Estimation in the Presence of Group Actions

Alex Wein
MIT Mathematics

Joint work with:

Amelia Perry
1991 – 2018

Joint work with:

Afonso Bandeira

Ben Blum-Smith

Jonathan Weed

Ankur Moitra

Group actions

G – compact group, e.g.

- S_n (permutations of $\{1, 2, \ldots, n\}$)
- $\mathbb{Z}/n$ (cyclic / integers mod n)
- any finite group
- $SO(2)$ (2D rotations)
- $SO(3)$ (3D rotations)

Group action $G \circlearrowleft V$: map $G \times V \to V$, write $g \cdot x$
Axioms: $1 \cdot x = x$ and $g \cdot (h \cdot x) = (gh) \cdot x$

- $S_n \circlearrowleft \mathbb{R}^n$ (permute coordinates)
- $\mathbb{Z}/n \circlearrowleft \mathbb{R}^n$ (permute coordinates cyclically)
- $SO(2) \circlearrowleft \mathbb{R}^2$ (rotate vector)
- $SO(3) \circlearrowleft \mathbb{R}^3$ (rotate vector)
- $SO(3) \circlearrowleft \mathbb{R}^n$ (rotate some object...)

Motivation: cryo-electron microscopy (cryo-EM)

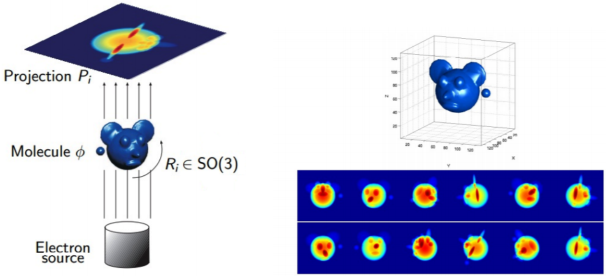

Image credit: [Singer, Shkolnisky '11]

- Biological imaging method: determine structure of molecule
- 2017 Nobel Prize in Chemistry
- Given many noisy 2D images of a 3D molecule, taken from different unknown angles
- Goal is to reconstruct the 3D structure of the molecule
- Group action $SO(3) \circlearrowleft \mathbb{R}^n$

Other examples

Other problems involving random group actions:

- Image registration

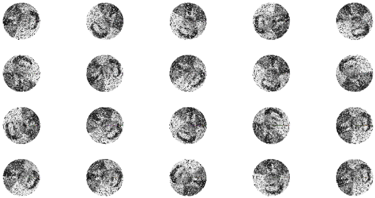

Image credit: [Bandeira, PhD thesis '15]

Group: SO(2) (2D rotations)

- Multi-reference alignment

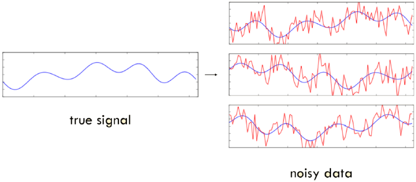

Image credit: Jonathan Weed

Group: $\mathbb{Z}/p$ (cyclic shifts)

- Applications: computer vision, radar, structural biology, robotics, geology, paleontology, ...
- Methods used in practice often lack provable guarantees...

Orbit recovery problem

Let G be a compact group acting linearly on a finite-dimensional real vector space $V = \mathbb{R}^p$.

- Linear: homomorphism $\rho : G \to \mathrm{GL}(V)$

  $\mathrm{GL}(V) = \{\text{invertible } p \times p \text{ matrices}\}$

- Action: $g \cdot x = \rho(g)x \qquad \text{for } g \in G, x \in V$

- Equivalently: G is a subgroup of matrices $\mathrm{GL}(V)$

Orbit recovery problem

Let G be a compact group acting linearly on a finite-dimensional real vector space $V = \mathbb{R}^p$.

Unknown signal $x \in V$ (e.g. the molecule)

For $i = 1, \dots, n$ observe $y_i = g_i \cdot x + \varepsilon_i$ where...

