Orbit Retrieval with Applications to Cryo-Electron Microcopy

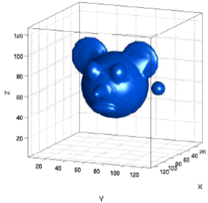
Joe Kileel, *Princeton University*

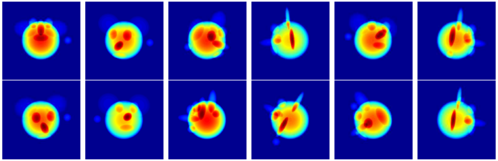
New York, December 2018

Thanks to many collaborators

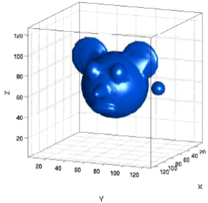
- Amit Singer (Princeton)
- Afonso Bandeira (NYU)
- Alex Wein (NYU)
- Ben Blum-Smith (NYU)
- Amelia Perry (MIT)
- Jon Weed (MIT)
- Nir Sharon (Tel Aviv)
- Yuehaw Khoo (Stanford)
- Tamir Bendory (Princeton)
- Nicolas Boumal (Princeton)
- João Pereira (Princeton)
- Emmanuel Abbe (Princeton)
- Eitan Levin (Princeton)
- Boris Landa (Tel Aviv)

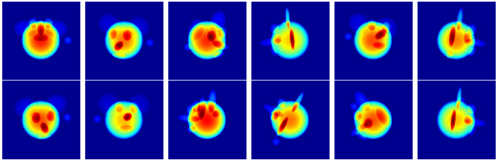
Cryo-EM: 3D protein structure from 2D images

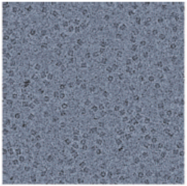




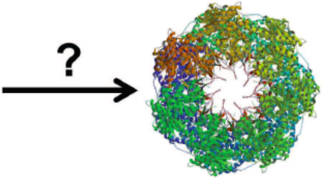
Cryo-EM: 3D protein structure from 2D images







tomographic image

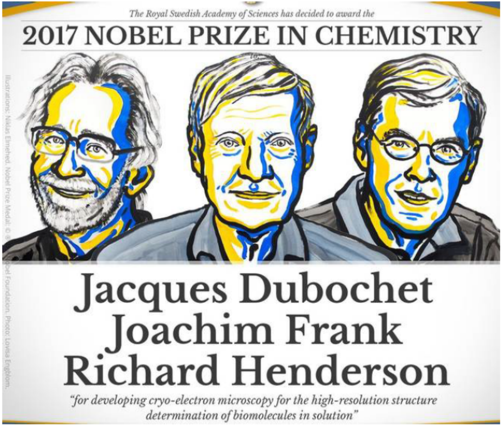


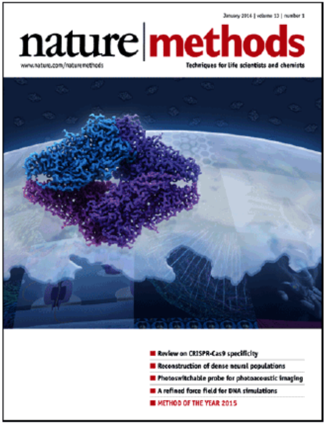
3-D structure

Given: $\sim 10^5$ very noisy 2D images from unknown viewing directions

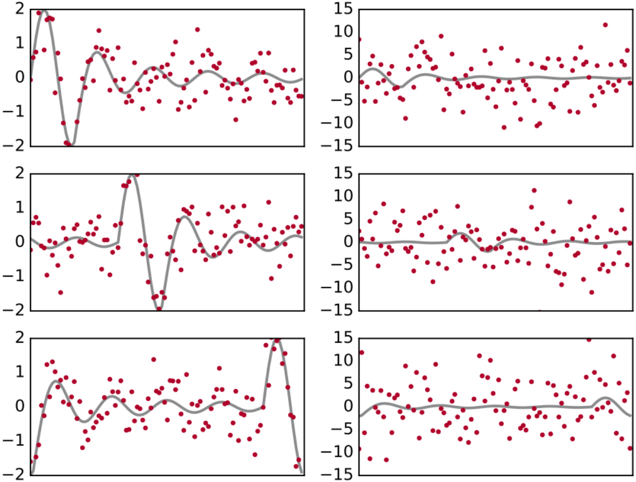
Want: 3D structure at resolution $\sim 3 \times 10^{-10}$ m

Much promise for cryo-EM





Multi-reference alignment: cyclically shifted, noisy signals



Orbit retrieval: a common abstraction

Let $G \curvearrowright V$ where G is a compact group acting linearly, continuously and orthogonally on $V = \mathbb{R}^L$.

