

Statistical estimation under group actions: the sample complexity of multi-reference alignment

Afonso S. Bandeira



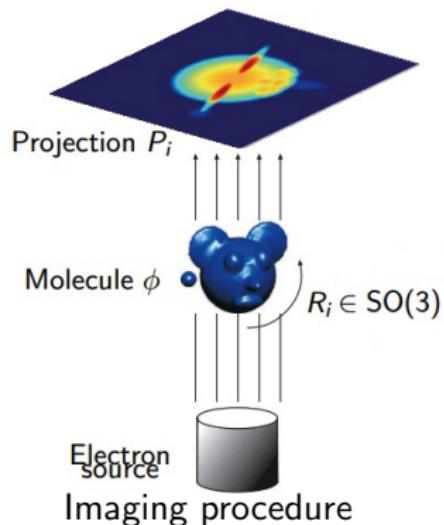
In memory of Amelia Perry

Cargese, France, August 22, 2018

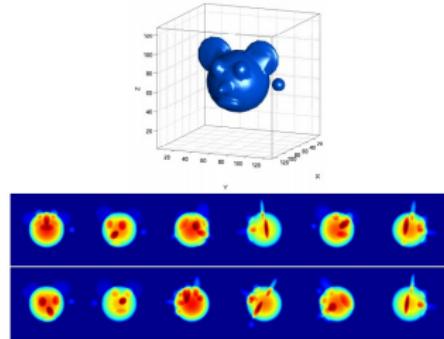
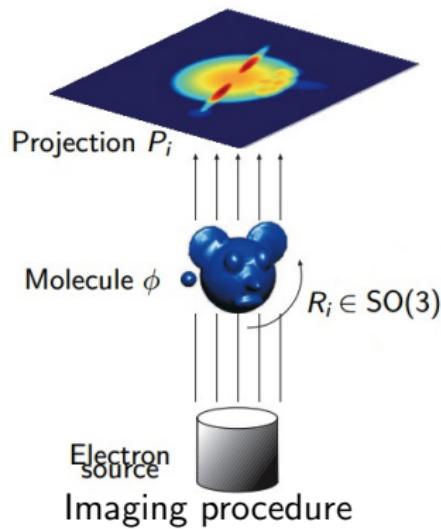
Joint work with:

- ▶ Amelia Perry (MIT)
- ▶ Jonathan Weed (MIT)
- ▶ Philippe Rigollet (MIT)
- ▶ Amit Singer (Princeton)
- ▶ Alex Wein (MIT)
- ▶ Ben Blum-Smith (NYU)
- ▶ Joe Kileel (Princeton)

Cryo-Electron Microscopy

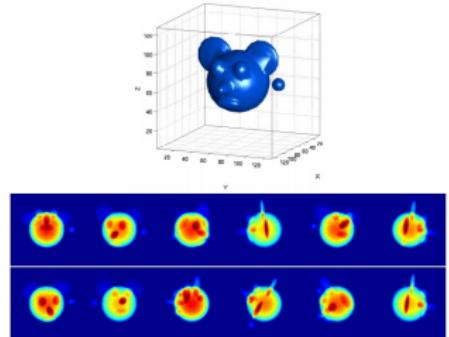
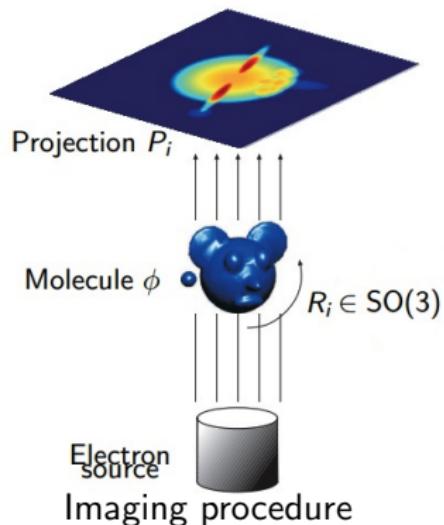


Cryo-Electron Microscopy



Task: Reconstruct the 3d molecule from **noisy** projections taken from **unknown** directions

Cryo-Electron Microscopy



Task: Reconstruct the 3d molecule from **noisy** projections taken from **unknown** directions



Images courtesy of Amit Singer, Yoel Shkolnisky, and Fred Sigworth

Cryo-Electron Microscopy

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video |

Archive > Volume 525 > Issue 7968 > News Feature > Article

NATURE | NEWS FEATURE

The revolution will not be crystallized: a new method sweeps through structural biology

Move over X-ray crystallography. Cryo-electron microscopy is kicking up a storm by revealing the hidden machinery of the cell.

Ewen Callaway

09 September 2015

[PDF](#) [Rights & Permissions](#)



In a basement room, deep in the bowels of a steel-clad building in Cambridge, a major insurgency is under way.

nature methods Techniques for life scientists and chemists

Home | Current Issue | Comment | Research | Archive | Authors & referees | About the journal |

Home > current issue > editorial > full text

NATURE METHODS | EDITORIAL

< >

Method of the Year 2015

Nature Methods 13, 1 (2016) | doi:10.1038/nmeth.3730

Published online 30 December 2015

[PDF](#) [Citation](#) [Reprints](#) [Rights & permissions](#) [Article metrics](#)

The end of ‘Meth-ology’: single-particle cryo-electron microscopy (cryo-EM) is now being used to solve macromolecular structures at high resolution.

The three-dimensional structure of a protein or protein complex provides crucial insights into its biological function. As a structure-determination technique, cryo-EM has played second fiddle to X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. This is rapidly changing. However, thanks to recent technical advances that allow near-atomic-resolution structures to be solved using cryo-EM, the time is right to celebrate single-particle cryo-EM as our Method of the Year.

For decades, X-ray crystallography has been the go-to approach for solving protein structures. However, many proteins—especially membrane proteins and protein complexes—are stubbornly resistant to crystallization. In contrast, cryo-EM is less limited by these technical difficulties, but it faces different limitations. For example, the technique of serial femtosecond crystallography, carried out using an X-ray free-electron laser (XFEL), requires a slew of easier-to-produce microcrystals, rather than single large protein crystals, but the competition for beamtime at a highly specialized XFEL is fierce. NMR spectroscopy is useful for solving the structures of small proteins, but it remains quite difficult to apply it to larger ones.

In contrast to crystallography, cryo-EM is particularly well suited for obtaining structural information for large protein complexes and for systems that exhibit multiple conformational or compositional states. Researchers in this initially small field have made steady advances to improve the resolution and, by extension, the biological applicability of cryo-EM over the past few decades. A brief history of the key milestones in cryo-EM is given in a Historical Commentary by Eva Nogales on page 24.

Cryo-Electron Microscopy

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video |

Archive > Volume 525 > Issue 7968 > News Feature > Article

NATURE | NEWS FEATURE

The revolution will not be crystallized: a new method sweeps through structural biology

Move over X-ray crystallography. Cryo-electron microscopy is kicking up a storm by revealing the hidden machinery of the cell.

Ewen Callaway

09 September 2015

[PDF](#) [Rights & Permissions](#)



Illustration by Victor Hoven

In a basement room, deep in the bowels of a steel-clad building in Cambridge, a major insurgency is under way.

nature methods 

Techniques for life scientists and chemists

Home | Current issue | Comment | Research | Archive | Authors & referees | About the journal |

Home | Current issue | Editorial | Full text

NATURE METHODS | EDITORIAL

< >

Method of the Year 2015

Nature Methods 13, 1 (2016) | doi:10.1038/nmeth.3730

Published online 30 December 2015

[PDF](#) [Citation](#) [Reprints](#) [Rights & permissions](#) [Article metrics](#)

The end of 'Meth-ology': single-particle cryo-electron microscopy (cryo-EM) is now being used to solve macromolecular structures at high resolution.

The three-dimensional structure of a protein or protein complex provides crucial insights into its biological function. As a structure-determination technique, cryo-EM has played second fiddle to X-ray crystallography for many years, especially in the field of nuclear magnetic resonance (NMR) spectroscopy. This is rapidly changing. However, thanks to recent technical advances that allow near-atomic-resolution structures to be solved using cryo-EM, the time is right to celebrate single-particle cryo-EM as our Method of the Year.

For decades, X-ray crystallography has been the go-to approach for solving protein structures. However, many proteins—especially membrane proteins and protein complexes—are stubbornly resistant to crystallization. In contrast, cryo-EM is well suited to solving structures of proteins with different limitations. For example, the technique of serial femtosecond crystallography, carried out using an X-ray free-electron laser (XFEL), requires a slew of easier-to-produce microcrystals rather than single large protein crystals, but the competition for beamtime at a highly specialized XFEL is fierce. NMR spectroscopy is useful for solving the structures of small proteins, but it remains quite difficult to apply it to larger ones.

In contrast to crystallography, cryo-EM is particularly well suited for obtaining structural information for large protein complexes and for systems that exhibit multiple conformational or compositional states. Researchers in this initially small field have made steady advances to improve the resolution and, by extension, the biological applicability of cryo-EM over the past few decades. A brief history of the key milestones in cryo-EM is given in a Historical Commentary by Eva Nogales on page 24.



2017 Chemistry Laureates. III: N. Elmehed.
© Nobel Media 2017

2017 Nobel Prize in Chemistry

The **Nobel Prize in Chemistry 2017** was awarded to **Jacques Dubochet**, **Joachim Frank** and **Richard Henderson** "for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution".

