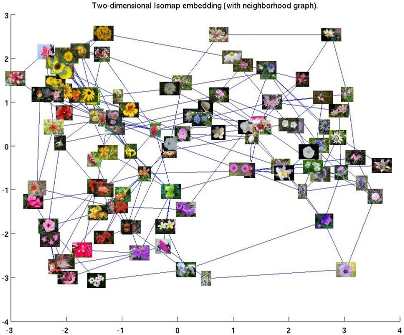
Robert Delaunay (1885–1941)

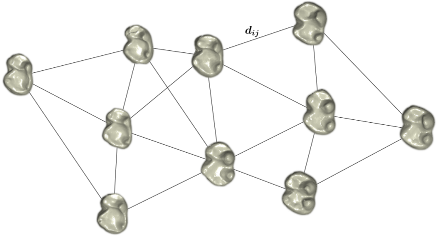


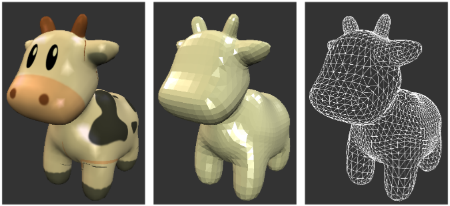

Word2Vec

# Graph: A Flexible Data Representation





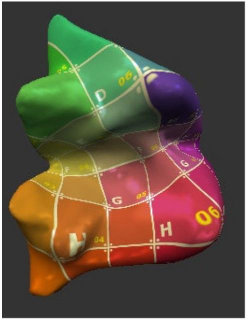



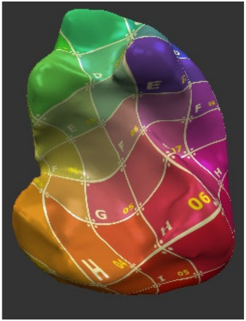
# Non-Scalar Edge Weights

$$d_{\mathrm{cP}}\left(S_i, S_j\right) = \inf_{\mathcal{C}\in\mathcal{A}\left(S_i,S_j\right)} \inf_{R\in\mathbb{E}(3)} \left(\int_{S_i} \left\| R\left(x\right) - \mathcal{C}\left(x\right)\right\|^2 d\mathrm{vol}_{S_i}\left(x\right)\right)^{\frac{1}{2}}$$

$S_i$

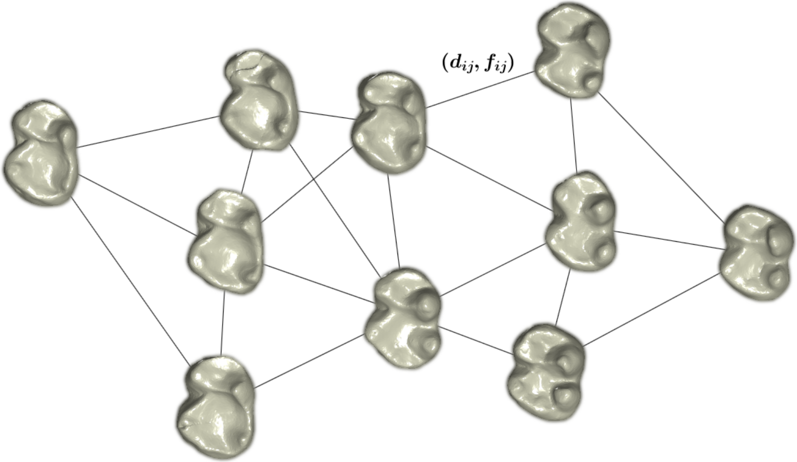


$\xrightarrow[f_{ij}]{d_{ij}}$

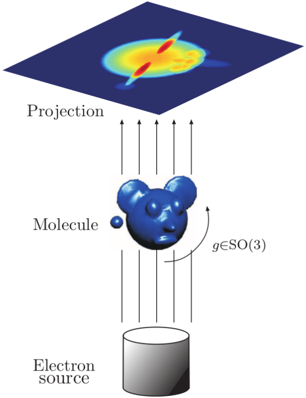


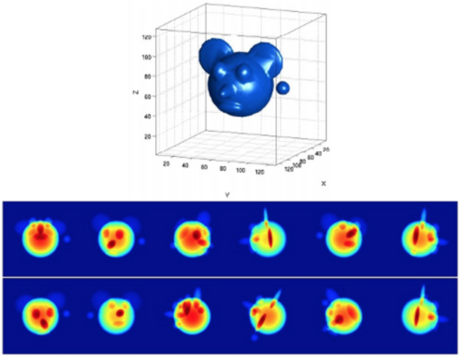
$S_j$

# Do Graph Representations Have Enough Expressive Power?



# Cryo-Electron Microscopy



• Singer et al. "Viewing Angle Classification of Cryo-Electron Microscopy Images using Eigenvectors", *SIAM Journal on Imaging Sciences*, 4 (2), pp. 543–572 (2011).

# Cryo-Electron Microscopy: Real Challenge is Low SNR

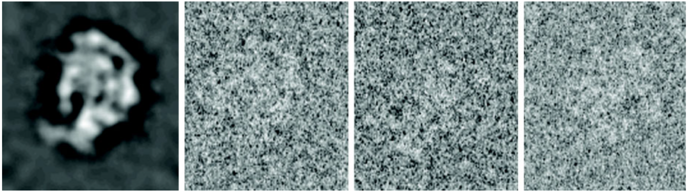


**Fig. 3** The left most image is a clean simulated projection image of the E.coli 50S ribosomal subunit. The other three images are real electron microscope images of the same subunit

Apply **Class Averaging** to improve SNR!

- For each image, identify nearest neighbors in terms of similar viewing directions
- Average out the image with the identified neighbor images (with respect to the correct pairwise rotations)

• Hadani & Singer. "Representation Theoretic Patterns in Three-Dimensional Cryo-Electron Microscopy II – The Class Averaging Problem," *Foundations of Computational Mathematics*, 11 (5), pp. 589–616 (2011).

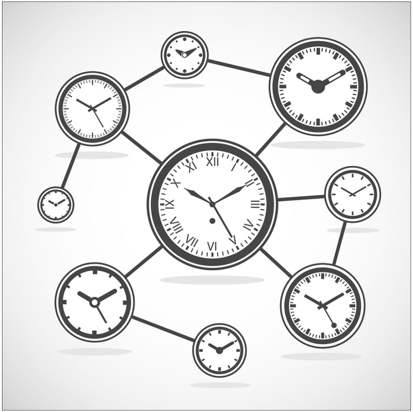
# Phase Synchronization

- **Problem:** Recover rotation angles $\theta_1, \ldots, \theta_n \in [0, 2\pi]$ from noisy measurements of their pairwise offsets

$$\theta_{ij} \equiv \theta_i - \theta_j + \text{noise}$$

for some or all pairs of $(i, j)$

- **Examples:** Class averaging in cryo-EM image analysis, shape registration and community detection