Let $\Pi : V \to W$ be a linear map where $W = \mathbb{R}^K$.

Let $x \in V$ be fixed.

We observe projected, rotated, noisy copies, precisely i.i.d. realizations of:

$$y = \Pi(g \cdot x) + \epsilon$$

where $g \sim \mathrm{Haar}(G)$ and $\epsilon \sim \mathcal{N}(0, \sigma^2 I_K)$.

The goal is to estimate the orbit $G \cdot x$.

Orbit retrieval: a common abstraction

Let $G \curvearrowright V$ where G is a compact group acting linearly, continuously and orthogonally on $V = \mathbb{R}^L$.

Let $\Pi : V \to W$ be a linear map where $W = \mathbb{R}^K$.

Let $x \in V$ be fixed.

We observe projected, rotated, noisy copies, precisely i.i.d. realizations of:

$$y = \Pi(g \cdot x) + \epsilon$$

where $g \sim \mathrm{Haar}(G)$ and $\epsilon \sim \mathcal{N}(0, \sigma^2 I_K)$.

The goal is to estimate the orbit $G \cdot x$.

<u>MRA</u>: $\mathbb{Z}/L\mathbb{Z} \curvearrowright \mathbb{R}^L$ by cyclic shifts, $\Pi = \mathrm{id}$

<u>Cryo-EM</u>: $\mathrm{SO}(3) \curvearrowright \{$band-limited molecular densities$\}$, $\Pi =$ tomographic projection

Fundamental obstacle at low SNR & state-of-the-art

Estimation of rotations is impossible when σ^2 is very big

Consider an oracle that knows x (the 3D structure). The oracle would estimate g's (the rotations) by generating projections and matching them to observations. The oracle would suffer large errors at very low SNR.

Fundamental obstacle at low SNR & state-of-the-art

Estimation of rotations is impossible when σ^2 is very big

Consider an oracle that knows x (the 3D structure). The oracle would estimate g's (the rotations) by generating projections and matching them to observations. The oracle would suffer large errors at very low SNR.

RELION, the leading reconstruction software for cryo-EM, does not attempt to estimate one rotation per each experimental image. Instead, RELION takes a Bayesian approach, estimating a probability distribution of rotations per each image. Holding these fixed, it estimates the 3D structure. Then the rotational distributions are updated with the 3D structure fixed, and so on . . .

Iterative refinement has model bias, is computationally intensive, and lacks rigorous guarantees.

Invariant features approach / Kam's method

To bypass estimation of g, we could average it out by calculating moments of the data and equating with population moments:

Invariant features approach / Kam's method

To bypass estimation of g, we could average it out by calculating moments of the data and equating with population moments:

samples are i.i.d. draws of $y = \Pi(g \cdot x) + \epsilon$

sample average $\approx \mathbb{E}_{g,\epsilon}[y] = \Pi \mathbb{E}_g[g \cdot x]$

sample 2nd *moment* $\approx \mathbb{E}_{g,\epsilon}[y^{\otimes 2}] = \Pi^{\otimes 2} \mathbb{E}_g[(g \cdot x)^{\otimes 2}] + \text{"bias"}$

sample 3rd *moment* $\approx \mathbb{E}_{g,\epsilon}[y^{\otimes 3}] = \Pi^{\otimes 3} \mathbb{E}_g[(g \cdot x)^{\otimes 3}] + \text{"bias"}$

... [should have $O(\sigma^{2d})$ samples to estimate the d^{th} population moment]