Cryo-Electron Microscopy

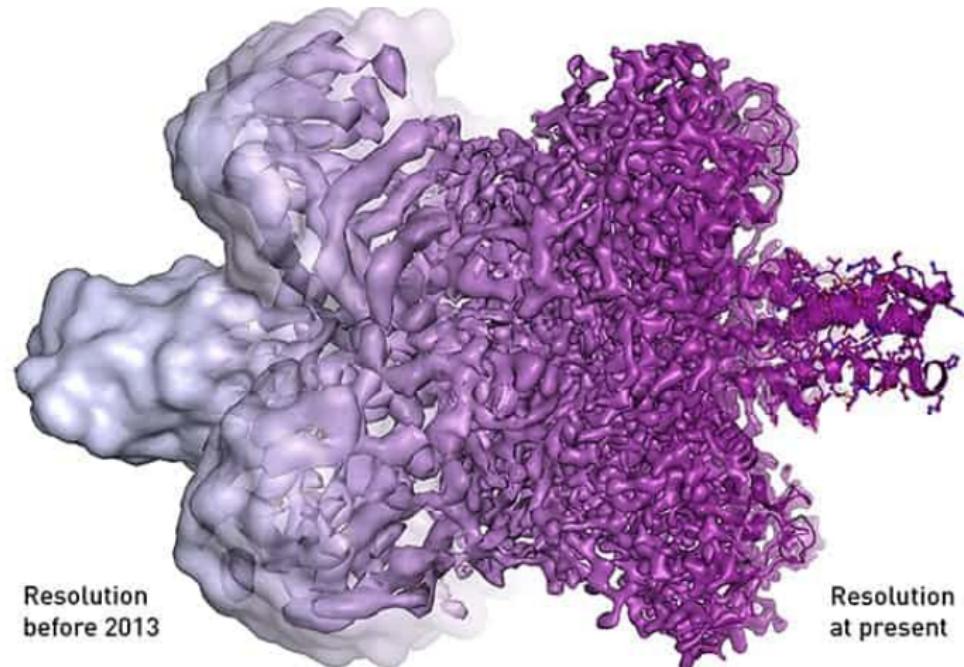


Illustration: ©Martin Hägglom/The Royal Swedish Academy of Sciences

Multi-reference Alignment

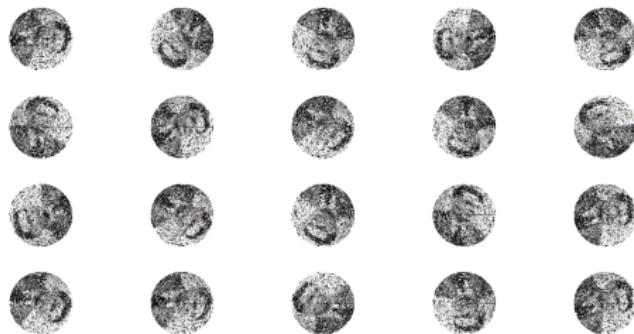
- ▶ G compact group, $G \curvearrowright \mathbb{R}^d$ ($G \leqslant O(d)$)
- ▶ Signal/Parameter of interest: $\theta \in \mathbb{R}^d$ (say $\|\theta\|=1$)
- ▶ Measurements: $Y_i = g_i \circ \theta + \sigma \xi_i, \quad i = 1, \dots, n$
 - $g_i \sim \text{Haar}(G)$ i.i.d.
 - $\xi_i \sim \mathcal{N}(0, I_d)$ i.i.d.
 - $\sigma \geq 0$ is the noise level

Multi-reference Alignment

- ▶ G compact group, $G \curvearrowright \mathbb{R}^d$ ($G \leqslant O(d)$)
- ▶ Signal/Parameter of interest: $\theta \in \mathbb{R}^d$ (say $\|\theta\|=1$)
- ▶ Measurements: $Y_i = \mathcal{P}(g_i \circ \theta + \sigma \xi_i), \quad i = 1, \dots, n$
 $g_i \sim \text{Haar}(G)$ i.i.d.
 $\xi_i \sim \mathcal{N}(0, I_d)$ i.i.d.
 $\sigma \geq 0$ is the noise level
 \mathcal{P} is a projection

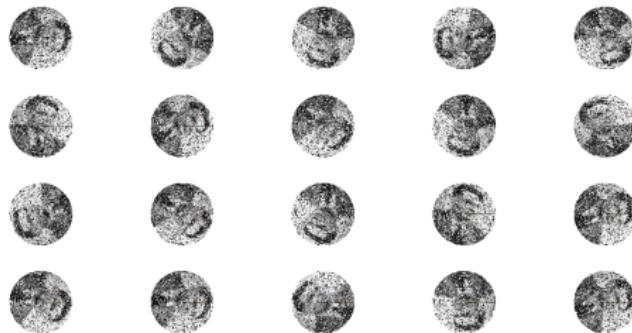
Other Applications

- ▶ Image Registration

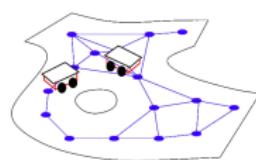


Other Applications

- ▶ Image Registration



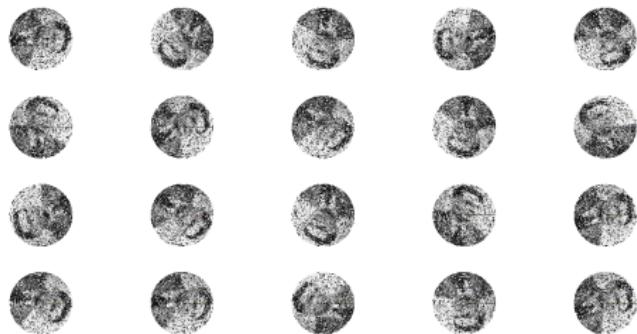
- ▶ Simultaneous Localization And Mapping



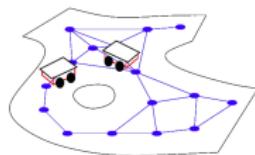
D. M. Rosen, L. Carlone, A. S. Bandeira, J. J. Leonard,
WAFR 2016

Other Applications

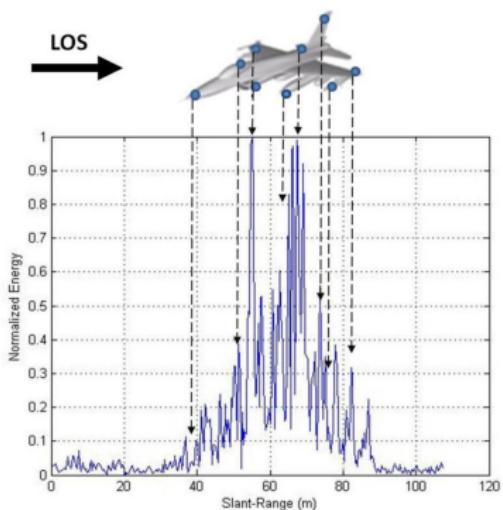
- ▶ Image Registration



- ▶ Simultaneous Localization And Mapping



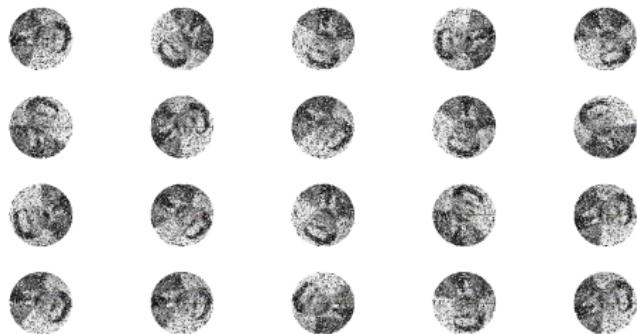
- ▶ Radar



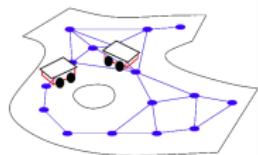
López-Rodríguez et al., 2013.

Other Applications

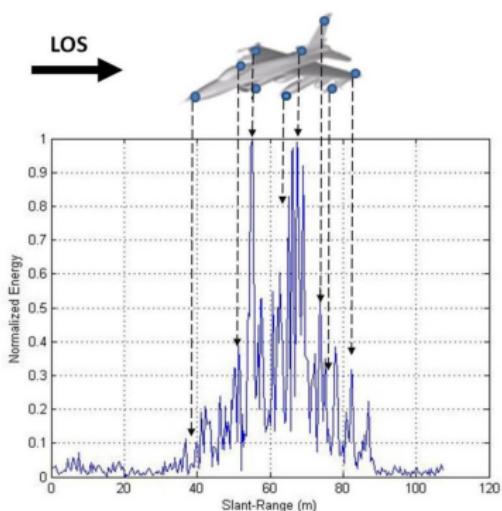
- ▶ Image Registration



- ▶ Simultaneous Localization And Mapping



- ▶ Radar



López-Rodríguez et al., 2013.

- ▶ And many others...

Cyclic Shifts

- ▶ Signal/Parameter of interest: $\theta \in \mathbb{R}^d$ (say $\|\theta\|=1$)
- ▶ Measurements: $Y_i = R_{\ell_i}\theta + \sigma\xi_i$

R_{ℓ_i} is a cyclic shift by ℓ_i coordinates
 $\ell_i \sim \text{Unif}[d]$ i.i.d. (harmless assumption)

$\xi_i \sim \mathcal{N}(0, I_d)$ i.i.d.