# Phase Synchronization (Notation: $\mathbb{C}_1^n := [\mathrm{U}(1)]^n$)

- **Setup**: Phase vector $z = \left(e^{\iota\theta_1}, \ldots, e^{\iota\theta_n}\right)^\top \in \mathbb{C}_1^n$, noisy pairwise measurements in $n$-by-$n$ Hermitian matrix

$$H_{ij} = \begin{cases} e^{\iota\left(\theta_i - \theta_j\right)} = z_i \bar{z}_j & \text{with prob. } r \in [0,1] \\ \text{Uniform}\left(\mathrm{U}(1)\right) & \text{with prob. } 1 - r \end{cases}$$

  and $H_{ij} = \overline{H_{ji}}$. This is known as a random corruption model.
- **Goal**: Recover the true phase vector $z$ (up to a global multiplicative factor)
- **Existing method**: Rank-1 recovery (e.g. convex relaxations)

$$\hat{x} := \underset{x \in \mathbb{C}_1^n}{\arg\min} \|xx^* - H\|_{\mathrm{F}}^2 \quad \Leftrightarrow \quad \hat{x} := \underset{x \in \mathbb{C}_1^n}{\arg\max}\, x^* H x$$

# Phase Synchronization (Notation: $\mathbb{C}_1^n := [\mathrm{U}(1)]^n$)

- **Setup**: Phase vector $z = \left(e^{\iota\theta_1}, \ldots, e^{\iota\theta_n}\right)^\top \in \mathbb{C}_1^n$, noisy pairwise measurements in $n$-by-$n$ Hermitian matrix

$$H_{ij} = \begin{cases} e^{\iota\left(\theta_i - \theta_j\right)} = z_i \bar{z}_j & \text{with prob. } r \in [0,1] \\ \mathrm{Uniform}\left(\mathrm{U}(1)\right) & \text{with prob. } 1 - r \end{cases}$$

  and $H_{ij} = \overline{H_{ji}}$. This is known as a random corruption model.
- **Goal**: Recover the true phase vector $z$ (up to a global multiplicative factor)
- **Spectral Relaxation**: solve for the top eigenvector of $H$, denoted as $\tilde{x}$ (scaled to $\|\tilde{x}\|_2 = \sqrt{n}$), then define $\hat{x} \in \mathbb{C}_1^n$ by

$$\hat{x}_i := \tilde{x}_i / \left|\tilde{x}_i\right|$$

# Outline

Motivation

- Class Averaging and Phase Synchronization

**Multi-Frequency Phase Synchronization**

- **A Multi-Frequency Formulation**
- A Proof by Picture

Multi-Frequency Class Averaging

- Some Representation Theoretic Patterns

From a Fibre Bundle Point of View

Joint work with **Yifeng Fan (UIUC)** & **Zhizhen Zhao (UIUC)**

# Multi-Frequency Phase Synchronization: Main Idea

- The rank-1 recovery formulation
$$\hat{x} := \underset{x \in \mathbb{C}_1^n}{\arg\min} \|xx^* - H\|_{\mathrm{F}}^2 \quad \Leftrightarrow \quad \hat{x} := \underset{x \in \mathbb{C}_1^n}{\arg\max}\, x^* H x$$
does not fully exploit that entries of $x$ and $H$ are **phases**!

# Multi-Frequency Phase Synchronization: Main Idea

- The rank-1 recovery formulation
$$\hat{x} := \arg\min_{x \in \mathbb{C}_1^n} \|xx^* - H\|_{\mathrm{F}}^2 \quad \Leftrightarrow \quad \hat{x} := \arg\max_{x \in \mathbb{C}_1^n} x^* H x$$
does not fully exploit that entries of $x$ and $H$ are **phases**!
- **Key Observation:** Raising a phase to any power yields another phase! $e^{\iota\theta} \longmapsto e^{\iota k\theta}, \quad k = 1, 2, \ldots$

# Multi-Frequency Phase Synchronization: Main Idea

- The rank-1 recovery formulation
$$\hat{x} := \operatorname*{arg\,min}_{x \in \mathbb{C}_1^n} \|xx^* - H\|_{\mathrm{F}}^2 \quad \Leftrightarrow \quad \hat{x} := \operatorname*{arg\,max}_{x \in \mathbb{C}_1^n} x^* H x$$
does not fully exploit that entries of $x$ and $H$ are **phases**!
- **Key Observation:** Raising a phase to any power yields another phase! $e^{\iota\theta} \longmapsto e^{\iota k\theta}, \quad k = 1, 2, \ldots$
- Solve a family of coupled matrix factorization problems jointly
$$\hat{x}^{(k)} := \operatorname*{arg\,max}_{x \in \mathbb{C}_1^n} (x^k)^* H^{(k)} x^k, \qquad k = 1, 2, \ldots, k_{\max}$$
where $x^k := \left(x_1^k, \ldots, x_n^k\right)^\top \in \mathbb{C}_1^n$, and $H^{(k)}$ is the $n$-by-$n$ Hermitian matrix with $H_{ij}^{(k)} := H_{ij}^k$, and then "**stitch up**" the individual estimates $\hat{x}^{(1)}, \ldots, \hat{x}^{(k_{\max})}$ to recover $\hat{x}$

# Multi-Frequency Phase Synchronization: Main Idea

- The rank-1 recovery formulation

$$\hat{x} := \arg\min_{x\in\mathbb{C}_1^n} \|xx^* - H\|_{\mathrm{F}}^2 \quad \Leftrightarrow \quad \hat{x} := \arg\max_{x\in\mathbb{C}_1^n} x^* H x$$

does not fully exploit that entries of $x$ and $H$ are **phases**!

- **Key Observation:** Raising a phase to any power yields another phase! $e^{\iota\theta} \longmapsto e^{\iota k\theta}, \quad k = 1, 2, \ldots$
- Solve a family of coupled matrix factorization problems jointly

$$\hat{x}^{(k)} := \arg\max_{x\in\mathbb{C}_1^n} (x^k)^* H^{(k)} x^k, \qquad k = 1, 2, \ldots, k_{\max}$$

where $x^k := \left(x_1^k, \ldots, x_n^k\right)^\top \in \mathbb{C}_1^n$, and $H^{(k)}$ is the $n$-by-$n$ Hermitian matrix with $H_{ij}^{(k)} := H_{ij}^k$, and then "**stitch up**" the individual estimates $\hat{x}^{(1)}, \ldots, \hat{x}^{(k_{\max})}$ to recover $\hat{x}$

- The "**stitching**" step strives to recover $e^{\iota\theta}$ from noisy measurements of $e^{\iota\theta}, \ldots, e^{\iota k_{\max}\theta}$, which is a version of **harmonic retrieval**

# Multi-Frequency Phase Synchronization

- **Multi-Frequency Formulation:**

$$\boxed{\max_{x\in\mathbb{C}_1^n} \sum_{k=1}^{k_{\max}} (x^k)^* H^{(k)} x^k}$$

where $x^k := \left(x_1^k, \ldots, x_n^k\right)^\top \in \mathbb{C}_1^n$, and $H^{(k)}$ is the $n$-by-$n$ Hermitian matrix with $H_{ij}^{(k)} := H_{ij}^k$

- **Intuition:** Matching higher trigonometric moments
- **Two-stage Algorithm:** (i) Good initialization (ii) Local methods e.g. gradient descent or (generalized) power iteration

# Initialization: Inspired by Harmonic Retrieval

- Fix $k_{\max} \geq 1$, build $H^{(2)}, \ldots, H^{(k_{\max})}$ out of $H = H^{(1)}$ by taking entrywise powers of $H$