- $g_i \sim \mathrm{Haar}(G)$ (“uniform distribution” on G)
- $\varepsilon_i \sim \mathcal{N}(0, \sigma^2 \mathrm{I}_p)$ (noise)

Goal: Recover some $\tilde{x}$ in the orbit $\{g \cdot x \,:\, g \in G\}$ of x

Special case: multi-reference alignment (MRA)

$G = \mathbb{Z}/p$ acts on $\mathbb{R}^p$ via cyclic shifts

For $i = 1, \ldots, n$ observe $y_i = g_i \cdot x + \varepsilon_i$ with $\varepsilon_i \sim \mathcal{N}(0, \sigma^2 \mathrm{I})$

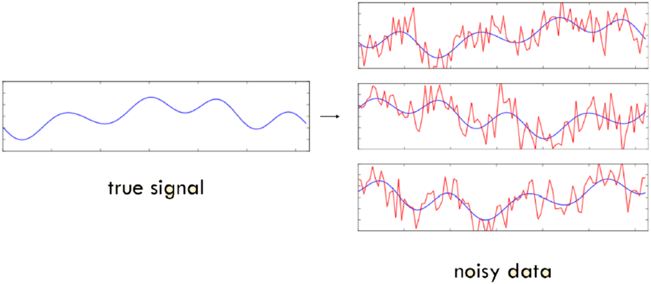

Image credit: Jonathan Weed

Special case: multi-reference alignment (MRA)

$G = \mathbb{Z}/p$ acts on $\mathbb{R}^p$ via cyclic shifts

For $i = 1, \ldots, n$ observe $y_i = g_i \cdot x + \varepsilon_i$ with $\varepsilon_i \sim \mathcal{N}(0, \sigma^2 \mathrm{I})$

How to solve this?

Maximum likelihood?

- Optimal rate but computationally intractable [1]

Synchronization? (learn the group elements / align the samples) [2]

- Can't learn the group elements if noise is too large

Iterative method? (EM, belief propagation)

- Not sure how to analyze...

[1] Bandeira, Rigollet, Weed, *Optimal rates of estimation for multi-reference alignment*, 2017

[2] Singer, *Angular Synchronization by Eigenvectors and Semidefinite Programming*, 2011

Method of invariants

Idea: measure features of the signal x that are shift-invariant [1,2]

Degree-1: $\sum_i x_i$ (mean)

Degree-2: $\sum_i x_i^2, \quad x_1x_2 + x_2x_3 + \cdots + x_px_1, \ \ldots$ (autocorrelation)

Degree-3: $x_1x_2x_4 + x_2x_3x_5 + \ldots$ (triple correlation)

Invariant features are easy to estimate from the samples

[1] Bandeira, Rigollet, Weed, *Optimal rates of estimation for multi-reference alignment*, 2017

[2] Perry, Weed, Bandeira, Rigollet, Singer, *The sample complexity of multi-reference alignment*, 2017

Sample complexity

Theorem [1]:
(Upper bound) With noise level σ, can estimate degree-d invariants using $n = O(\sigma^{2d})$ samples.
(Lower bound) If $x^{(1)}, x^{(2)}$ agree on all invariants of degree $\leq d - 1$ then $\Omega(\sigma^{2d})$ samples are required to distinguish them.

- Method of invariants is optimal

Question: What degree d^* of invariants do we need to learn before we can recover x (up to orbit)?

- Optimal sample complexity is $n = \Theta(\sigma^{2d^*})$

Answer (for MRA) [1]:

- For "generic" x, degree 3 is sufficient, so sample complexity $n = \Theta(\sigma^6)$
- But for a measure-zero set of "bad" signals, need much higher degree (as high as p)

---

[1] Bandeira, Rigollet, Weed, *Optimal rates of estimation for multi-reference alignment*, 2017

Another viewpoint: mixtures of Gaussians

MRA sample: $y = g \cdot x + \varepsilon$ with $g \sim G$, $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathrm{I})$

The distribution of y is a (uniform) mixture of $|G|$ Gaussians centered at $\{g \cdot x \; : \; g \in G\}$