$\sigma \geq 0$ is the noise level

$$R_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$

Cyclic Shifts

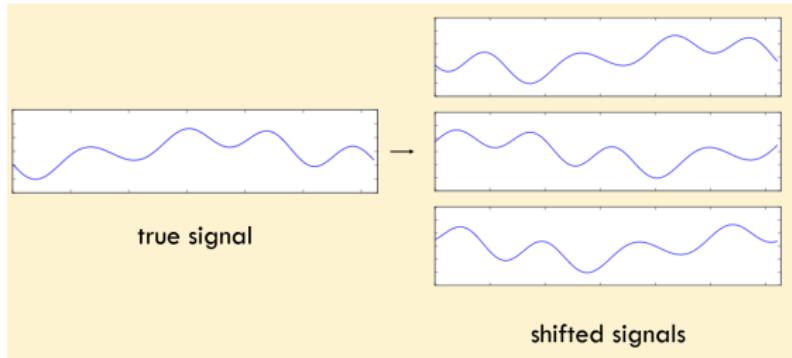
- ▶ Signal/Parameter of interest: $\theta \in \mathbb{R}^d$ (say $\|\theta\|=1$)
- ▶ Measurements: $Y_i = R_{\ell_i}\theta + \sigma\xi_i$

R_{ℓ_i} is a cyclic shift by ℓ_i coordinates
 $\ell_i \sim \text{Unif}[d]$ i.i.d. (harmless assumption)

$\xi_i \sim \mathcal{N}(0, I_d)$ i.i.d.

$\sigma \geq 0$ is the noise level

$$R_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$



A. S. Bandeira, M. Charikar, A. Singer, A. Zhu, ITCS 2014.

E. Abbe, J. Pereira, A. Singer, 2017.

Cyclic Shifts

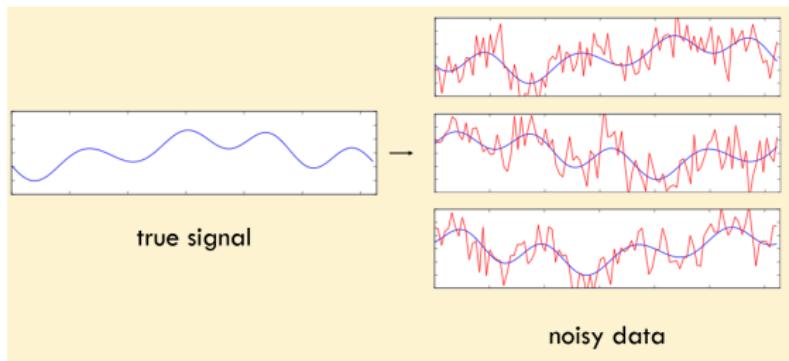
- ▶ Signal/Parameter of interest: $\theta \in \mathbb{R}^d$ (say $\|\theta\|=1$)
- ▶ Measurements: $Y_i = R_{\ell_i}\theta + \sigma\xi_i$

R_{ℓ_i} is a cyclic shift by ℓ_i coordinates
 $\ell_i \sim \text{Unif}[d]$ i.i.d. (harmless assumption)

$\xi_i \sim \mathcal{N}(0, I_d)$ i.i.d.

$\sigma \geq 0$ is the noise level

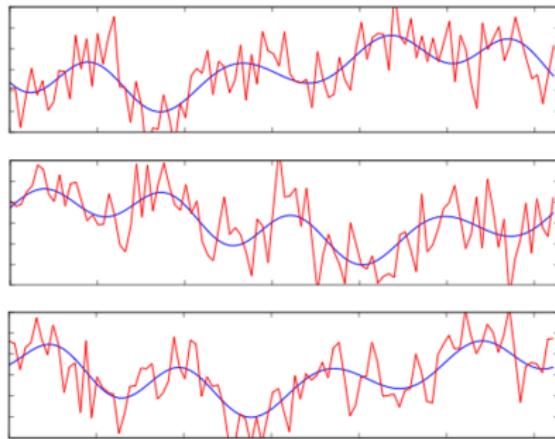
$$R_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$



A. S. Bandeira, M. Charikar, A. Singer, A. Zhu, ITCS 2014.

E. Abbe, J. Pereira, A. Singer, 2017.

Estimating the shifts

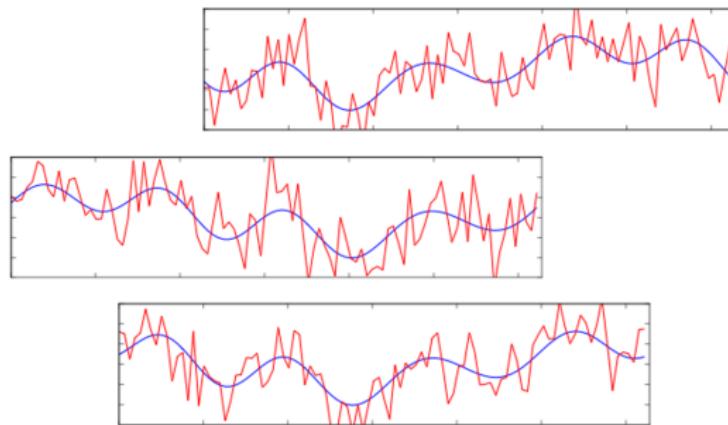


A. S. Bandeira, M. Charikar, A. Singer, A. Zhu, ITCS 2014.

A. S. Bandeira, Y. Chen, A. Singer, 2015.

Images courtesy of J. Weed

Estimating the shifts

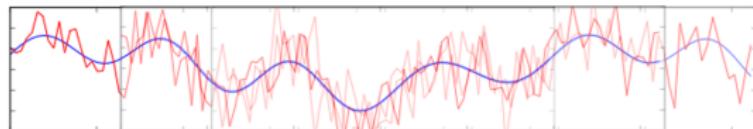


A. S. Bandeira, M. Charikar, A. Singer, A. Zhu, ITCS 2014.

A. S. Bandeira, Y. Chen, A. Singer, 2015.

Images courtesy of J. Weed

Estimating the shifts

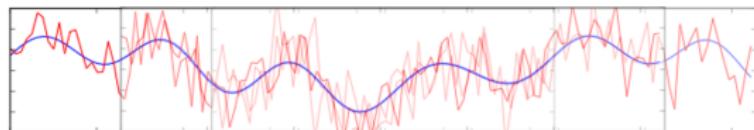


A. S. Bandeira, M. Charikar, A. Singer, A. Zhu, ITCS 2014.

A. S. Bandeira, Y. Chen, A. Singer, 2015.

Images courtesy of J. Weed

Estimating the shifts



If we can estimate shifts well then:

$$\tilde{\theta} = \frac{1}{n} \sum_{i=1}^n R_{-\ell_i^*} Y_i \quad \text{satisfies} \quad \mathbb{E} \min_{\ell \in \mathbb{Z}_d} \|\tilde{\theta} - R_\ell \theta\| \lesssim_d \frac{\sigma}{\sqrt{n}},$$

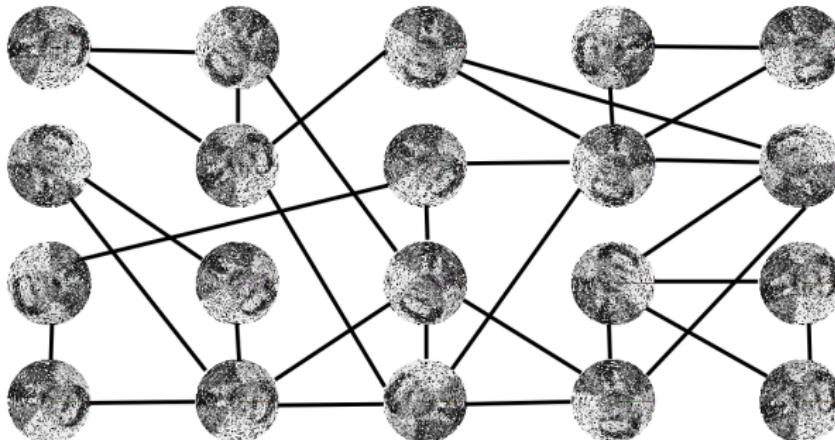
$$\text{Orbit distance: } \rho(\tilde{\theta}, \theta) = \min_{g \in G} \|\tilde{\theta} - g \circ \theta\|$$

A. S. Bandeira, M. Charikar, A. Singer, A. Zhu, ITCS 2014.

A. S. Bandeira, Y. Chen, A. Singer, 2015.

Images courtesy of J. Weed

The Synchronization Approach

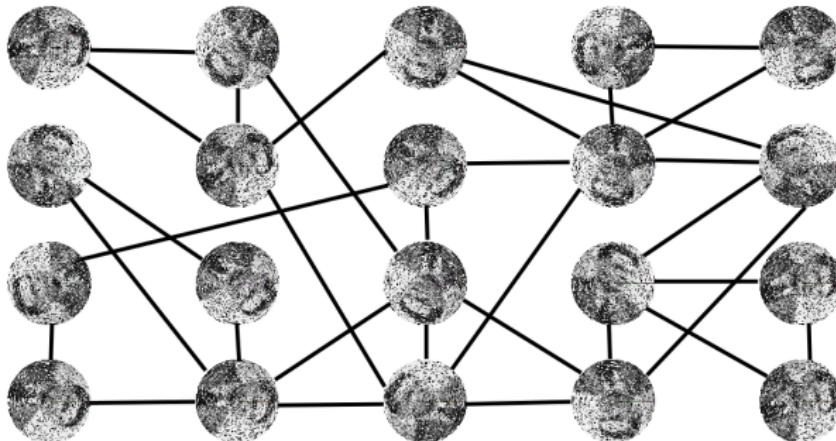


A. S. Bandeira, M. Charikar, A. Singer, A. Zhu, ITCS 2014.

A. S. Bandeira, Y. Chen, A. Singer, 2015.

C. Aguerrebere, M. Delbracio, A. Bartesaghi, G. Sapiro. IEEE ToSP 2016.

The Synchronization Approach



Potential Pitfall: Below a certain SNR threshold it is impossible to reliably estimate the shifts