# Initialization: Inspired by Harmonic Retrieval

- Fix $k_{\max} \geq 1$, build $H^{(2)}, \ldots, H^{(k_{\max})}$ out of $H = H^{(1)}$ by taking entrywise powers of $H$
- For each $k = 1, \ldots, k_{\max}$, find a reasonably good symmetric rank-1 approximation

$$W^{(k)} := \underset{\substack{Y = Y^\top \\ \mathrm{rank}(Y) = 1}}{\arg\max} \left\| H^{(k)} - Y \right\|_F^2$$

using e.g. spectral method

# Initialization: Inspired by Harmonic Retrieval

- Fix $k_{\max} \geq 1$, build $H^{(2)}, \ldots, H^{(k_{\max})}$ out of $H = H^{(1)}$ by taking entrywise powers of $H$
- For each $k = 1, \ldots, k_{\max}$, find a reasonably good symmetric rank-1 approximation

$$W^{(k)} := \underset{\substack{Y=Y^\top \\ \mathrm{rank}(Y)=1}}{\arg\max} \left\| H^{(k)} - Y \right\|_F^2$$

  using e.g. spectral method
- For all $1 \leq i, j \leq n$, find the "peak location" of the spectrogram

$$\hat{\theta}_{ij} := \underset{\phi \in [0, 2\pi]}{\arg\max} \left| \frac{1}{2} \sum_{k=-k_{\max}}^{k_{\max}} W_{ij}^{(k)} e^{-\iota k \phi} \right|$$

# Initialization: Inspired by Harmonic Retrieval

- Fix $k_{\max} \geq 1$, build $H^{(2)}, \ldots, H^{(k_{\max})}$ out of $H = H^{(1)}$ by taking entrywise powers of $H$
- For each $k = 1, \ldots, k_{\max}$, find a reasonably good symmetric rank-1 approximation

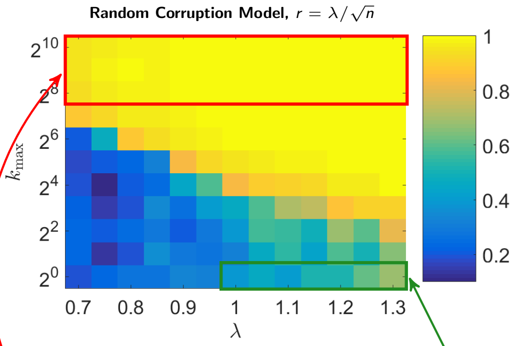
$$W^{(k)} := \operatorname*{arg\,max}_{\substack{Y = Y^\top \\ \mathrm{rank}(Y)=1}} \left\| H^{(k)} - Y \right\|_F^2$$

  using e.g. spectral method
- For all $1 \leq i, j \leq n$, find the "peak location" of the spectrogram

$$\hat{\theta}_{ij} := \operatorname*{arg\,max}_{\phi \in [0, 2\pi]} \left| \frac{1}{2} \sum_{k=-k_{\max}}^{k_{\max}} W_{ij}^{(k)} e^{-\iota k \phi} \right|$$

- Apply the spectral method yet again to the Hermitian matrix $\widehat{H}$ to get $\hat{x} \in \mathbb{C}_1^n$, where $\widehat{H}_{ij} = e^{\iota \hat{\theta}_{ij}}$

# How well does it work? Evaluate correlation $|\mathrm{Corr}\,(\hat{x}, z)|$




**Our Method:** $|\mathrm{Corr}\,(\hat{x}, z)| \longrightarrow 1$ as $k_{\max} \gg 1$, even for $\lambda < 1$!

**Previous Art:** Only ensures $|\mathrm{Corr}\,(\hat{x}, z)| > \frac{1}{\sqrt{n}}$ for $\lambda > 1$

# Outline

Motivation

- Class Averaging and Phase Synchronization

**Multi-Frequency Phase Synchronization**

- A Multi-Frequency Formulation
- **A Proof by Picture**

Multi-Frequency Class Averaging

- Some Representation Theoretic Patterns

From a Fibre Bundle Point of View

Joint work with **Yifeng Fan (UIUC)** & **Zhizhen Zhao (UIUC)**

# New Theory: Strong Recovery

**Theorem (G.–Zhao 2019).** Under (mild) assumptions, with high probability, the multi-frequency phase synchronization algorithm produces an estimate $\hat{x}$ satisfying

$$|\mathrm{Corr}\,(\hat{x}, z)| \geq 1 - \frac{C'}{k_{\max}^2}$$

provided that

$$k_{\max} > \max\left\{5, \frac{1}{\sqrt{2}\,\pi - 2 - 4\sqrt{2}\,\pi C_2 \sigma \sqrt{\log n / n}}\right\}.$$

**In particular, $|\mathrm{Corr}\,(\hat{x}, z)| \to 1$ as $k_{\max} \to \infty$.**

# Why does it work?

- By a perturbation analysis, after solving each subproblem

$$W^{(k)} := \underset{\substack{Y=Y^\top \\ \mathrm{rank}(Y)=1}}{\arg\max} \left\| H^{(k)} - Y \right\|_F^2$$

we expect $W^{(k)} = z^k (z^k)^* + E^{(k)} \approx z^k (z^k)^*$, where $z_i = e^{\iota \theta_i}$

- The peak finding step is expected to ensure

$$\begin{aligned} \hat{\theta}_{ij} &= \underset{\phi \in [0, 2\pi]}{\arg\max} \left| \frac{1}{2} \sum_{k=1}^{k_{\max}} W_{ij}^{(k)} e^{-\iota k \phi} \right| \\ &\approx \underset{\phi \in [0, 2\pi]}{\arg\max} \left| \frac{1}{2} \sum_{k=1}^{k_{\max}} e^{\iota k (\theta_i - \theta_j)} e^{-\iota k \phi} \right| = \theta_i - \theta_j \end{aligned}$$

provided that $E^{(k)}$ does not “perturb away” the maximum!

# Landscape Analysis of the Dirichlet Kernel

$$\sum_{k=-k_{\max}}^{k_{\max}} W_{ij}^{(k)} e^{-\iota k \phi}$$

$$= \sum_{k=-k_{\max}}^{k_{\max}} e^{\iota k \left(\theta_i - \theta_j\right)} e^{-\iota k \phi} + \sum_{k=-k_{\max}}^{k_{\max}} E_{ij}^{(k)} e^{-\iota k \phi}$$

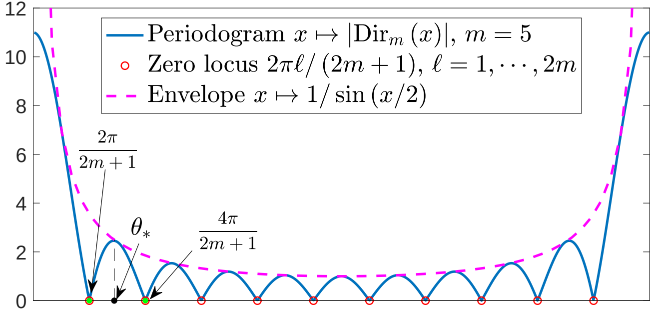
# Landscape Analysis of the Dirichlet Kernel

$$
\begin{aligned}
&\sum_{k=-k_{\max}}^{k_{\max}} W_{ij}^{(k)} e^{-\iota k \phi} \\
&= \sum_{k=-k_{\max}}^{k_{\max}} e^{\iota k \left(\theta_i - \theta_j\right)} e^{-\iota k \phi} + \sum_{k=-k_{\max}}^{k_{\max}} {\color{red} E_{ij}^{(k)}} e^{-\iota k \phi} \\
&= \frac{\sin\left[\left(k_{\max} + \frac{1}{2}\right)\left(\theta_i - \theta_j - \phi\right)\right]}{\sin\left[\frac{1}{2}\left(\theta_i - \theta_j - \phi\right)\right]} + \sum_{k=-k_{\max}}^{k_{\max}} {\color{red} E_{ij}^{(k)}} e^{-\iota k \phi} \\
&=: \mathrm{Dir}_{k_{\max}}\left(\phi - \left(\theta_i - \theta_j\right)\right) + R_{k_{\max}}\left(\phi\right)
\end{aligned}
$$