- For infinite groups, a mixture of infinitely-many Gaussians

Method of moments: Estimate moments $\mathbb{E}[y], \mathbb{E}[yy^\top], \ldots, \mathbb{E}[y^{\otimes d}]$

De-bias to get moments of signal term: $\mathbb{E}[y^{\otimes k}] \rightsquigarrow \mathbb{E}_g[(g \cdot x)^{\otimes k}]$

Fact: Moments are equivalent to invariants

- $\mathbb{E}_g[(g \cdot x)^{\otimes k}]$ contains the same information as the degree-k invariant polynomials

Our contributions

Joint work with Ben Blum-Smith, Afonso Bandeira, Amelia Perry, Jonathan Weed [1]

- We generalize from MRA to any compact group
- Again, the method of invariants/moments is optimal
  - Independently by [2]
- We give an (inefficient) algorithm that achieves optimal sample complexity: solve polynomial system
- To determine what degree of invariants are required, we use invariant theory and algebraic geometry

---

[1] Bandeira, Blum-Smith, Perry, Weed, W., *Estimation under group actions: recovering orbits from invariants*, 2017

[2] Abbe, Pereira, Singer, *Estimation in the group action channel*, 2018

Invariant theory

Variables $x_1, \ldots, x_p$ (corresponding to the coordinates of x)

The invariant ring $\mathbb{R}[\mathbf{x}]^G$ is the subring of $\mathbb{R}[\mathbf{x}] := \mathbb{R}[x_1, \ldots, x_p]$ consisting of polynomials f such that $f(g \cdot \mathbf{x}) = f(\mathbf{x})\ \forall g \in G$.

- Aside: A main result of invariant theory is that $\mathbb{R}[\mathbf{x}]^G$ is finitely-generated

$\mathbb{R}[\mathbf{x}]^G_{\leq d}$ – invariants of degree $\leq d$

(Simple) algorithm:

- Pick d^* (to be chosen later)
- Using $\Theta(\sigma^{2d^*})$ samples, estimate invariants up to degree d^*: learn value $f(x)$ for all $f \in \mathbb{R}[\mathbf{x}]^G_{\leq d}$
- Solve for an $\hat{x}$ that is consistent with those values: $f(\hat{x}) = f(x)\ \forall f \in \mathbb{R}[\mathbf{x}]^G_{\leq d}$ (polynomial system of equations)

Example: norm recovery

$G = SO(3)$ acting on $\mathbb{R}^3$ (by rotation)

Samples: noisy, randomly-rotated copies of $x \in \mathbb{R}^3$

To learn orbit, need to learn $\|x\|$

Invariant ring is generated by $\|x\|^2 = \sum_i x_i^2$

- $d^* = 2$

Sample complexity $\Theta(\sigma^{2d^*}) = \Theta(\sigma^4)$

Example: learning a "bag of numbers"

$G = S_p$ acting on $\mathbb{R}^p$ (by permuting coordinates)

Samples: noisy copes of $x \in \mathbb{R}^p$ with entries permuted randomly

To learn orbit, need to learn the multiset $\{x_i\}_{i \in [p]}$

Invariants are the symmetric polynomials

- Generated by elementary symmetric polynomials:

$$e_1 = \sum_i x_i,\ e_2 = \sum_{i<j} x_i x_j,\ e_3 = \sum_{i<j<k} x_i x_j x_k,\ \ldots$$

Can't learn $e_p = \prod_{i=1}^p x_i$ until degree p

- $d^* = p$ so sample complexity $\Theta(\sigma^{2p})$

All invariants determine orbit

Theorem [1]: If G is compact, for every $x \in V$, the full invariant ring $\mathbb{R}[\mathbf{x}]^G$ determines x up to orbit.

- In the sense that if x, x' do not lie in the same orbit, there exists $f \in \mathbb{R}[\mathbf{x}]^G$ that separates them: $f(x) \neq f(x')$

Corollary: Suppose that for some d, $\mathbb{R}[\mathbf{x}]^G_{\leq d}$ generates $\mathbb{R}[\mathbf{x}]^G$ (as an $\mathbb{R}$-algebra). Then $\mathbb{R}[\mathbf{x}]^G_{\leq d}$ determines x up to orbit and so sample complexity is $O(\sigma^{2d})$.