A. S. Bandeira, M. Charikar, A. Singer, A. Zhu, ITCS 2014.

A. S. Bandeira, Y. Chen, A. Singer, 2015.

C. Aguerrebere, M. Delbracio, A. Bartesaghi, G. Sapiro. IEEE ToSP 2016.

Mixture of Gaussians Interpretation

Gaussian mixture model:

WLOG: $\ell_i \sim \text{Unif}[d]$

$$Y_i \sim P_\theta = \frac{1}{d} \sum_{R \in \mathbb{Z}_d} \mathcal{N}(R\theta, \sigma^2 I_d)$$

Centers correspond to point on orbit – algebraic constraints

Mixture of Gaussians Interpretation

Gaussian mixture model:

WLOG: $\ell_i \sim \text{Unif}[d]$

$$Y_i \sim P_\theta = \frac{1}{d} \sum_{R \in \mathbb{Z}_d} \mathcal{N}(R\theta, \sigma^2 I_d)$$

Centers correspond to point on orbit – algebraic constraints

- ▶ **Curse of Dimensionality:** Mixture of Gaussian with d centers:

$$\min_{\tilde{\theta}(Y)} \max_{\theta} \mathbb{E} \rho(\tilde{\theta}, \theta) \asymp C(\sigma, d) n^{-\frac{1}{2d}}$$

Mixture of Gaussians Interpretation

Gaussian mixture model:

WLOG: $\ell_i \sim \text{Unif}[d]$

$$Y_i \sim P_\theta = \frac{1}{d} \sum_{R \in \mathbb{Z}_d} \mathcal{N}(R\theta, \sigma^2 I_d)$$

Centers correspond to point on orbit – algebraic constraints

- ▶ **Curse of Dimensionality:** Mixture of Gaussian with d centers:

$$\min_{\tilde{\theta}(Y)} \max_{\theta} \mathbb{E} \rho(\tilde{\theta}, \theta) \asymp C(\sigma, d) n^{-\frac{1}{2d}}$$

- ▶ **Parametric Rate:** Assume $\forall j, |\hat{\theta}_j| = 0$ or $|\hat{\theta}_j| > c$, then:

$$\min_{\tilde{\theta}(Y)} \max_{\theta \in \mathcal{S}} \mathbb{E} \rho(\tilde{\theta}, \theta) \asymp C(\sigma, d) n^{-\frac{1}{2}}$$

Mixture of Gaussians Interpretation

Gaussian mixture model:

WLOG: $\ell_i \sim \text{Unif}[d]$

$$Y_i \sim P_\theta = \frac{1}{d} \sum_{R \in \mathbb{Z}_d} \mathcal{N}(R\theta, \sigma^2 I_d)$$

Centers correspond to point on orbit – algebraic constraints

- ▶ **Curse of Dimensionality:** Mixture of Gaussian with d centers:

$$\min_{\tilde{\theta}(Y)} \max_{\theta} \mathbb{E} \rho(\tilde{\theta}, \theta) \asymp C(\sigma, d) n^{-\frac{1}{2d}}$$

- ▶ **Parametric Rate:** Assume $\forall j, |\hat{\theta}_j| = 0$ or $|\hat{\theta}_j| > c$, then:

$$\min_{\tilde{\theta}(Y)} \max_{\theta \in \mathcal{S}} \mathbb{E} \rho(\tilde{\theta}, \theta) \asymp C(\sigma, d) n^{-\frac{1}{2}}$$

What is the dependence on σ ?

Method of Invariants

Invariant features of the signal / moments of the corresponding gaussian mixture
are easy to learn/estimate

$$Y_i \sim P_\theta = \frac{1}{d} \sum_{R \in \mathbb{Z}_d} \mathcal{N}(R\theta, \sigma^2 I_d)$$

Method of Invariants

Invariant features of the signal / moments of the corresponding gaussian mixture
are easy to learn/estimate

$$Y_i \sim P_\theta = \frac{1}{d} \sum_{R \in \mathbb{Z}_d} \mathcal{N}(R\theta, \sigma^2 I_d)$$

$$T_1 = \mathbb{E}_R(R\theta) = (\mathbf{1}^T \theta) \mathbf{1} \quad (\text{Mean})$$

Method of Invariants

Invariant features of the signal / moments of the corresponding gaussian mixture
are easy to learn/estimate

$$Y_i \sim P_\theta = \frac{1}{d} \sum_{R \in \mathbb{Z}_d} \mathcal{N}(R\theta, \sigma^2 I_d) \quad \hat{\theta} = \mathcal{F}(\theta)$$

$$T_1 = \mathbb{E}_R(R\theta) = (\mathbf{1}^T \theta) \mathbf{1} \quad (\text{Mean}) \quad \rightarrow \quad \hat{\theta}_0 \quad (\text{DC Component})$$

Method of Invariants

Invariant features of the signal / moments of the corresponding gaussian mixture
are easy to learn/estimate

$$Y_i \sim P_\theta = \frac{1}{d} \sum_{R \in \mathbb{Z}_d} \mathcal{N}(R\theta, \sigma^2 I_d) \quad \hat{\theta} = \mathcal{F}(\theta)$$

$$T_1 = \mathbb{E}_R(R\theta) = (\mathbf{1}^T \theta) \mathbf{1} \quad (\text{Mean}) \quad \rightarrow \quad \hat{\theta}_0 \quad (\text{DC Component})$$

$$T_2 = \mathbb{E}_R(R\theta)(R\theta)^T \quad (\text{Auto-Correlation}) \quad \rightarrow \quad |\hat{\theta}_k|^2 \quad (\text{Power Spectrum})$$

Method of Invariants

Invariant features of the signal / moments of the corresponding gaussian mixture
are easy to learn/estimate

$$Y_i \sim P_\theta = \frac{1}{d} \sum_{R \in \mathbb{Z}_d} \mathcal{N}(R\theta, \sigma^2 I_d) \quad \hat{\theta} = \mathcal{F}(\theta)$$

$$T_1 = \mathbb{E}_R(R\theta) = (\mathbf{1}^T \theta) \mathbf{1} \quad (\text{Mean}) \quad \rightarrow \hat{\theta}_0 \quad (\text{DC Component})$$

$$T_2 = \mathbb{E}_R(R\theta)(R\theta)^T \quad (\text{Auto-Correlation}) \quad \rightarrow |\hat{\theta}_k|^2 \quad (\text{Power Spectrum})$$

$$T_3 = \mathbb{E}_R(R\theta)^{\otimes 3} \quad \rightarrow \hat{\theta}_{k_1} \hat{\theta}_{k_2} \hat{\theta}_{-k_1 - k_2} \quad (\text{Bispectrum})$$

...

→ ...

Inverting the bispectrum

If $\hat{\theta}_k \neq 0 \forall_k$ then θ can be reconstructed from $\theta_{k_1} \theta_{k_2} \theta_{-k_1-k_2}$

Assume $\hat{\theta}_1 > 0$

$$\hat{\theta}_1 = \sqrt{|\hat{\theta}_1|^2}, \quad \hat{\theta}_2 = \frac{\hat{\theta}_2 \hat{\theta}_1^2}{\hat{\theta}_1^2}, \quad \hat{\theta}_3 = \frac{\hat{\theta}_3 \hat{\theta}_{-2} \hat{\theta}_{-1}}{\hat{\theta}_{-2} \hat{\theta}_{-1}}, \quad \dots$$

J. W. Tukey, 1953.

G. B. Giannakis, 1989.

D. R. Brillinger, 1991.

B. M. Sadler and G. B. Giannakis, 1992.

T. Bendory, N. Boumal, C. Ma, Z. Zhao, A. Singer, 2017.

Inverting the bispectrum

If $\hat{\theta}_k \neq 0 \forall_k$ then θ can be reconstructed from $\theta_{k_1} \theta_{k_2} \theta_{-k_1-k_2}$

Assume $\hat{\theta}_1 > 0$

$$\hat{\theta}_1 = \sqrt{|\hat{\theta}_1|^2}, \quad \hat{\theta}_2 = \frac{\hat{\theta}_2 \hat{\theta}_1^2}{\hat{\theta}_1^2}, \quad \hat{\theta}_3 = \frac{\hat{\theta}_3 \hat{\theta}_{-2} \hat{\theta}_{-1}}{\hat{\theta}_{-2} \hat{\theta}_{-1}}, \quad \dots$$

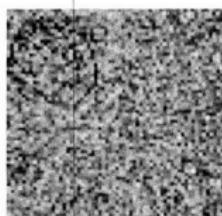


Figure 3
Object "Calvin" in white Gaussian noise.
SNR = -16dB.

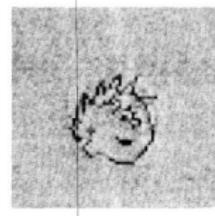


Figure 4
Reconstruction from 10 frames with random translation.

-
- J. W. Tukey, 1953.
G. B. Giannakis, 1989.
D. R. Brillinger, 1991.
B. M. Sadler and G. B. Giannakis, 1992.
T. Bendory, N. Boumal, C. Ma, Z. Zhao, A. Singer, 2017.

Inverting the bispectrum

If $\hat{\theta}_k \neq 0 \forall_k$ then θ can be reconstructed from $\theta_{k_1} \theta_{k_2} \theta_{-k_1-k_2}$

Assume $\hat{\theta}_1 > 0$

$$\hat{\theta}_1 = \sqrt{|\hat{\theta}_1|^2}, \quad \hat{\theta}_2 = \frac{\hat{\theta}_2 \hat{\theta}_1^2}{\hat{\theta}_1^2}, \quad \hat{\theta}_3 = \frac{\hat{\theta}_3 \hat{\theta}_{-2} \hat{\theta}_{-1}}{\hat{\theta}_{-2} \hat{\theta}_{-1}}, \quad \dots$$



Figure 3
Object "Calvin" in white Gaussian noise.
SNR = $\sim 10dB$.