# A Proof by Picture (Notation: $\Delta\theta_{ij} := \hat{\theta}_{ij} - (\theta_i - \theta_j)$)

**Key Observation:** $|\Delta\theta_{ij}| \leq \theta_* < \frac{4\pi}{2k_{\max}+1}$ whenever

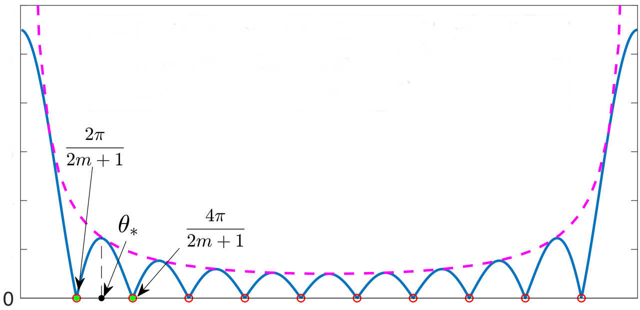
$$2k_{\max} + 1 - \|R_{k_{\max}}\|_\infty > \frac{1}{\sin(\theta_*/2)} + \|R_{k_{\max}}\|_\infty$$



# A Proof by Picture (Notation: $\Delta\theta_{ij} := \hat{\theta}_{ij} - (\theta_i - \theta_j)$)

**Key Observation:** $|\Delta\theta_{ij}| \leq \theta_* < \frac{4\pi}{2k_{\max}+1}$ whenever

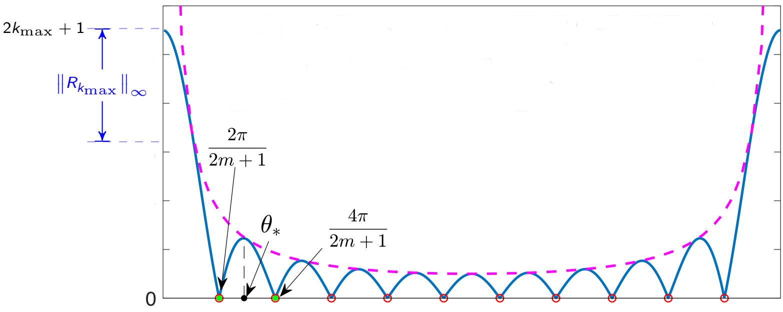
$$2k_{\max} + 1 - \|R_{k_{\max}}\|_\infty > \frac{1}{\sin(\theta_*/2)} + \|R_{k_{\max}}\|_\infty$$



# A Proof by Picture (Notation: $\Delta\theta_{ij} := \hat{\theta}_{ij} - (\theta_i - \theta_j)$)

**Key Observation:** $|\Delta\theta_{ij}| \leq \theta_* < \frac{4\pi}{2k_{\max}+1}$ whenever

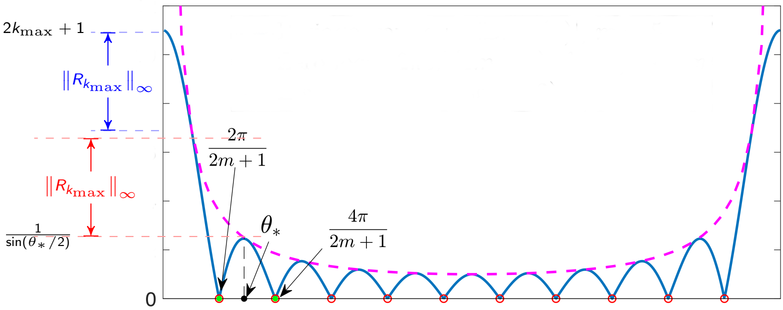
$$2k_{\max} + 1 - \|R_{k_{\max}}\|_\infty > \frac{1}{\sin(\theta_*/2)} + \|R_{k_{\max}}\|_\infty$$



# A Proof by Picture (Notation: $\Delta\theta_{ij} := \hat{\theta}_{ij} - (\theta_i - \theta_j)$)

**Key Observation:** $|\Delta\theta_{ij}| \leq \theta_* < \frac{4\pi}{2k_{\max}+1}$ whenever

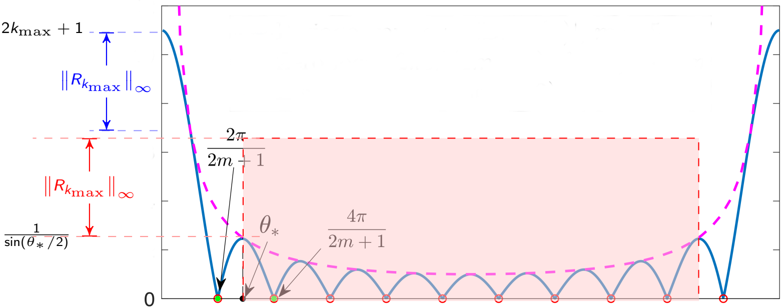
$$2k_{\max} + 1 - \|R_{k_{\max}}\|_\infty > \frac{1}{\sin(\theta_*/2)} + \|R_{k_{\max}}\|_\infty$$



# A Proof by Picture (Notation: $\Delta\theta_{ij} := \hat{\theta}_{ij} - (\theta_i - \theta_j)$)

**Key Observation:** $|\Delta\theta_{ij}| \leq \theta_* < \frac{4\pi}{2k_{\max}+1}$ whenever

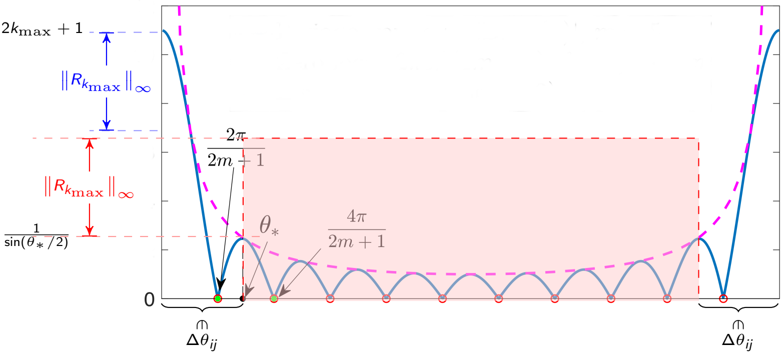
$$2k_{\max} + 1 - \|R_{k_{\max}}\|_\infty > \frac{1}{\sin(\theta_*/2)} + \|R_{k_{\max}}\|_\infty$$



# A Proof by Picture (Notation: $\Delta\theta_{ij} := \hat{\theta}_{ij} - (\theta_i - \theta_j)$)

**Key Observation:** $|\Delta\theta_{ij}| \leq \theta_* < \frac{4\pi}{2k_{\max}+1}$ whenever

$$2k_{\max} + 1 - \|R_{k_{\max}}\|_\infty > \frac{1}{\sin(\theta_*/2)} + \|R_{k_{\max}}\|_\infty$$



