Estimation under group actions: recovering orbits from invariants

Afonso S. Bandeira, Ben Blum-Smith, Amelia Perry, Jonathan Weed, Alexander S. Wein

November 10, 2017

Abstract

We define and study a class of *orbit recovery* problems wherein we observe many noisy copies of an unknown signal, except each has been acted upon by a random element of some compact group. The goal is to recover the orbit of the signal under the group action. This generalizes various "synchronization" problems such as multi-reference alignment (MRA) and the reconstruction problem from cryo-electron microscopy (cryo-EM). Extending recent work on MRA [BRW17], we apply the method of moments and show that it yields optimal sample complexity. Using tools from algebraic geometry and invariant theory, we characterize the difficulty of orbit recovery problems under various success criteria. Our main focus is on the case of generic list recovery, where the signal is assumed to be generic and the goal is to output a finite list of candidate signals, one of which lies in the orbit of the true signal. For this case we give an efficient procedure that takes the problem setup as input (i.e. the group and its action) and determines the number of moments required to solve the problem, which in turn dictates the sample complexity. We also consider extensions of the basic orbit recover problem such post-projection and heterogeneous mixtures of signals.

1 Introduction

1.1 Background and prior work

Fix a compact group G acting (linearly, continuously, and orthogonally) on a vector space V. Let $\theta \in V$ be an unknown signal and suppose we observe many samples of the form $y_i = g_i \cdot \theta + \xi_i$ where g_i is drawn randomly from (the Haar measure on) G and $\xi_i \sim \mathcal{N}(0, \sigma^2 I)$ is noise. The goal is to recover the *orbit* of θ under the action of G. We call this an *orbit recovery* problem.

Many special cases of the orbit recovery problem have been studied extensively for their theoretical and practical interest (see, e.g., [Ban15] for an account of many such problems, which that work calls "synchronization" problems). One such problem is *multi-reference alignment* (MRA) [BCSZ14, BRW17, PWB$^+$17], which is motivated by applications in signal processing [ZvdHGG03, PZAF05] and structural biology [Dia92, TS12]. In this problem, one observes noisy copies of a signal $\theta \in \mathbb{R}^p$, each of which has its coordinates permuted by a random cyclic shift. This is an example of the orbit recovery problem when G is taken to be the cyclic group $\mathbb{Z}/p$ acting by circular shift on the standard basis vectors.

Another important example is the reconstruction problem arising in cryo-electron microscopy (cryo-EM) (see, e.g., [ADLM84, SS11, Nog16]), an imaging technique in structural biology that was recently awarded the 2017 Nobel Prize in Chemistry. This technique seeks to estimate the structure of a large biological molecule such as a protein on the basis of many noisy images of the molecule from random directions in 3-dimensional space. Here the signal θ is the molecule and we consider the group $G = SO(3)$ of rotations in $\mathbb{R}^3$. Each image is a 2-dimensional projection of the 3-dimensional molecule; we will consider a generalized orbit recovery problem that allows this type of observation. We also consider an additional extension called *heterogeneity* wherein each sample actually comes from one of K different molecules (but we don't know which) and we want to reconstruct all of them. This extension is extremely relevant in practice, since it is often impossible to guarantee that a biological sample contains molecules of only one type.

Many prior methods for orbit recovery problems employ the *synchronization* approach, meaning that they attempt to recover the group element associated with each sample (up to global right-multiplication by some group element). The signal itself is either reconstructed simultaneously with the group elements, or is easily recovered at the end once the group elements are known.

Methods using the synchronization approach include spectral methods [Sin11, SS11], semidefinite programming [Sin11, SS11, BCSZ14, BCS15], and approximate message passing (AMP) [PWBM16a]. A general Gaussian model for synchronization problems over any compact group is studied in [PWBM16b, PWBM16a]; ideas from statistical physics suggest that for this model, AMP is optimal among all efficient algorithms [PWBM16a]. However, the model does not model an underlying signal θ and instead assumes that for every pair i, j of samples, we observe an independent noisy measurement of the relative group element $g_i g_j^{-1}$. Thus, this model does not correctly capture problems like MRA and cryo-EM.