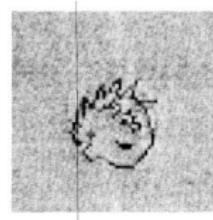


Figure 4
Reconstruction from 10 frames with random translation.

Drawback: $\text{Var}[Y_i^{\otimes 3}] = \sigma^6$

so $\frac{1}{n} \sum_{i=1}^n Y_i^{\otimes 3} \rightarrow \mathbb{E}[Y_i^{\otimes 3}]$ at rate σ^3/\sqrt{n}

$n \sim \sigma^6$ samples needed

J. W. Tukey, 1953.

G. B. Giannakis, 1989.

D. R. Brillinger, 1991.

B. M. Sadler and G. B. Giannakis, 1992.

T. Bendory, N. Boumal, C. Ma, Z. Zhao, A. Singer, 2017.

Statistical Lower Bounds and Information Geometry

$$Y_i \sim P_\theta = \frac{1}{d} \sum_{R \in \mathbb{Z}_d} \mathcal{N}(R\theta, \sigma^2 I_d)$$

S. Kullback, R. A. Leibler, AMS 1951.

M. S. Pinsker, 1964.

L. LeCam, AoS 1973.

Statistical Lower Bounds and Information Geometry

LeCam two-point argument: $Y_i \sim P_\theta = \frac{1}{d} \sum_{R \in \mathbb{Z}_d} \mathcal{N}(R\theta, \sigma^2 I_d)$ VS P_τ

S. Kullback, R. A. Leibler, AMS 1951.

M. S. Pinsker, 1964.

L. LeCam, AoS 1973.

Statistical Lower Bounds and Information Geometry

LeCam two-point argument: $Y_i \sim P_\theta = \frac{1}{d} \sum_{R \in \mathbb{Z}_d} \mathcal{N}(R\theta, \sigma^2 I_d)$ VS P_τ

- ▶ Kullback-Liebler Divergence [KL'51]

$$D_{KL}(\theta || \tau) = \int \log \left(\frac{dP_\theta}{dP_\tau}(y) \right) dP_\theta(y)$$

S. Kullback, R. A. Leibler, AMS 1951.

M. S. Pinsker, 1964.

L. LeCam, AoS 1973.

Statistical Lower Bounds and Information Geometry

LeCam two-point argument: $Y_i \sim P_\theta = \frac{1}{d} \sum_{R \in \mathbb{Z}_d} \mathcal{N}(R\theta, \sigma^2 I_d)$ VS P_τ

- ▶ Kullback-Liebler Divergence [KL'51]

$$D_{KL}(\theta || \tau) = \int \log \left(\frac{dP_\theta}{dP_\tau}(y) \right) dP_\theta(y)$$

- ▶ Invariant Moments: If $\rho(\theta, \tau) < \varepsilon$

$$D_{KL}(\theta || \tau) = \frac{1}{2\sigma^2} \|\mathbb{E}R\theta - \mathbb{E}R\tau\|^2 + \frac{1}{4\sigma^4} \|\mathbb{E}(R\theta)^{\otimes 2} - \mathbb{E}(R\tau)^{\otimes 2}\|^2 + \frac{O(\varepsilon^2)}{\sigma^6}$$

S. Kullback, R. A. Leibler, AMS 1951.

M. S. Pinsker, 1964.

L. LeCam, AoS 1973.

Statistical Lower Bounds and Information Geometry

LeCam two-point argument: $Y_i \sim P_\theta = \frac{1}{d} \sum_{R \in \mathbb{Z}_d} \mathcal{N}(R\theta, \sigma^2 I_d)$ VS P_τ

- ▶ Kullback-Liebler Divergence [KL'51]

$$D_{KL}(\theta || \tau) = \int \log \left(\frac{dP_\theta}{dP_\tau}(y) \right) dP_\theta(y)$$

- ▶ Invariant Moments: If $\rho(\theta, \tau) < \varepsilon$

$$D_{KL}(\theta || \tau) = \frac{1}{2\sigma^2} \|\mathbb{E}R\theta - \mathbb{E}R\tau\|^2 + \frac{1}{4\sigma^4} \|\mathbb{E}(R\theta)^{\otimes 2} - \mathbb{E}(R\tau)^{\otimes 2}\|^2 + \frac{O(\varepsilon^2)}{\sigma^6}$$

- ▶ Chain Rule: $D_{KL}(P_\theta^{\otimes n} || P_\tau^{\otimes n}) = n D_{KL}(P_\theta || P_\tau)$

S. Kullback, R. A. Leibler, AMS 1951.

M. S. Pinsker, 1964.

L. LeCam, AoS 1973.

Statistical Lower Bounds and Information Geometry

LeCam two-point argument: $Y_i \sim P_\theta = \frac{1}{d} \sum_{R \in \mathbb{Z}_d} \mathcal{N}(R\theta, \sigma^2 I_d)$ VS P_τ

- ▶ Kullback-Liebler Divergence [KL'51]

$$D_{KL}(\theta || \tau) = \int \log \left(\frac{dP_\theta}{dP_\tau}(y) \right) dP_\theta(y)$$

- ▶ Invariant Moments: If $\rho(\theta, \tau) < \varepsilon$

$$D_{KL}(\theta || \tau) = \frac{1}{2\sigma^2} \|\mathbb{E}R\theta - \mathbb{E}R\tau\|^2 + \frac{1}{4\sigma^4} \|\mathbb{E}(R\theta)^{\otimes 2} - \mathbb{E}(R\tau)^{\otimes 2}\|^2 + \frac{O(\varepsilon^2)}{\sigma^6}$$

- ▶ Chain Rule: $D_{KL}(P_\theta^{\otimes n} || P_\tau^{\otimes n}) = n D_{KL}(P_\theta || P_\tau)$

- ▶ Pinsker's inequality [P'64]

$$\sup\{|P(A) - Q(A)|, A \text{ event}\} =: \|P - Q\|_{TV} \leq \sqrt{\frac{1}{2} D_{KL}(P || Q)}$$

S. Kullback, R. A. Leibler, AMS 1951.

M. S. Pinsker, 1964.

L. LeCam, AoS 1973.

Statistical Lower Bounds and Information Geometry

LeCam two-point argument: $Y_i \sim P_\theta = \frac{1}{d} \sum_{R \in \mathbb{Z}_d} \mathcal{N}(R\theta, \sigma^2 I_d)$ VS P_τ

- ▶ Kullback-Liebler Divergence [KL'51]

$$D_{KL}(\theta || \tau) = \int \log \left(\frac{dP_\theta}{dP_\tau}(y) \right) dP_\theta(y)$$

- ▶ Invariant Moments: If $\rho(\theta, \tau) < \varepsilon$

$$D_{KL}(\theta || \tau) = \frac{1}{2\sigma^2} \|\mathbb{E}R\theta - \mathbb{E}R\tau\|^2 + \frac{1}{4\sigma^4} \|\mathbb{E}(R\theta)^{\otimes 2} - \mathbb{E}(R\tau)^{\otimes 2}\|^2 + \frac{O(\varepsilon^2)}{\sigma^6}$$

- ▶ Chain Rule: $D_{KL}(P_\theta^{\otimes n} || P_\tau^{\otimes n}) = n D_{KL}(P_\theta || P_\tau)$

- ▶ Pinsker's inequality [P'64]

$$\sup\{|P(A) - Q(A)|, A \text{ event}\} =: \|P - Q\|_{TV} \leq \sqrt{\frac{1}{2} D_{KL}(P || Q)}$$

In this particular case, control of $D_{KL}(\theta, \tau) \rightarrow$ guarantees for MLE

S. Kullback, R. A. Leibler, AMS 1951.

M. S. Pinsker, 1964.

L. LeCam, AoS 1973.

Main Theorem (holds for any $G \leq O(d)$)

Theorem

$$\theta, \tau \in \mathbb{R}^d, \rho(\theta, \tau) = \varepsilon. \Delta_m = \mathbb{E}_R(R\theta)^{\otimes m} - \mathbb{E}_R(R\tau)^{\otimes m}.$$

If \exists_k s.t. for $\varepsilon \rightarrow 0$

$$\|\Delta_m\| = o(\varepsilon), \quad m < k$$

$$\|\Delta_k\| = \Omega(\varepsilon)$$

then $D_{KL}(P_\theta || P_\tau) \asymp \sigma^{-2k} \varepsilon^2$

Main Theorem (holds for any $G \leq O(d)$)

Theorem

$$\theta, \tau \in \mathbb{R}^d, \rho(\theta, \tau) = \varepsilon. \Delta_m = \mathbb{E}_R(R\theta)^{\otimes m} - \mathbb{E}_R(R\tau)^{\otimes m}.$$

If \exists_k s.t. for $\varepsilon \rightarrow 0$

$$\|\Delta_m\| = o(\varepsilon), \quad m < k$$

$$\|\Delta_k\| = \Omega(\varepsilon)$$

$$\text{then } D_{KL}(P_\theta || P_\tau) \asymp \sigma^{-2k} \varepsilon^2$$

$$\text{Chain rule: } D_{KL}\left(P_\theta^{\otimes n} || P_\tau^{\otimes n}\right) = n D_{KL}(P_\theta || P_\tau)$$

For $\varepsilon \sim \sigma^k / \sqrt{n}$ no statistical test can distinguish P_θ from P_τ
(LeCam '73 two-point argument)

If $k - 1$ -th moments match, no estimator has accuracy better than
 σ^k / \sqrt{n}

Main Theorem (holds for any $G \leq O(d)$)