# A Proof by Picture (Notation: $\Delta\theta_{ij} := \hat{\theta}_{ij} - (\theta_i - \theta_j)$)

**Key Observation:** $|\Delta\theta_{ij}| \leq \theta_* < \frac{4\pi}{2k_{\max}+1}$ whenever

$$2k_{\max} + 1 - \|R_{k_{\max}}\|_\infty > \frac{1}{\sin(\theta_*/2)} + \|R_{k_{\max}}\|_\infty$$

# A Proof by Picture (Notation: $\Delta\theta_{ij} := \hat{\theta}_{ij} - (\theta_i - \theta_j)$)

**Key Observation:** $|\Delta\theta_{ij}| \leq \theta_* < \frac{4\pi}{2k_{\max}+1}$ whenever

$$2k_{\max} + 1 - \|R_{k_{\max}}\|_\infty > \frac{1}{\sin\left(\theta_*/2\right)} + \|R_{k_{\max}}\|_\infty$$

$$\Leftarrow \; 2k_{\max} - \|R_{k_{\max}}\|_\infty > \frac{1}{\sin\left(\frac{\pi}{2k_{\max}+1}\right)} + \|R_{k_{\max}}\|_\infty$$

# A Proof by Picture (Notation: $\Delta\theta_{ij} := \hat{\theta}_{ij} - (\theta_i - \theta_j)$)

**Key Observation:** $|\Delta\theta_{ij}| \leq \theta_* < \frac{4\pi}{2k_{\max}+1}$ whenever

$$2k_{\max} + 1 - \|R_{k_{\max}}\|_\infty > \frac{1}{\sin(\theta_*/2)} + \|R_{k_{\max}}\|_\infty$$

$$\Leftarrow\ 2k_{\max} - \|R_{k_{\max}}\|_\infty > \frac{1}{\sin\left(\frac{\pi}{2k_{\max}+1}\right)} + \|R_{k_{\max}}\|_\infty$$

$$\Leftarrow\ 1 - \frac{1}{2k_{\max}\sin\left(\frac{\pi}{2k_{\max}+1}\right)} > \frac{1}{k_{\max}}\|R_{k_{\max}}\|_\infty \sim \left\|E_{ij}^{(k)}\right\|_\infty$$

**What are the odds?**
**Can be estimated with a uniform upper bound for the $E_{ij}^{(k)}$'s!**

# Eigenvector perturbation analysis in $\ell_\infty$-norm!

- Standard perturbation bounds use $\ell_2$-norm (e.g. Davis–Kahan), but most application scenarios of spectral methods require bounding the $\ell_\infty$-norm
- Active research area in recent years, e.g. **[Eldridge et al. (2017)]; [Abbe et al. (2017)]; [Fan et al. (2018)]; [Zhong & Boumal (2018)]**
- Sharpest results to date use a "leave-one-out" trick popularized by statisticians

**Lemma (G.–Zhao 2019).** For any $0 < \epsilon \leq 2$, with probability at least $1 - O\left(n^{-(2+\epsilon)}\right)$, there exists absolute constant $C_2 > 0$ s.t.

$$\left\| E_{ij}^{(k)} \right\|_\infty \leq C_2 \sigma \sqrt{\frac{\log n}{n}}.$$

# Putting Everything Together, with a Union Bound

With high probability, uniformly for all $i, j \in [n]$, $\hat{\theta}_{ij}$ is close to the true offset $\theta_i - \theta_j$ as long as

$$
\begin{aligned}
& 1 - \frac{1}{2k_{\max} \sin\left(\frac{\pi}{2k_{\max}+1}\right)} \geq \left\| E_{ij}^{(k)} \right\|_\infty \\
\Leftarrow\ & 1 - \frac{1}{2k_{\max} \sin\left(\frac{\pi}{2k_{\max}+1}\right)} > C_2 \sigma \sqrt{\frac{\log n}{n}} \\
\Leftarrow\ & k_{\max} > \max\left\{ 5, \frac{1}{\sqrt{2}\,\pi - 2 - 4\sqrt{2}\,\pi C_2 \sigma \sqrt{\log n / n}} \right\}
\end{aligned}
$$

# New Theory: Strong Recovery

**Theorem (G.–Zhao 2019).** Under (mild) assumptions, with high probability, the multi-frequency phase synchronization algorithm produces an estimate $\hat{x}$ satisfying

$$|\mathrm{Corr}\,(\hat{x}, z)| \geq 1 - \frac{C'}{k_{\mathrm{max}}^2}$$

provided that

$$k_{\mathrm{max}} > \max\left\{5, \frac{1}{\sqrt{2}\,\pi - 2 - 4\sqrt{2}\,\pi C_2 \sigma \sqrt{\log n / n}}\right\}.$$

**In particular, $|\mathrm{Corr}\,(\hat{x}, z)| \to 1$ as $k_{\mathrm{max}} \to \infty$.**

# Detour: Multi-Frequency Synchronization over $\mathrm{SO}(3)$

**Peter–Weyl:** $f(g) = \sum_{k=0}^{\infty} d_k \mathrm{Tr}\left[\hat{f}(k)\,\rho_k(g)\right], \ \forall f \in L^2(\mathrm{SO}(3))$

- Random corruption model:

$$g_{ij} = \begin{cases} g_i g_j^{-1} & \text{with probability } r \\ \mathrm{Unif}\left(\mathrm{SO}(3)\right) & \text{with probability } 1-r \end{cases}$$

- Use spectral methods to estimate $\rho_1(g_{ij}), \ldots, \rho_{k_{\max}}(g_{ij})$, denote $\widehat{H}_{ij}^{(k)}$ for the estimator of $\rho_k(g_{ij})$
- Solve a generalized harmonic retrieval problem on $\mathrm{SO}(3)$:

$$\hat{g}_{ij} = \underset{g\in\mathrm{SO}(3)}{\arg\max} \sum_{k=1}^{k_{\max}} d_k \mathrm{Tr}\left[\widehat{H}_{ij}^{(k)} \rho_k^{*}(g)\right]$$

- Works fantastic in practice, but no theory yet!