Invariant features approach / Kam's method

To bypass estimation of g, we could average it out by calculating moments of the data and equating with population moments:

samples are i.i.d. draws of $y = \Pi(g \cdot x) + \epsilon$

sample average $\approx \mathbb{E}_{g,\epsilon}[y] = \Pi \mathbb{E}_g[g \cdot x]$

sample 2$^{\text{nd}}$ *moment* $\approx \mathbb{E}_{g,\epsilon}[y^{\otimes 2}] = \Pi^{\otimes 2} \mathbb{E}_g[(g \cdot x)^{\otimes 2}] + \text{"bias"}$

sample 3$^{\text{rd}}$ *moment* $\approx \mathbb{E}_{g,\epsilon}[y^{\otimes 3}] = \Pi^{\otimes 3} \mathbb{E}_g[(g \cdot x)^{\otimes 3}] + \text{"bias"}$

... [should have $O(\sigma^{2d})$ samples to estimate the d^{th} population moment]

> **Invariant features approach**: estimate enough moments so that $G \cdot x$ is determined and then estimate $G \cdot x$ by (somehow) solving a (noisy) polynomial system

Note: the red terms are exactly the low-degree invariants in $\mathbb{R}[V]^G$

Invariant features for MRA

What are low-degree invariants for MRA? How to invert them?

Invariant features for MRA

What are low-degree invariants for MRA? How to invert them?

Since $\mathbb{Z}/L\mathbb{Z}$ is finite abelian, its action may be diagonalized over $\mathbb{C}$.

DFT does this. $\mathbb{Z}/L\mathbb{Z}$ acts by modulating the phase of each Fourier coefficient:

$$(s \cdot \hat{x})[k] = \exp\left(\frac{2\pi isk}{L}\right)\hat{x}[k]$$

Thus the invariants are monomial in this basis.

- *DC component*: $\hat{x}[0]$
- *Power spectrum*: $\hat{x}[k]\hat{x}[-k] \, : \, k = 0, \ldots, L-1$
- *Bispectrum*: $\hat{x}[k_1]\hat{x}[k_2]\hat{x}[k_3] \, : \, k_1 + k_2 + k_3 = 0 \,(\mathrm{mod}\ L)$

Invariant features for MRA

What are low-degree invariants for MRA? How to invert them?

Since $\mathbb{Z}/L\mathbb{Z}$ is finite abelian, its action may be diagonalized over $\mathbb{C}$.

DFT does this. $\mathbb{Z}/L\mathbb{Z}$ acts by modulating the phase of each Fourier coefficient:

$$(s \cdot \hat{x})[k] = \exp\left(\frac{2\pi i s k}{L}\right)\hat{x}[k]$$

Thus the invariants are monomial in this basis.

- *DC component*: $\hat{x}[0]$
- *Power spectrum*: $\hat{x}[k]\hat{x}[-k] \;:\; k = 0, \ldots, L-1$
- *Bispectrum*: $\hat{x}[k_1]\hat{x}[k_2]\hat{x}[k_3] \;:\; k_1 + k_2 + k_3 = 0 \pmod{L}$

Assuming all DFT coefficients are nonzero, recover $\hat{x}$ by multiplying/dividing:

$$\hat{x}[1]^L = \frac{\prod_{k=0}^{L-1} \hat{x}[1]\hat{x}[k]\hat{x}[-1-k]}{\prod_{k=0}^{L-1} \hat{x}[k]\hat{x}[-k]}, \quad \hat{x}[-1] = \frac{\hat{x}[1]\hat{x}[-1]}{\hat{x}[1]}, \quad \hat{x}[2] = \frac{\hat{x}[2]\hat{x}[-1]\hat{x}[-1]}{\hat{x}[-1]^2}, \ldots$$

More stable bispectrum inversion is by a certain eigendecomposition (provably) and even better (empirically) is by non-convex least squares moment-fitting.