Problem: This is for worst-case $x \in V$. For MRA (cyclic shifts) this requires $d = p$ whereas generic x only requires $d = 3$ [2].

Actually care about whether $\mathbb{R}[\mathbf{x}]^G_{\leq d}$ generically determines $\mathbb{R}[\mathbf{x}]^G$

- "Generic" means that x lies outside a particular measure-zero "bad" set.

[1] Kač, Invariant theory lecture notes, 1994

[2] Bandeira, Rigollet, Weed, *Optimal rates of estimation for multi-reference alignment*, 2017

Do polynomials <u>generically</u> determine other polynomials?

Say we have $A \subseteq B \subseteq \mathbb{R}[\mathbf{x}]$

- (Technically need to assume B is finitely generated)

Question: Do the values $\{a(x) \ : \ a \in A\}$ generically determine the values $\{b(x) \ : \ b \in B\}$?

- Formally: does there exist a full-measure set $S \subseteq V$ such that if $x \in S$ ("generic") then any $x' \in V$ satisfying $a(x) = a(x') \ \forall a \in A$ also satisfies $b(x) = b(x') \ \forall b \in B$

Definition: Polynomials $f_1, \ldots, f_m$ are algebraically independent if there is no $P \in \mathbb{R}[y_1, \ldots, y_m]$ with $P(f_1, \ldots, f_m) \equiv 0$.

Definition: For $U \subseteq \mathbb{R}[\mathbf{x}]$, the transcendence degree $\mathrm{trdeg}(U)$ is the number of algebraically independent polynomials in U.

Do polynomials generically determine other polynomials?

Definition: For $U \subseteq \mathbb{R}[\mathbf{x}]$, the transcendence degree $\mathrm{trdeg}(U)$ is the number of algebraically independent polynomials in U.

Answer: Suppose $\mathrm{trdeg}(A) = \mathrm{trdeg}(B)$. If x is generic then the values $\{a(x) \ : \ a \in A\}$ determine a finite number of possibilities for the entire collection $\{b(x) \ : \ b \in B\}$.

- Formally: for generic x there is a finite list $x^{(1)}, \ldots, x^{(s)}$ such that for any x' satisfying $a(x) = a(x') \ \forall a \in A$ there exists i such that $b(x^{(i)}) = b(x') \ \forall b \in B$

A determines B (up to finite ambiguity) if A has as many algebraically independent polynomials as B

- Intuition: algebraically independent polynomials are "degrees-of-freedom"

Testing algebraic independence

Given polynomials $f_1, \ldots, f_m \in \mathbb{R}[x_1, \ldots, x_p]$, can you efficiently test whether they are algebraically independent?

Answer: yes!

Theorem (Jacobian criterion):
Polynomials $f_1, \ldots, f_m \in \mathbb{R}[x_1, \ldots, x_p]$ are algebraically independent if and only if the $m \times p$ Jacobian matrix $J_{ij} = \frac{\partial f_i}{\partial x_j}$ has full row rank. (Still true if you evaluate J at a generic point x.)

- Why: Tests whether map $(x_1, \ldots, x_p) \mapsto (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$ is locally surjective

Generic list recovery

Our main result is an efficient procedure that takes the problem setup as input (group G and action on V) and outputs the degree d^* of invariants required for generic list recovery.

- List recovery: output a finite list $\hat{x}^{(1)}, \hat{x}^{(2)}, \ldots$, one of which (approximately) lies in the orbit of the true x
- List recovery may be good enough in practice?