# Outline


Motivation

- Class Averaging and Phase Synchronization

Multi-Frequency Phase Synchronization

- A Multi-Frequency Formulation
- A Proof by Picture

**Multi-Frequency Class Averaging**

- **Some Representation Theoretic Patterns**

From a Fibre Bundle Point of View


Joint work with **Yifeng Fan (UIUC)** & **Zhizhen Zhao (UIUC)**

# Class Averaging

- Compute the *rotation-invariant distance* between all pairs of images $d_{\mathrm{RID}}(I_i, I_j) := \min_{\alpha \in [0,2\pi]} \|I_i - e^{\iota\alpha} I_j\|_{\mathrm{F}}$, and denote $\alpha_{ij}$ for the optimal alignment angle

# Class Averaging

- Compute the *rotation-invariant distance* between all pairs of images $d_{\mathrm{RID}}(I_i, I_j) := \min_{\alpha\in[0,2\pi]} \left\| I_i - e^{\iota\alpha} I_j \right\|_{\mathrm{F}}$, and denote $\alpha_{ij}$ for the optimal alignment angle
- Fix threshold $\epsilon > 0$ and define Hermitian $W \in \mathbb{C}^{n\times n}$ by

$$W_{ij} := \begin{cases} \exp(\iota\alpha_{ij}) & \text{if } d_{\mathrm{RID}}(I_i, I_j) < \epsilon \\ 0 & \text{otherwise} \end{cases}$$

# Class Averaging

- Compute the *rotation-invariant distance* between all pairs of images $d_{\mathrm{RID}}\left(I_i, I_j\right) := \min_{\alpha\in[0,2\pi]} \left\|I_i - e^{\iota\alpha} I_j\right\|_{\mathrm{F}}$, and denote $\alpha_{ij}$ for the optimal alignment angle
- Fix threshold $\epsilon > 0$ and define Hermitian $W \in \mathbb{C}^{n\times n}$ by

$$W_{ij} := \begin{cases} \exp\left(\iota\alpha_{ij}\right) & \text{if } d_{\mathrm{RID}}\left(I_i, I_j\right) < \epsilon \\ 0 & \text{otherwise} \end{cases}$$

- Solve for the top 3 eigenvectors $\psi_1, \psi_2, \psi_3$ of $W$, which embeds $I_1, I_2, \ldots$ into $\mathbb{C}^3$ by

$$I_i \longmapsto \Psi\left(I_i\right) := \frac{\left(\psi_1\left(i\right), \psi_2\left(i\right), \psi_3\left(i\right)\right)}{\left\|\left(\psi_1\left(i\right), \psi_2\left(i\right), \psi_3\left(i\right)\right)\right\|}$$

# Class Averaging

- Compute the *rotation-invariant distance* between all pairs of images $d_{\mathrm{RID}}(I_i, I_j) := \min_{\alpha\in[0,2\pi]} \|I_i - e^{\iota\alpha} I_j\|_{\mathrm{F}}$, and denote $\alpha_{ij}$ for the optimal alignment angle
- Fix threshold $\epsilon > 0$ and define Hermitian $W \in \mathbb{C}^{n\times n}$ by

$$W_{ij} := \begin{cases} \exp(\iota\alpha_{ij}) & \text{if } d_{\mathrm{RID}}(I_i, I_j) < \epsilon \\ 0 & \text{otherwise} \end{cases}$$

- Solve for the top 3 eigenvectors $\psi_1, \psi_2, \psi_3$ of $W$, which embeds $I_1, I_2, \ldots$ into $\mathbb{C}^3$ by

$$I_i \longmapsto \Psi(I_i) := \frac{(\psi_1(i), \psi_2(i), \psi_3(i))}{\|(\psi_1(i), \psi_2(i), \psi_3(i))\|}$$

- Use correlation in the embedded $\mathbb{C}^3$ space to determine the closeness between viewing directions

# Multi-Frequency Class Averaging

- Compute the *rotation-invariant distance* between all pairs of images $d_{\mathrm{RID}}(I_i, I_j) := \min_{\alpha \in [0,2\pi]} \|I_i - e^{\iota\alpha} I_j\|_{\mathrm{F}}$, and denote $\alpha_{ij}$ for the optimal alignment angle
- Fix threshold $\epsilon > 0$ and define Hermitian $W \in \mathbb{C}^{n \times n}$ by

$$W_{ij}^{(k)} := \begin{cases} \exp(\iota k \alpha_{ij}) & \text{if } d_{\mathrm{RID}}(I_i, I_j) < \epsilon \\ 0 & \text{otherwise} \end{cases}$$

- Solve for the top $2k+1$ eigenvectors $\psi_1^{(k)}, \cdots, \psi_{2k+1}^{(k)}$ of $W^{(k)}$, which embeds $I_1, I_2, \ldots$ into $\mathbb{C}^{2k+1}$ by

$$I_i \longmapsto \Psi^{(k)}(I_i) := \frac{(\psi_1(i), \ldots, \psi_{2k+1}(i))}{\|(\psi_1(i), \ldots, \psi_{2k+1}(i))\|}$$

- Use all correlations in the embedded $\mathbb{C}^{2k+1}$ $(k = 1, \ldots, k_{\max})$ spaces to determine the closeness between viewing directions

# Why $2k+1$? Some Representation Theoretic Patterns

- **(G., Fan, Zhao 2019)** For sufficiently large sample size $n$ and appropriately small $\epsilon > 0$, the top eigenspace of $W^{(k)}$ is $(2k+1)$-dimensional, and the spectral gap grows linearly in $k$:

$$\lambda_k^{(k)} - \lambda_{k+1}^{(k)} \sim \frac{1+k}{4}\epsilon^2$$

- Larger $k$ $\Rightarrow$ larger spectral gap $\Rightarrow$ better numerical stability!

• Tingran Gao, Yifeng Fan, Zhizhen Zhao. *Representation Theoretic Patterns in Multi-Frequency Class Averaging for Three-Dimensional Cryo-Electron Microscopy.* arxiv:1906.01082.

# More Representation Theoretic Patterns......

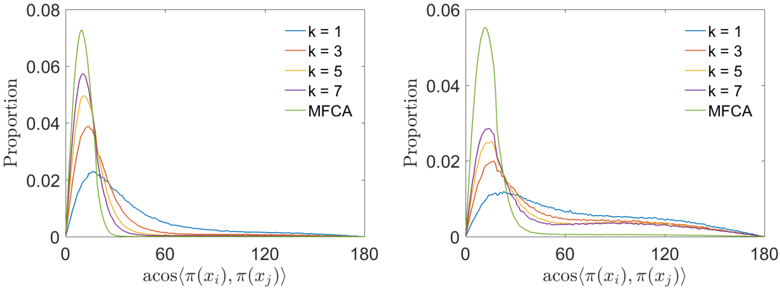
- **(G.–Fan–Zhao 2019)** The viewing angle $\theta_{ij}$ between image $i$ and image $j$ satisfies

$$\left|\left\langle \Psi^{(k)}\left(I_i\right), \Psi^{(k)}\left(I_j\right)\right\rangle\right| = \left(\frac{1+\cos\theta_{ij}}{2}\right)^k$$

- Larger $k \Rightarrow$ easier thresholding
- Can also jointly use $k = 1, \cdots, k_{\max}$ to construct polynomial filters for $\cos\theta_{ij}$

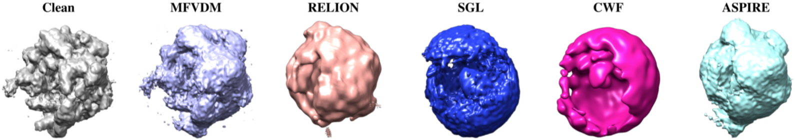
• Tingran Gao, Yifeng Fan, Zhizhen Zhao. *Representation Theoretic Patterns in Multi-Frequency Class Averaging for Three-Dimensional Cryo-Electron Microscopy.* arxiv:1906.01082.

# Multi-Frequency Information Improves Class Averaging




Histograms of true viewing angles between each image and its 50 nearest neighboring images

# Multi-Frequency Information Improves Class Averaging




- Yifeng Fan & Zhizhen Zhao. *Cryo-Electron Microscopy Image Analysis Using Multi-Frequency Vector Diffusion Maps.* arXiv:1904.07772.
- Tingran Gao, Yifeng Fan, Zhizhen Zhao. *Representation Theoretic Patterns in Multi-Frequency Class Averaging for Three-Dimensional Cryo-Electron Microscopy.* arxiv:1906.01082.