Statistical optimality for invariant features in large σ^2 limit

Theorem (2018)

Let $x \in V$. Let $d \in \mathbb{Z}_{>0}$ be the least degree such that for $x' \in V$:

$$\begin{cases} \Pi\mathbb{E}[g \cdot x'] = \Pi\mathbb{E}[g \cdot x] \\ \Pi^{\otimes 2}\mathbb{E}[(g \cdot x')^{\otimes 2}] = \Pi^{\otimes 2}\mathbb{E}[(g \cdot x)^{\otimes 2}] \\ \vdots \\ \Pi^{\otimes d}\mathbb{E}[(g \cdot x')^{\otimes d}] = \Pi^{\otimes d}\mathbb{E}[(g \cdot x)^{\otimes d}] \end{cases}$$

implies $G \cdot x' = G \cdot x$. Then **any** *estimation procedure requires $O(\sigma^{2d})$ samples to accurately estimate $G \cdot x$ with high probability, as $\sigma^2 \to \infty$.*

Statistical optimality for invariant features in large σ^2 limit

Theorem (2018)

Let $x \in V$. Let $d \in \mathbb{Z}_{>0}$ be the least degree such that for $x' \in V$:

$$\begin{cases} \Pi\mathbb{E}[g \cdot x'] = \Pi\mathbb{E}[g \cdot x] \\ \Pi^{\otimes 2}\mathbb{E}[(g \cdot x')^{\otimes 2}] = \Pi^{\otimes 2}\mathbb{E}[(g \cdot x)^{\otimes 2}] \\ \vdots \\ \Pi^{\otimes d}\mathbb{E}[(g \cdot x')^{\otimes d}] = \Pi^{\otimes d}\mathbb{E}[(g \cdot x)^{\otimes d}] \end{cases}$$

implies $G \cdot x' = G \cdot x$. Then **any** *estimation procedure requires* $O(\sigma^{2d})$ *samples to accurately estimate $G \cdot x$ with high probability, as $\sigma^2 \to \infty$.*

Sketch: for $x, x' \in V$ the *Kullback-Leibler divergence* between the probability distribution for an observation with x and with x' is bounded above by the following Taylor expansion in σ^{-2}:

$$e^2 \sum_k \frac{\sigma^{-2k}}{k!} \left\| \Pi^{\otimes k}\mathbb{E}_g[(g \cdot x)^{\otimes k}] - \Pi^{\otimes k}\mathbb{E}_g[(g \cdot x')^{\otimes k}] \right\|_{HS}^2$$

A. Bandeira, B. Blum-Smith, J. Kileel, A. Perry, J. Weed, A. Wein, *submitted 2018*

Purely algebraic questions

Sample complexity for orbit retrieval ($+$ various weakenings) is answered by purely algebraic questions about invariants. If $\Pi = \mathrm{id}$, these are

- *Unique recovery*: smallest d s.t. a separating subalgebra of $\mathbb{R}[V]^G$ is generated by $\mathbb{R}[V]^G_{\leq d}$
- *Generic unique recovery*: smallest d s.t. $\mathbb{R}(V)^G$ is generated by $\mathbb{R}[V]^G_{\leq d}$
- *Generic list recovery*: smallest d s.t. $\mathbb{R}[V]^G_{\leq d}$ generates a subfield of $\mathbb{R}(V)^G$ of full transcendence degree

Purely algebraic questions

Sample complexity for orbit retrieval (+ various weakenings) is answered by purely algebraic questions about invariants. If $\Pi = \mathrm{id}$, these are

- *Unique recovery*: smallest d s.t. a separating subalgebra of $\mathbb{R}[V]^G$ is generated by $\mathbb{R}[V]^G_{\leq d}$
- *Generic unique recovery*: smallest d s.t. $\mathbb{R}(V)^G$ is generated by $\mathbb{R}[V]^G_{\leq d}$
- *Generic list recovery*: smallest d s.t. $\mathbb{R}[V]^G_{\leq d}$ generates a subfield of $\mathbb{R}(V)^G$ of full transcendence degree

MRA has invariant fraction field generated at $d = 3$. Thus for generic $x \in V$, need $\mathrm{O}(\sigma^6)$ samples to uniquely recover $\mathbb{Z}/L\mathbb{Z} \cdot x$ (by any means)

The Jacobian criterion efficiently calculates the d for generic list recovery

Low-degree ring generation studied classically, field generation neglected

Back to cryo-EM

Theorem (2018)

Cryo-EM with projection has sample complexity $O(\sigma^6)$*, in the sense of generic list recovery.*

Cryo-EM with no projection has sample complexity $O(\sigma^6)$*, in the sense of generic unique recovery. The cubic, quadratic and linear invariant features may be birationally inverted* **efficiently***, via Cholesky factorizations of matrices, orthogonal Tucker factorizations of tensors and frequency marching.*