Procedure:

- Need to test whether $\mathbb{R}[\mathbf{x}]^G_{\leq d}$ determines $\mathbb{R}[\mathbf{x}]^G$ (generically)
- So need to check if $\text{trdeg}(\mathbb{R}[\mathbf{x}]^G_{\leq d}) = \text{trdeg}(\mathbb{R}[\mathbf{x}]^G)$
- $\text{trdeg}(\mathbb{R}[\mathbf{x}]^G) = \dim(x) - \dim(\text{orbit})$ (d.o.f. needed)
- $\text{trdeg}(\mathbb{R}[\mathbf{x}]^G_{\leq d})$ via Jacobian criterion (d.o.f. have)

Generic list recovery

Our main result is an efficient procedure that takes the problem setup as input (group G and action on V) and outputs the degree d^* of invariants required for generic list recovery.

- List recovery: output a finite list $\hat{x}^{(1)}, \hat{x}^{(2)}, \ldots$, one of which (approximately) lies in the orbit of the true x
- List recovery may be good enough in practice?

Comments:

- For e.g. MRA (cyclic shifts), need to test each p separately on a computer
- Not an efficient algorithm to solve any particular instance
- There is also an algorithm to bound the size of the list (or test for unique recovery), but it is not efficient (Gröbner bases)

Generalized orbit recovery problem

Extensions:

- Post-projection (e.g. cryo-EM):
  - Observe $y_i = \Pi(g_i \cdot x) + \varepsilon_i$
  - $\Pi : V \to W$ linear
  - $\varepsilon_i \sim \mathcal{N}(0, \sigma^2 \mathrm{I})$
- Heterogeneity (mixture of signals):
  - K signals $x^{(1)}, \dots, x^{(K)}$
  - Mixing weights $(w_1, \dots, w_K) \in \Delta_K$
  - Observe $y_i = \Pi(g_i \cdot x^{(k_i)}) + \varepsilon_i$
  - $k_i \sim \{1, \dots, K\}$ according to w

Same methods apply!

- Order-d moments now only give access to a particular subspace of $\mathbb{R}[\mathbf{x}]^G$
- For heterogeneity, work over a bigger group G^K acting on $(x^{(1)}, \dots, x^{(K)}) \in V^{\oplus K}$

Results: cryo-EM

Our methods show that for cryo-EM, generic list recovery is possible at degree 3

So information-theoretic sample complexity is $\Theta(\sigma^6)$

Open: polynomial time algorithm for cryo-EM

Efficient recovery: tensor decomposition

Restrict to finite group

Recall: with $O(\sigma^6)$ samples, can estimate the third moment:

$$T_3(x) = \sum_{g \in G} (g \cdot x)^{\otimes 3}$$

This is an instance of tensor decomposition: Given $\sum_{i=1}^{m} a_i^{\otimes 3}$ for some $a_1, \ldots, a_m \in \mathbb{R}^p$, recover $\{a_i\}$

For MRA: since $m \leq p$ ("undercomplete") can apply Jennrich's algorithm to decompose tensor efficiently [1]

- Note: unique (not list) recovery

---

[1] Perry, Weed, Bandeira, Rigollet, Singer, *The sample complexity of multi-reference alignment*, 2017

Example: heterogeneous MRA

MRA with multiple signals $x^{(1)}, \dots, x^{(K)}$

$$T_d(x) = \sum_{k=1}^{K} \sum_{g \in G} (g \cdot x^{(k)})^{\otimes d}$$

Jennrich's algorithm works if given 5th moment $\rightsquigarrow n = O(\sigma^{10})$ [1]

Information-theoretically, 3rd moment suffices if $K \le p/6$

- Can even show unique recovery (upcoming with Joe Kileel)

If signals $x^{(k)}$ are random (i.i.d. Gaussian), conjectured that efficient recovery is possible from 3rd moment iff $K \le \sqrt{p}$ [2]

Theorem (with A. Moitra): if $K \le \sqrt{p}/\mathrm{polylog}(p)$ then for random signals, efficient recovery is possible from 3rd moment

- Based on random overcomplete 3-tensor decomposition [3]

---

[1] Perry, Weed, Bandeira, Rigollet, Singer '17

[2] Boumal, Bendory, Lederman, Singer '17

[3] Ma, Shi, Steurer '16

Open problems

- Analytic results for all problem sizes
- Efficiently test if unique recovery is possible
- Determine the computational limits
- Polynomial-time recovery for all groups

Thanks!