# Outline

Motivation

- Class Averaging and Phase Synchronization

Multi-Frequency Phase Synchronization

- A Multi-Frequency Formulation
- A Proof by Picture

Multi-Frequency Class Averaging

- Some Representation Theoretic Patterns

**From a Fibre Bundle Point of View**

Joint work with **Yifeng Fan (UIUC)** & **Zhizhen Zhao (UIUC)**

# Synchronization Problems

- Data:
  - graph $\Gamma = (V, E)$
  - topological group $G$, equipped with a norm $\|\cdot\|$, and a $G$-module $F$
  - **edge potential** $g : E \to G$ satisfying $g_{ij} = g_{ji}^{-1}$, $\forall (i, j) \in E$

# Synchronization Problems

- Data:
  - graph $\Gamma = (V, E)$
  - topological group $G$, equipped with a norm $\|\cdot\|$, and a $G$-module $F$
  - **edge potential** $g : E \to G$ satisfying $g_{ij} = g_{ji}^{-1}, \ \forall (i,j) \in E$
- Goal:
  - find a **vertex potential** $f : V \to G$ or $F$ such that

$$f_i = g_{ij} f_j, \quad \forall (i,j) \in E$$

- The goal can be achieved if and only if $g_{ij} = f_i f_j^{-1}$
- **Not** always feasible!

# Synchronization Problems

- Data:
  - graph $\Gamma = (V, E)$
  - topological group $G$, equipped with a norm $\|\cdot\|$, and a $G$-module $F$
  - **edge potential** $g : E \to G$ satisfying $g_{ij} = g_{ji}^{-1}, \ \forall (i,j) \in E$
- Goal:
  - find a **vertex potential** $f : V \to G$ or $F$ such that

$$f_i = g_{ij} f_j, \quad \forall (i,j) \in E$$

- The goal can be achieved if and only if $g_{ij} = f_i f_j^{-1}$
- **Not** always feasible!
- If infeasible, find the "closest solution" in the sense of

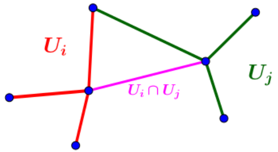
$$\min_{\substack{f:V\to G\\ \|f\|\neq 0}} \frac{1}{2} \frac{\sum\limits_{i,j\in V} \|f_i - g_{ij} f_j\|^2}{\sum\limits_{i\in V} \|f_i\|^2} (=: \eta(f))$$

# Geometric Picture

- Data:
  - graph $\Gamma = (V, E)$
  - topological group $G$, equipped with a norm $\|\cdot\|$, and a $G$-module $F$
  - **edge potential** $g : E \to G$ satisfying $g_{ij} = g_{ji}^{-1}$, $\forall (i, j) \in E$

# Geometric Picture

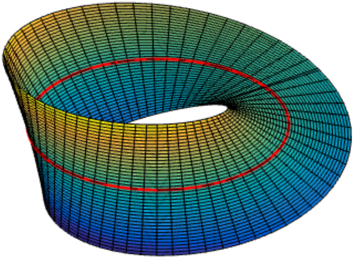
- Data:
  - graph $\Gamma = (V, E)$
  - topological group $G$, equipped with a norm $\|\cdot\|$, and a $G$-module $F$
  - **edge potential** $g : E \to G$ satisfying $g_{ij} = g_{ji}^{-1}, \ \forall (i,j) \in E$
- Flat Principal $G$-Bundle:
  - Let $\mathfrak{U} = \{U_i \mid 1 \le i \le |V|\}$ be an open cover of $\Gamma$ (viewed as a 1-dimensional simplicial complex), where $U_i$ is the *(open) star neighborhood* of vertex $i$.

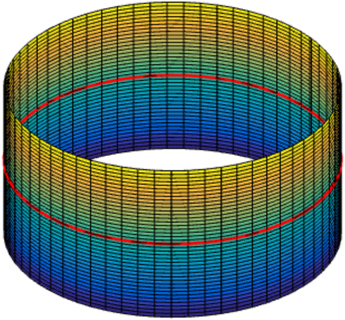


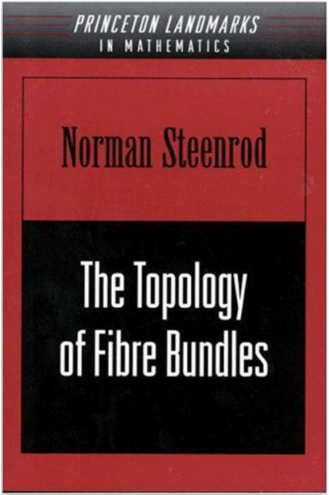

  - Triplet $(g, G, \Gamma)$ defines a *flat principal $G$-bundle $\mathscr{B}_\rho$* over $\Gamma$

Fibre Bundle: $\mathscr{E} = (E, M, F, \pi)$ is called a $F$-bundle over $M$ if

- $M$: *base* manifold
- $F$: *fibre* manifold
- $E$: *total* manifold
- $\pi : E \to M$: smooth surjective map (*bundle projection*)
- *local triviality*: for "small" open set $U \subset M$, $\pi^{-1}(U)$ is diffeomorphic to $U \times F$



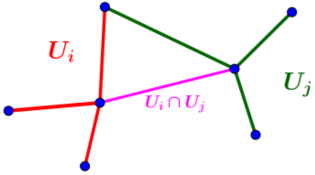





**Theorem** (Steenrod 1951, §2).
If Lie group $G$ acts on $Y$,
$\mathfrak{U} = \{U_i\}$ is an open cover of $X$,
$\{g_{ij} \in G \mid U_i \cap U_j \neq \emptyset\}$ satisfies

$$\begin{aligned} g_{ii} &= e \in G \quad \text{for all } U_i \\ g_{ij} &= g_{ji}^{-1} \quad \text{if } U_i \cap U_j \neq \emptyset \\ g_{ij} g_{jk} &= g_{ik} \quad \text{if } U_i \cap U_j \cap U_k \neq \emptyset \end{aligned}$$

then there exists a fibre bundle $\mathscr{B}$ with base space $X$, fibre $Y$, group $G$, and bundle transformations $\{g_{ij}\}$.




**No triple intersections!**

**Theorem** (Steenrod 1951, §2).
If Lie group $G$ acts on $Y$,
$\mathfrak{U} = \{U_i\}$ is an open cover of $X$,
$\{g_{ij} \in G \mid U_i \cap U_j \neq \emptyset\}$ satisfies

$$g_{ii} = e \in G \quad \text{for all } U_i$$

$$g_{ij} = g_{ji}^{-1} \quad \text{if } U_i \cap U_j \neq \emptyset$$

$$\cancel{g_{ij}g_{jk} = g_{ik} \quad \text{if } U_i \cap U_j \cap U_k \neq \emptyset}$$

then there exists a fibre bundle $\mathscr{B}$ with base space $X$, fibre $Y$, group $G$, and bundle transformations $\{g_{ij}\}$.