Back to cryo-EM

Theorem (2018)

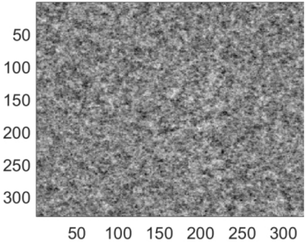
Cryo-EM with projection has sample complexity $O(\sigma^6)$*, in the sense of generic list recovery.*

Cryo-EM with no projection has sample complexity $O(\sigma^6)$*, in the sense of generic unique recovery. The cubic, quadratic and linear invariant features may be birationally inverted* **efficiently***, via Cholesky factorizations of matrices, orthogonal Tucker factorizations of tensors and frequency marching.*

- *cubic invariants for cryo-EM*: $\mathcal{P}_3(s_1, m_1, s_2, m_2, s_3, m_3) =$
  $\sum_\ell N_{\ell_1 m_1} N_{\ell_2 m_2} N_{\ell_3 m_3} P_{\ell_1}^{m_1}(0) P_{\ell_2}^{m_2}(0) P_{\ell_3}^{m_3}(0) \langle \ell_2 m_2 \ell_3 m_3 \mid \ell_1(-m_1)\rangle \mathcal{I}_3(s_1, \ell_1, s_2, \ell_2, s_3, \ell_3)$
- *cubic invariants for cryo-ET*:
  $\mathcal{I}_3(s_1, \ell_1, s_2, \ell_2, s_3, \ell_3) = \frac{1}{2\ell_1+1} \sum_k (-1)^{k_1} \langle \ell_2 k_2 \ell_3 k_3 \mid \ell_1(-k_1)\rangle x_{s_1 \ell_1 k_1} x_{s_2 \ell_2 k_2} x_{s_3 \ell_3 k_3}$
- *variables*: $x_{s\ell k}$, spherical harmonic coefficients for the molecule
- *special constants*: Clebsch-Gordan coefficients for SO(3), specializations of associated Legendre polynomials, normalizations for spherical hamonics

A. Bandeira, B. Blum-Smith, J. Kileel, A. Perry, J. Weed, A. Wein, *submitted 2018*

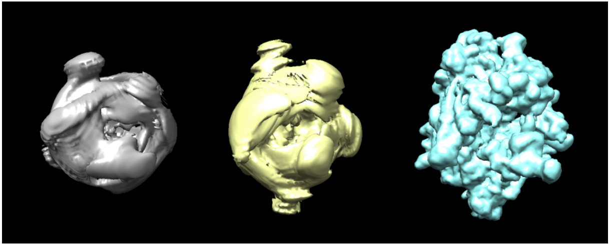
Results on real data: *ab initio* modeling



$\times \;\; 3.5 \times 10^5$

Yeast Mitochondrial Ribosome

`EMPIAR-10107`



reconstruction *low-pass molecule* *molecule*

E. Levin, T. Bendory, N. Boumal, J. Kileel, A. Singer, *IEEE ISBI 2018*

Extension: cryo-EM with non-uniform rotations

Often molecules have preferred orientations. It is possible to learn both the rotational distribution and the molecule from moments, by solving a polynomial system with more variables.

- Molecule in spherical harmonics: $x_{s\ell k}$ coefficients
- Rotational pdf in Wigner entries: $\rho(R) = \sum_{\ell,m,n} b^{\ell}_{m,n} D^{\ell}_{m,n}(R)$
- Both $x_{s\ell k}$ and $b^{\ell}_{m,n}$ are unknown variables

[3 band-limits: molecule's angular frequency, molecule's radial resolution, rotational distribution's band-limit]

Non-uniform rotations may be easier

Theorem (2019)

Cryo-EM with projection and unknown distribution on $\mathrm{SO}(3)$ *has sample complexity* $\mathrm{O}(\sigma^4)$, *in the sense of generic list recovery, for certain band-limits on the rotational distribution and molecule.*

Non-uniform rotations may be easier

Theorem (2019)

Cryo-EM with projection and unknown distribution on SO(3) *has sample complexity* $O(\sigma^4)$*, in the sense of generic list recovery, for certain band-limits on the rotational distribution and molecule.*