Theorem

$$\theta, \tau \in \mathbb{R}^d, \rho(\theta, \tau) = \varepsilon. \Delta_m = \mathbb{E}_R(R\theta)^{\otimes m} - \mathbb{E}_R(R\tau)^{\otimes m}.$$

If \exists_k s.t. for $\varepsilon \rightarrow 0$

$$\|\Delta_m\| = o(\varepsilon), \quad m < k$$

$$\|\Delta_k\| = \Omega(\varepsilon)$$

$$\text{then } D_{KL}(P_\theta || P_\tau) \asymp \sigma^{-2k} \varepsilon^2$$

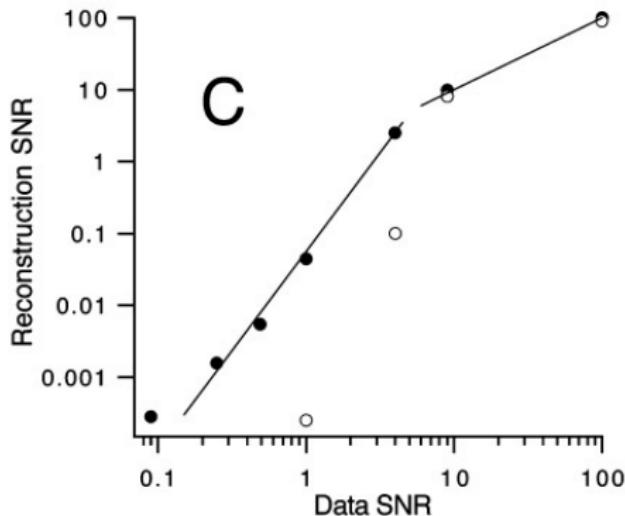
$$\text{Chain rule: } D_{KL}\left(P_\theta^{\otimes n} || P_\tau^{\otimes n}\right) = n D_{KL}(P_\theta || P_\tau)$$

For $\varepsilon \sim \sigma^k / \sqrt{n}$ no statistical test can distinguish P_θ from P_τ
(LeCam '73 two-point argument)

If $k-1$ -th moments match, no estimator has accuracy better than
 σ^k / \sqrt{n}

Power spectrum does not determine signals
⇒ σ^3 / \sqrt{n} is optimal for cyclic shifts

Behavior observed in cryo-EM reconstruction 20 years ago!



- ▶ Surprising $1/\text{SNR}^3$ (σ^6) scaling at low SNR observed in '98

Some signals are much harder

- ▶ **Worst-case:** There are signals θ that can only be estimated at rate σ^{d-2}/\sqrt{n}

Hardness depends on additive-combinatorial properties of the support of $\hat{\theta}$

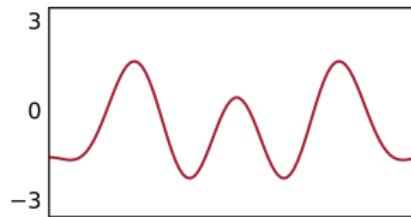
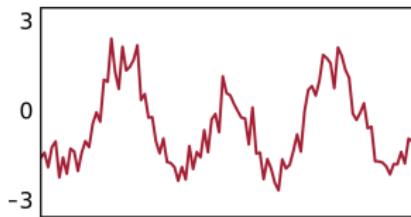
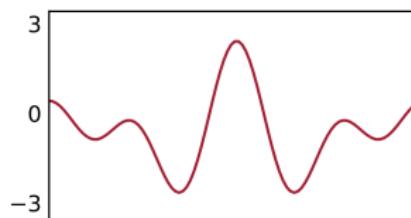
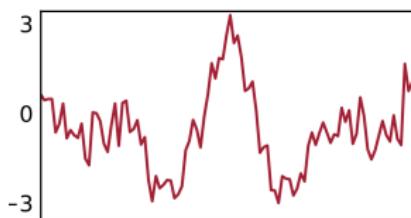
Some signals are much harder

- ▶ **Worst-case:** There are signals θ that can only be estimated at rate σ^{d-2}/\sqrt{n}

Hardness depends on additive-combinatoric properties of the support of $\hat{\theta}$

Sometimes low-passing may make the problem harder:

In following example, original signals can be recovered at $O(\sigma^6)$ samples, but low-pass ones need $\Omega(\sigma^8)$



Tensor-based Bispectrum inversion: Jennrich Algorithm

$$\mathbb{E}[Y^{\otimes 3}] \rightsquigarrow \mathbf{T}_3 = \frac{1}{d} \sum_{\ell=1}^d (R_\ell \theta)^{\otimes 3}$$

$\mathbf{T}_3 = \sum_{\ell=1}^d u_\ell^{\otimes 3}$ is a $d \times d \times d$ tensor with rank d

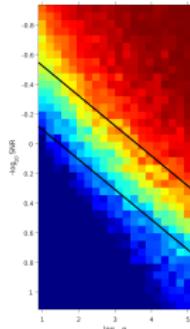
Tensor-based Bispectrum inversion: Jennrich Algorithm

$$\mathbb{E}[Y^{\otimes 3}] \rightsquigarrow \mathbf{T}_3 = \frac{1}{d} \sum_{\ell=1}^d (R_\ell \theta)^{\otimes 3}$$

$\mathbf{T}_3 = \sum_{\ell=1}^d u_\ell^{\otimes 3}$ is a $d \times d \times d$ tensor with rank d

Idea: Leverage low-rank structure

For “generic” signals θ : $\|\tilde{\theta}_{Jen} - \theta\| \sim_d \sigma^3 / \sqrt{n}$



relative error (red is good)

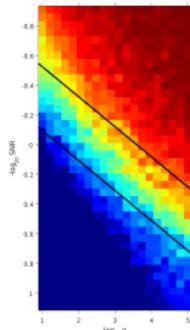
Tensor-based Bispectrum inversion: Jennrich Algorithm

$$\mathbb{E}[Y^{\otimes 3}] \rightsquigarrow \mathbf{T}_3 = \frac{1}{d} \sum_{\ell=1}^d (R_\ell \theta)^{\otimes 3}$$

$\mathbf{T}_3 = \sum_{\ell=1}^d u_\ell^{\otimes 3}$ is a $d \times d \times d$ tensor with rank d

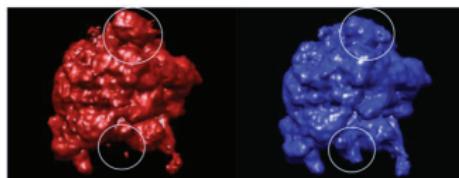
Idea: Leverage low-rank structure

For “generic” signals θ : $\|\tilde{\theta}_{Jen} - \theta\| \sim_d \sigma^3 / \sqrt{n}$



relative error (red is good)

Heterogeneity: Multiple Conformations in Cryo-EM



R. Harshman, 1970 (credits R. Jennrich).

H. Liao, J. Frank, IEEE ICBI 2010.

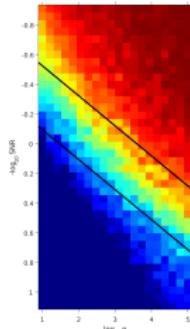
Tensor-based Bispectrum inversion: Jennrich Algorithm

$$\mathbb{E}[Y^{\otimes 3}] \rightsquigarrow \mathbf{T}_3 = \frac{1}{d} \sum_{\ell=1}^d (R_\ell \theta)^{\otimes 3}$$

$\mathbf{T}_3 = \sum_{\ell=1}^d u_\ell^{\otimes 3}$ is a $d \times d \times d$ tensor with rank d

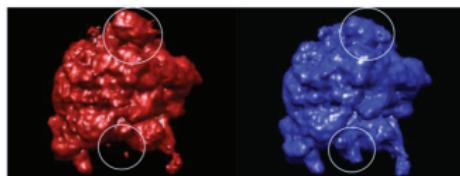
Idea: Leverage low-rank structure

For “generic” signals θ : $\|\tilde{\theta}_{Jen} - \theta\| \sim_d \sigma^3 / \sqrt{n}$



relative error (red is good)

Heterogeneity: Multiple Conformations in Cryo-EM



Samples from a **mixture** $\{\theta^{(k)}\}_{k=1}^K$

$$Y_i = R_{\ell_i} \theta^{(\pi_i)} + \sigma \xi_i$$

R. Harshman, 1970 (credits R. Jennrich).

H. Liao, J. Frank, IEEE ICBI 2010.

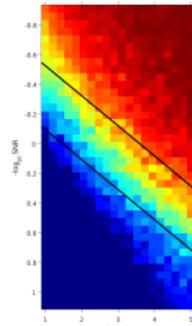
Tensor-based Bispectrum inversion: Jennrich Algorithm

$$\mathbb{E}[Y^{\otimes 3}] \rightsquigarrow \mathbf{T}_3 = \frac{1}{d} \sum_{\ell=1}^d (R_\ell \theta)^{\otimes 3}$$

$\mathbf{T}_3 = \sum_{\ell=1}^d u_\ell^{\otimes 3}$ is a $d \times d \times d$ tensor with rank d

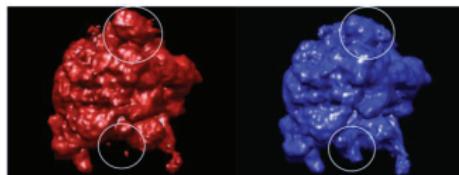
Idea: Leverage low-rank structure

For “generic” signals θ : $\|\tilde{\theta}_{Jen} - \theta\| \sim_d \sigma^3 / \sqrt{n}$



relative error (red is good)

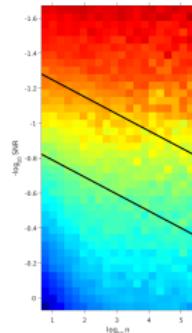
Heterogeneity: Multiple Conformations in Cryo-EM



Samples from a **mixture** $\{\theta^{(k)}\}_{k=1}^K$

$$Y_i = R_{\ell_i} \theta^{(\pi_i)} + \sigma \xi_i$$

Can be applied on \mathbf{T}_5 and gives $\|\tilde{\theta}_{Jen} - \theta\| \sim_d \sigma^5 / \sqrt{n}$

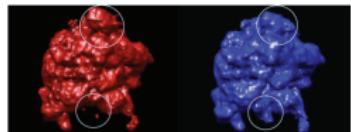


R. Harshman, 1970 (credits R. Jennrich).

H. Liao, J. Frank, IEEE ICBI 2010.

Heterogeneity: Computational-to-Statistical Gaps?