# Representation and Associated Bundles

- If $M$ is a principal $G$-bundle over $B$, $\rho : G \to \mathrm{Aut}(F)$ is a representation of $G$ over a vector space $F$. Then $\rho$ induces an **associated** $F$-bundle over $B$:

$$M \times_\rho F := M \times F \Big/ \sim$$

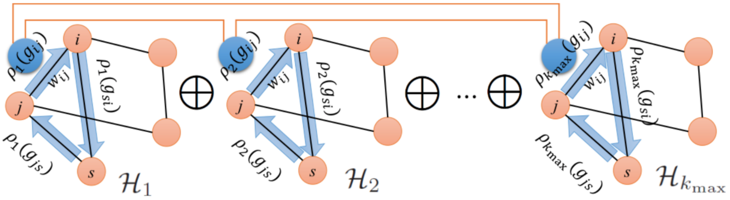
  where the equivalence relation is defined by

$$(m \cdot g, v) \sim (m, \rho(g) v)$$

- Non-equivalent irreducible representations gives rise to distinct associated bundles

# Multi-Frequency $\leftrightarrow$ Multiple Associated Bundles

- All irreducible representations of $\mathrm{U}(1)$: $\{\theta \to e^{\iota k\theta} \mid k \in \mathbb{Z}\}$
- Entrywise power: Inducing multiple irreducible representations, effectively creating many associated bundles
- Multi-frequency phase synchronization and class averaging both strive to distill features across multiple associate bundles (associated with different principal bundles)



# How Far Can We Push This idea Further?

- Faster ways than arg max Peter–Weyl?

$$\hat{g}_{ij} = \arg\max_{g \in G} \sum_{k=1}^{k_{\max}} d_k \mathrm{Tr}\left[\widehat{H}_{ij}^{(k)} \rho_k^* (g)\right]$$

- Alternative ways to leverage the algebraic consistency across irreducible representations / associated bundles?

# Bispectrum

- **Example:** Construct invariant features such as the bispectrum, using algebraic constraints such as
$$e^{\iota k_1 \theta_{ij}} \cdot e^{\iota k_2 \theta_{ij}} = e^{\iota (k_1 + k_2) \theta_{ij}}$$

# Bispectrum

- **Example:** Construct invariant features such as the bispectrum, using algebraic constraints such as
$$e^{\iota k_1 \theta_{ij}} \cdot e^{\iota k_2 \theta_{ij}} = e^{\iota (k_1 + k_2) \theta_{ij}}$$
- In phase synchronization, this amounts to comparing 1 with $\left| W_{ij}^{(k_1)} W_{ij}^{(k_2)} \overline{W_{ij}^{(k_1+k_2)}} \right|$ for each pair $(i, j)$

# Bispectrum

- **Example:** Construct invariant features such as the bispectrum, using algebraic constraints such as

$$e^{\iota k_1 \theta_{ij}} \cdot e^{\iota k_2 \theta_{ij}} = e^{\iota (k_1+k_2)\theta_{ij}}$$

- In phase synchronization, this amounts to comparing 1 with $\left| W_{ij}^{(k_1)} W_{ij}^{(k_2)} \overline{W_{ij}^{(k_1+k_2)}} \right|$ for each pair $(i,j)$
- In general, with the help of Clebsch–Gordan coefficients $C_{k_1,k_2}$, compare the following with the identity matrix $I_{d_1 d_2}$:

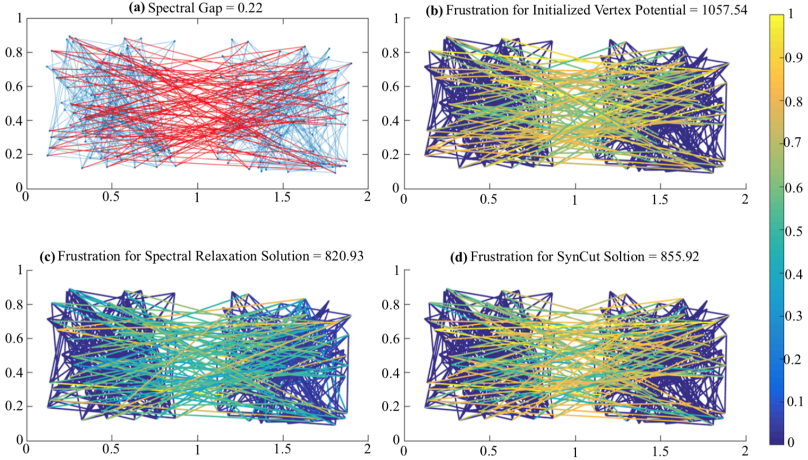
$$\left[ F_{ij}^{(k_1)} \bigotimes F_{ij}^{(k_2)} \right] C_{k_1,k_2} \left[ \bigoplus_{k \in k_1 \bigotimes k_2} F_{ij}^{(k)} \right] C_{k_1,k_2}^*$$

where $F_{ij}^{(k_1)}$, $F_{ij}^{(k_2)}$, $F_{ij}^{(k)}$ estimate $\rho_{k_1}(g_{ij})$, $\rho_{k_2}(g_{ij})$, $\rho_k(g_{ij})$

# Bispectrum

- **Example:** Construct invariant features such as the bispectrum, using algebraic constraints such as
$$e^{\iota k_1 \theta_{ij}} \cdot e^{\iota k_2 \theta_{ij}} = e^{\iota (k_1+k_2) \theta_{ij}}$$
- In phase synchronization, this amounts to comparing 1 with $\left| W_{ij}^{(k_1)} W_{ij}^{(k_2)} \overline{W_{ij}^{(k_1+k_2)}} \right|$ for each pair $(i, j)$
- In general, with the help of Clebsch–Gordan coefficients $C_{k_1,k_2}$, compare the following with the identity matrix $I_{d_1 d_2}$:
$$\left[ F_{ij}^{(k_1)} \bigotimes F_{ij}^{(k_2)} \right] C_{k_1,k_2} \left[ \bigoplus_{k \in k_1 \bigotimes k_2} F_{ij}^{(k)} \right] C_{k_1,k_2}^{*}$$
where $F_{ij}^{(k_1)}$, $F_{ij}^{(k_2)}$, $F_{ij}^{(k)}$ estimate $\rho_{k_1}(g_{ij})$, $\rho_{k_2}(g_{ij})$, $\rho_k(g_{ij})$
- Can use this to construct invariant moment features of arbitrary order; more details in **(Fan–G.–Zhao 2019)**

# Detour: A Variant of Community Detection



# Open Problems

- Algorithm works for $\mathrm{SO}(3)$ seamlessly, but theory?
- Possible extension to other synchronization and multireference alignment problems over compact/noncompact Lie groups?
- Fundamental statistical/computational limits in the community detection setting?
- Mult-frequency vector diffusion maps?
- **A learning paradigm on sheaves?**

# Thank You!


- AMS–Simons Travel Grant
- UChicago CDAC Data Science Discovery Seed Grant
- NSF CDS&E-MSS DMS-1854831


Center for Data and Computing


• Tingran Gao and Zhizhen Zhao, *Multi-Frequency Phase Synchronization.* Proceedings of the 36th International Conference on Machine Learning (ICML 2019), PMLR 97:2132–2141, 2019.
• Tingran Gao, Yifeng Fan, and Zhizhen Zhao. *Representation Theoretic Patterns in Multi-Frequency Class Averaging for Three-Dimensional Cryo-Electron Microscopy.* arxiv:1906.01082.
• Yifeng Fan, Tingran Gao, and Zhizhen Zhao, *Unsupervised Co-Learning on $\mathcal{G}$-Manifolds Across Irreducible Representations.* 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), arxiv:1906.02707