We are developing a non-convex optimization solver, using just first and second moments and least squares:

$$\operatorname{argmin}_{x,b} \quad \left(\int (R \cdot x)\, d\rho(R) \;-\; \text{sample } 1^{\text{st}} \text{ moment} \right)^2$$

$$+ \;\; \lambda \left(\int (R \cdot x)^{\otimes 2}\, d\rho(R) \;-\; \text{sample } 2^{\text{nd}} \text{ moment de-biased} \right)^2$$

N. Sharon, Y. Khoo, J. Kileel, B. Landa, A. Singer, *preliminary results at Princeton–Nature Conference 2018*

A. Singer et al., *USPTO application 2018*

Extension: continuous heterogeneity

Often datasets contain images of various molecular conformations. How to deal with a continuum of conformations?

Testing ground: continuously heterogeneous MRA

Observe samples $y = R_\ell(\theta_1 s_1 + \ldots + \theta_k s_k) + \epsilon$

$k \ll n$

$s_1, \ldots, s_k \in \mathbb{R}^n$ latent signals (generating a subspace)

$\theta \in \mathbb{R}^k$ random vector (some probability distribution)

$\ell \sim \mathrm{Unif}(\mathbb{Z}/n\mathbb{Z})$

$\epsilon \sim \mathcal{N}(0, \sigma^2 I),\ \sigma^2 \gg 0$

Goals: recover subspace, recover $s_1, \ldots, s_k$ (up to shifts)

Continuous heterogeneity and tensor decomposition

Continuously heterogeneous MRA relates to the following kind of symmetric tensor decomposition.

Let $F \in \mathbb{R}[X_1, \ldots, X_n]$ be a polynomial of degree d.

Write $F = f_1 + \ldots + f_r$, with r minimal, such that each summand f_i equals a degree d polynomial in k linear forms.

For $k = 1$, this is CP decomposition. For higher k, this seems new.

Continuous heterogeneity and tensor decomposition

Continuously heterogeneous MRA relates to the following kind of symmetric tensor decomposition.

Let $F \in \mathbb{R}[X_1, \ldots, X_n]$ be a polynomial of degree d.

Write $F = f_1 + \ldots + f_r$, with r minimal, such that each summand f_i equals a degree d polynomial in k linear forms.

For $k = 1$, this is CP decomposition. For higher k, this seems new.

> We are developing ways to decompose this way (small enough r).
>
> Key idea: *apolarity* from classical CP tensor decomposition, i.e., exploiting the linear differential operators that annihilate F.

E. Abbe, T. Bendory, J. Kileel, J. Pereira, A. Singer, *in-preparation 2019*

Wrap-up

- Defined general orbit retrieval, to model estimation problems with group actions in image/signal processing, computer vision and robotics, especially cryo-EM.
- At low SNR, rotation estimation is impossible and invariant features is sample-efficient.
- Get various generation problems about low-degree invariants.
- Inverting moments is polynomial solving/optimization.
- Add-ons to the orbit retrieval model are work-in-progress, as is implementation with real data.

Wrap-up

- Defined general orbit retrieval, to model estimation problems with group actions in image/signal processing, computer vision and robotics, especially cryo-EM.
- At low SNR, rotation estimation is impossible and invariant features is sample-efficient.
- Get various generation problems about low-degree invariants.
- Inverting moments is polynomial solving/optimization.
- Add-ons to the orbit retrieval model are work-in-progress, as is implementation with real data.

Thank you for your attention!

References

- A. Bandeira, B. Blum-Smith, J. Kileel, A. Perry, J. Weed, A. Wein, *Estimation under group actions: recovering orbits from invariants*, submitted 2018

- E. Levin, T. Bendory, N. Boumal, J. Kileel, A. Singer, *3D* ab initio *modeling in cryo-EM by autocorrelation analysis*, *IEEE International Symposium on Biomedical Imaging 2018*, pp. 1569-1573

- N. Sharon, Y. Khoo, J. Kileel, B. Landa, A. Singer, *Work-in-progress on non-uniform rotational distributions*, 2019

- E. Abbe, T. Bendory, J. Kileel, J. Pereira, A. Singer, *Work-in-progress on continuous heterogeneity*, 2019