Heterogeneity: recovery from 3rd moments



Samples from a **mixture** $\{\theta^{(k)}\}_{k=1}^K$

$$Y_i = R_{\ell_i} \theta^{(\pi_i)} + \sigma \xi_i$$

Plot Courtesy of: Boumal, Bendory, Lederman, Singer, 2017

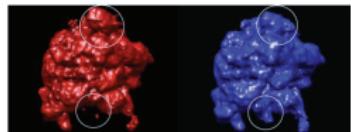
Boumal, Bendory, Lederman, Singer, 2017.

Wein, 2018.

Bandeira, Blum-Smith, Kileel, Perry, Weed, Wein, 2017.

Heterogeneity: Computational-to-Statistical Gaps?

Heterogeneity: recovery from 3rd moments



Samples from a **mixture** $\{\theta^{(k)}\}_{k=1}^K$

$$Y_i = R_{\ell_i} \theta^{(\pi_i)} + \sigma \xi_i \quad \mathbf{T}_3 = \frac{1}{d} \sum_{k=1}^K \sum_{\ell=1}^d (R_\ell \theta^{(k)})^{\otimes 3}$$

Plot Courtesy of: Boumal, Bendory, Lederman, Singer, 2017

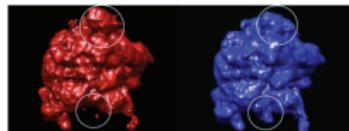
Boumal, Bendory, Lederman, Singer, 2017.

Wein, 2018.

Bandeira, Blum-Smith, Kileel, Perry, Weed, Wein, 2017.

Heterogeneity: Computational-to-Statistical Gaps?

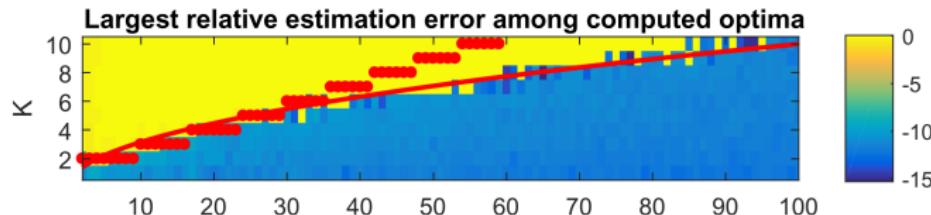
Heterogeneity: recovery from 3rd moments



Samples from a **mixture** $\{\theta^{(k)}\}_{k=1}^K$

$$Y_i = R_{\ell_i} \theta^{(\pi_i)} + \sigma \xi_i \quad \mathbf{T}_3 = \frac{1}{d} \sum_{k=1}^K \sum_{\ell=1}^d (R_{\ell} \theta^{(k)})^{\otimes 3}$$

Least squares on 3rd moments tensor:



Plot Courtesy of: Boumal, Bendory, Lederman, Singer, 2017

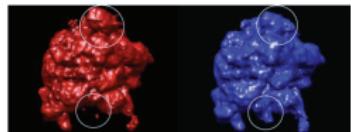
Boumal, Bendory, Lederman, Singer, 2017.

Wein, 2018.

Bandeira, Blum-Smith, Kileel, Perry, Weed, Wein, 2017.

Heterogeneity: Computational-to-Statistical Gaps?

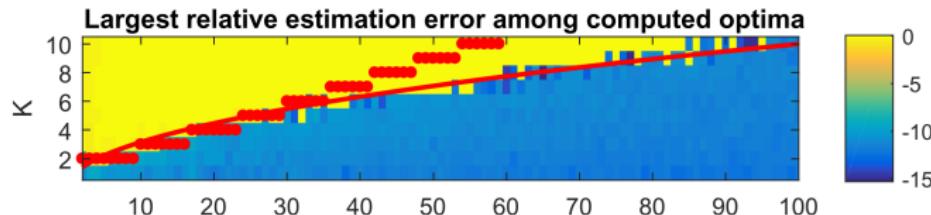
Heterogeneity: recovery from 3rd moments



Samples from a **mixture** $\{\theta^{(k)}\}_{k=1}^K$

$$Y_i = R_{\ell_i} \theta^{(\pi_i)} + \sigma \xi_i \quad \mathbf{T}_3 = \frac{1}{d} \sum_{k=1}^K \sum_{\ell=1}^d (R_{\ell} \theta^{(k)})^{\otimes 3}$$

Least squares on 3rd moments tensor:



- ▶ **Statistical-to-Computational Gap?** Preliminary results suggest computational threshold at $k \lesssim \sqrt{d}$; statistical threshold $k \lesssim d$; **related to conjectured gaps in tensor decomposition**

Plot Courtesy of: Boumal, Bendory, Lederman, Singer, 2017

Boumal, Bendory, Lederman, Singer, 2017.

Wein, 2018.

Bandeira, Blum-Smith, Kileel, Perry, Weed, Wein, 2017.

What about other groups? Invariant Theory

$$\Delta_m = \mathbb{E}_R(R\theta)^{\otimes m} - \mathbb{E}_R(R\tau)^{\otimes m}$$

$p(\theta) = \mathbb{E}_R(R\theta)^{\otimes m}$ is a degree m **invariant polynomial** in $\theta_1, \dots, \theta_d$,
$$p(\theta) = p(R\theta)$$

What about other groups? Invariant Theory

$$\Delta_m = \mathbb{E}_R(R\theta)^{\otimes m} - \mathbb{E}_R(R\tau)^{\otimes m}$$

$p(\theta) = \mathbb{E}_R(R\theta)^{\otimes m}$ is a degree m **invariant polynomial** in $\theta_1, \dots, \theta_d$,
$$p(\theta) = p(R\theta)$$

If degree m invariant polynomials generate (as an algebra) the **ring of invariant polynomials** then rate $\lesssim_d \sigma^m / \sqrt{n}$
(actually enough to generate the separating subalgebra)

For particular θ enough to generate subalgebra that is able to identify orbit of θ

What about other groups? Invariant Theory

$$\Delta_m = \mathbb{E}_R(R\theta)^{\otimes m} - \mathbb{E}_R(R\tau)^{\otimes m}$$

$p(\theta) = \mathbb{E}_R(R\theta)^{\otimes m}$ is a degree m **invariant polynomial** in $\theta_1, \dots, \theta_d$,
 $p(\theta) = p(R\theta)$

If degree m invariant polynomials generate (as an algebra) the **ring of invariant polynomials** then rate $\lesssim_d \sigma^m / \sqrt{n}$
(actually enough to generate the separating subalgebra)

For particular θ enough to generate subalgebra that is able to identify orbit of θ

Main take-home: **Statistical rates correspond to properties of the ring of invariant polynomials, an object well studied in invariant theory**

Counting Degrees-of-Freedom?

If ρ is an irreducible non-trivial representation, then $\mathbb{E}_g \rho_g \theta = 0$.

Counting Degrees-of-Freedom?

If ρ is an irreducible non-trivial representation, then $\mathbb{E}_g \rho_g \theta = 0$.

$\phi_r \theta = (R\theta)^{\otimes r}$ is a d^r -dimensional representation.

Peter-Weyl Theorem: $\phi_r = \rho_{k_1} \otimes \rho_{k_2} \otimes \cdots \otimes \rho_{k_t}$,
 ρ_k irreducible representation

Counting Degrees-of-Freedom?

If ρ is an irreducible non-trivial representation, then $\mathbb{E}_g \rho_g \theta = 0$.

$\phi_r \theta = (R\theta)^{\otimes r}$ is a d^r -dimensional representation.

Peter-Weyl Theorem: $\phi_r = \rho_{k_1} \otimes \rho_{k_2} \otimes \cdots \otimes \rho_{k_t}$,
 ρ_k irreducible representation

T_r “resolves” at most m DoF

m is number of times the trivial representation appears

Counting Degrees-of-Freedom?

If ρ is an irreducible non-trivial representation, then $\mathbb{E}_g \rho_g \theta = 0$.

$\phi_r \theta = (R\theta)^{\otimes r}$ is a d^r -dimensional representation.

Peter-Weyl Theorem: $\phi_r = \rho_{k_1} \otimes \rho_{k_2} \otimes \cdots \otimes \rho_{k_t}$,
 ρ_k irreducible representation

T_r “resolves” at most m DoF

m is number of times the trivial representation appears

Example: $SO(3) \curvearrowright \mathbb{R}^3$

$$T_1 = \rho_3 \quad 0 \text{ DoF}$$

Counting Degrees-of-Freedom?

If ρ is an irreducible non-trivial representation, then $\mathbb{E}_g \rho_g \theta = 0$.

$\phi_r \theta = (R\theta)^{\otimes r}$ is a d^r -dimensional representation.

Peter-Weyl Theorem: $\phi_r = \rho_{k_1} \otimes \rho_{k_2} \otimes \cdots \otimes \rho_{k_t}$,
 ρ_k irreducible representation

T_r “resolves” at most m DoF

m is number of times the trivial representation appears

Example: $SO(3) \curvearrowright \mathbb{R}^3$

	$T_1 = \rho_3$	0 DoF
(Clebsch-Gordon rule:)	$T_2 = \rho_1 \otimes \rho_3 \otimes \rho_5$	1 DoF

Indeed, $\text{Trace}(\mathbb{E}_\phi(\phi\theta)(\phi\theta)^T) = \|\theta\|^2$, only parameter to learn

Counting Degrees-of-Freedom?

If ρ is an irreducible non-trivial representation, then $\mathbb{E}_g \rho_g \theta = 0$.

$\phi_r \theta = (R\theta)^{\otimes r}$ is a d^r -dimensional representation.

Peter-Weyl Theorem: $\phi_r = \rho_{k_1} \otimes \rho_{k_2} \otimes \cdots \otimes \rho_{k_t}$,
 ρ_k irreducible representation

T_r “resolves” at most m DoF

m is number of times the trivial representation appears

Example: $SO(3) \curvearrowright \mathbb{R}^3$

(Clebsch-Gordon rule:)	$T_1 = \rho_3$	0 DoF
	$T_2 = \rho_1 \otimes \rho_3 \otimes \rho_5$	1 DoF

Indeed, $\text{Trace}(\mathbb{E}_\phi(\phi\theta)(\phi\theta)^T) = \|\theta\|^2$, only parameter to learn

Statistical Rate: σ^2/\sqrt{n} Sample Complexity $n \sim_d \sigma^4$

Learning most things “list decoding”

$$SO(d) \curvearrowright \left(\mathbb{R}^d\right)^k \quad \phi[\theta_1, \dots, \theta_k] = [\phi\theta_1, \dots, \phi\theta_k]$$

Learning most things “list decoding”

$$SO(d) \curvearrowright \left(\mathbb{R}^d\right)^k \quad \phi[\theta_1, \dots, \theta_k] = [\phi\theta_1, \dots, \phi\theta_k]$$

“Frequency” 2 learns Gram matrix $\sum_{i=1}^k \theta_i \theta_i^T$,
gives orbit up to Chirality.

Learning most things “list decoding”

$$SO(d) \curvearrowright (\mathbb{R}^d)^k \quad \phi[\theta_1, \dots, \theta_k] = [\phi\theta_1, \dots, \phi\theta_k]$$

“Frequency” 2 learns Gram matrix $\sum_{i=1}^k \theta_i \theta_i^T$,
gives orbit up to Chirality.

Chirality needs “Frequency” d (compute $d \times d$ determinant)

- ▶ With $\sim \sigma^4$ samples one can reduce to two orbits,
but to resolve the rest $\sim \sigma^{2d}$ samples are needed

Learning most things “list decoding”

$$SO(d) \curvearrowright (\mathbb{R}^d)^k \quad \phi[\theta_1, \dots, \theta_k] = [\phi\theta_1, \dots, \phi\theta_k]$$

“Frequency” 2 learns Gram matrix $\sum_{i=1}^k \theta_i \theta_i^T$,
gives orbit up to Chirality.

Chirality needs “Frequency” d (compute $d \times d$ determinant)

- ▶ With $\sim \sigma^4$ samples one can reduce to two orbits,
but to resolve the rest $\sim \sigma^{2d}$ samples are needed
- ▶ **“Computer certifies”:** **(Generic) List recovery is possible for Cryo-EM at $n \sim \sigma^6$, even for heterogeneous mixtures.**

Some take home messages

- ▶ Latent transformations such as shifts and rotations fundamentally change the problem
- ▶ Invariant method gives optimal dependency at high Signal-to-Noise Ratio
- ▶ Rates of estimation / Sample complexity correspond to properties of algebra of invariant polynomials
 - a well-studied object in group theory

What I didn't talk about (much)

- ▶ Heterogeneity
- ▶ Projections
- ▶ Algorithms
- ▶ Algebra
- ▶ ...

- ▶ Afonso S. Bandeira, Jonathan Weed, and Philippe Rigollet **Optimal rates of estimation for multi-reference alignment**, available on arxiv, 2017.
- ▶ Amelia Perry, Jonathan Weed, Afonso S. Bandeira, Philippe Rigollet, and Amit Singer, **The sample complexity of multi-reference alignment**, available on arxiv, 2017.
- ▶ Afonso S. Bandeira, Ben Blum-Smith, Joe Kileel, Amelia Perry, Jonathan Weed, Alex S. Wein, **Estimation under group actions: recovering orbits from invariants**, available on arxiv, 2017.

Questions?

Shameless plug: Take a look at **Ten Lectures and Forty-Two Open Problems in the Mathematics of Data Science** for some open problems

Add-on: Model bias – “*Einstein from noise*”

Naive Approach:

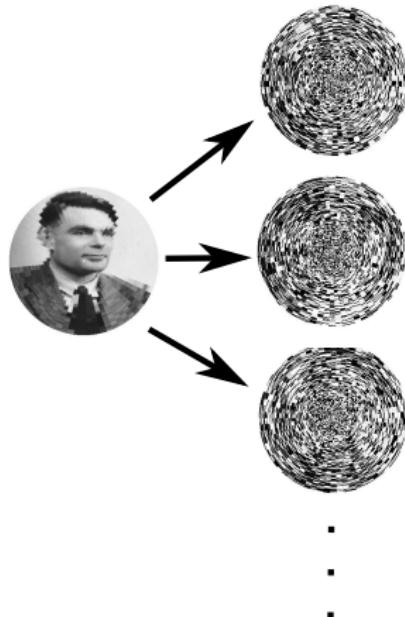
Use prior beliefs to construct a **model image** and align observations to it



Add-on: Model bias – “Einstein from noise”

Naive Approach:

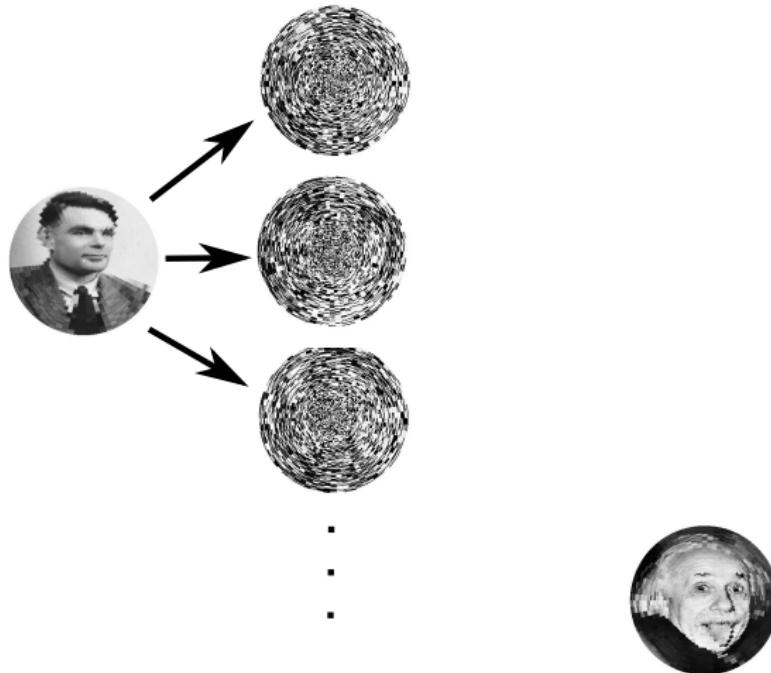
Use prior beliefs to construct a **model image** and align observations to it



Add-on: Model bias – “Einstein from noise”

Naive Approach:

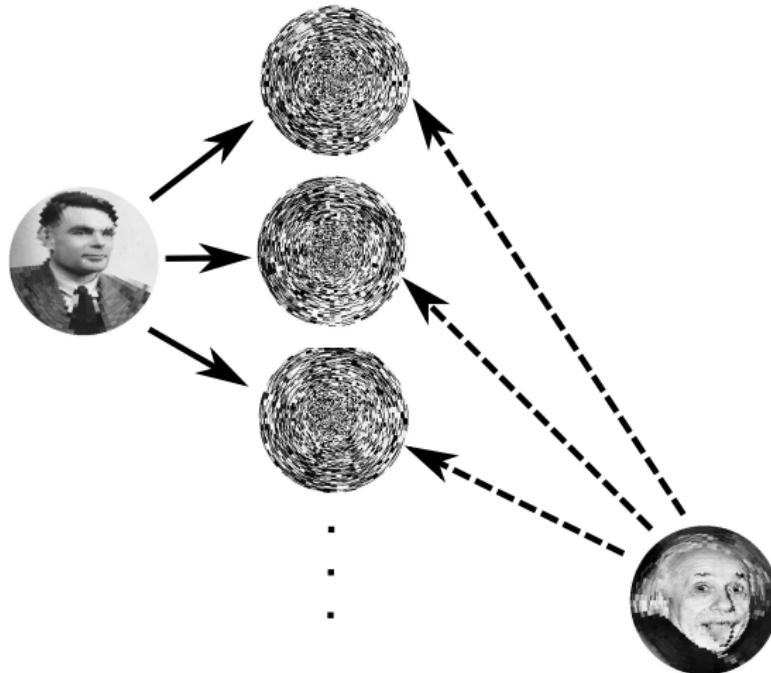
Use prior beliefs to construct a **model image** and align observations to it



Add-on: Model bias – “Einstein from noise”

Naive Approach:

Use prior beliefs to construct a **model image** and align observations to it



Add-on: Model bias – “Einstein from noise”

Naive Approach:

Use prior beliefs to construct a **model image** and align observations to it