The synchronization approach has proven to be effective both in theory and practice when the noise is sufficiently small. However, once the noise level is large, no consistent estimation of group elements themselves is possible [ADBS16]. Moreover, it is the high-noise regime that is the practically relevant one for many applications, including cryo-EM, where the presence of large noise is a primary obstruction to current techniques [Sig16]. As a result, recent work has focused on approaches which provably succeed even in the large-noise limit. One striking finding of this line of work is that the sample complexity of the statistical estimation problem increases drastically as the signal-to-noise ratio (SNR) decreases. For instance, for the multi-reference alignment problem, consistent estimation of typical signals requires $\Omega(1/\mathrm{SNR}^3)$ samples [BRW17, APS17], with significantly worse rates for atypical signals. By contrast, when the SNR is larger than some threshold, only $O(1/\mathrm{SNR})$ samples are required. Moreover, in contrast with the $O(1/\mathrm{SNR})$ rate—which would hold even in the absence of a group action—the $\Omega(1/\mathrm{SNR}^3)$ bound obtained in previous works depends on particular properties of the cyclic group. In this work, we significantly extend this prior work by determining the sample complexity of the estimation problem in the high-noise regime for *general* groups.

The leading theoretical framework for the high-noise regime is the *invariant features* approach [BRW17, BBM$^+$17, PWB$^+$17, BBLS17]. This approach has a long history in the signal processing literature [Kam80, Sad89, SG92] and is analogous to the well known "method of moments" in statistics [vdV98]. In brief, the invariant features approach bypasses entirely the problem of estimating the group elements and focuses instead on estimating features of the signal which are preserved by the action of the group. So long as these invariant features uniquely specify the orbit of the original signal, the invariants are sufficient statistics for the problem of recovering the orbit of the original signal. Surprisingly, this simple approach yields optimal dependence on the SNR for the multi-reference alignment problem [BRW17, PWB$^+$17].

In this paper we extend the results of [BRW17] and show that the method of moments yields optimal sample complexity for orbit recovery problems over *any* compact group. Specifically, we show that optimal sample complexity is achieved by an algorithm that estimates the moments from the samples and then solves a polynomial system of equations in order to find a signal θ that would produce such moments. This gives rise to the algebraic question of how many moments suffice to determine the orbit of θ. Using tools from algebraic geometry and invariant theory, we investigate this question for various recovery criteria and obtain sharp results in a number of settings. We also give an algorithm to determine the number of moments required (and thus the sample complexity) for any compact group and action.

We note that ours is an information-theoretic result rather than a computational one because even with knowledge of the optimal sample complexity, estimating the original signal still requires to solving a polynomial system of equations. There are fast heuristic methods to solve these systems in practice (see e.g. [BBLS17]) but we leave for future work the question of analyzing such methods rigorously and exploring whether or not they reach the information-theoretic limits.

1.2 Problem statement

Throughout, we consider a compact (topological) group G acting linearly, continuously, and orthogonally on a finite-dimensional real vector space $V = \mathbb{R}^p$. In other words, G acts on V via a linear representation $\rho : G \to \mathrm{O}(V)$, and ρ itself is a continuous function. Here $\mathrm{O}(V)$ denotes the space of orthogonal $p \times p$ matrices. Let $\mathrm{Haar}(G)$ denote Haar measure (i.e., the uniform distribution) on G. We define the *orbit recovery* problem as follows.

Problem 1.1 (orbit recovery)**.** *Let $V = \mathbb{R}^p$ and let $\theta \in V$ be the unknown signal. Let G be a compact group that acts linearly, continuously, and orthogonally on V. For $i \in [n] = \{1, 2, \dots, n\}$ we observe*

$$y_i = g_i \cdot \theta + \xi_i$$

where $g_i \sim \mathrm{Haar}(G)$ and $\xi_i \sim \mathcal{N}(0, \sigma^2 I_{p\times p})$, all independently. The goal is to estimate θ. Note that we can only hope to recover θ up to action by G; thus we aim to recover the orbit $\{g \cdot \theta \;:\; g \in G\}$ *of θ.*

In practical applications, σ is often known in advance and, when it is not, it can generally be estimated accurately on the basis of the samples. We therefore assume throughout that σ is known and do not pursue the question of its estimation in this work.

Our primary goal is to study the sample complexity of this problem: how must the number of samples n scale with the noise level σ (as $\sigma \to \infty$ with G and V fixed) in order for orbit recovery to be statistically possible. All of our results will actually apply to a generalized orbit recovery problem (Problem 2.4) allowing for *projection* and *heterogeneity* (see Section 1.7).

1.3 Motivating examples

It is helpful to have the following motivating examples in mind:

1. Multi-reference alignment (MRA): Let G be the cyclic group $\mathbb{Z}/p$, acting on $V = \mathbb{R}^p$ via cyclic shifts. Thus we observe many noisy copies of a fixed signal, each with a random cyclic shift applied.

2. S^2 registration: Let $S^2 \subseteq \mathbb{R}^3$ be the unit sphere. Let V be the vector space of functions on S^2 that are *band-limited*, i.e. linear combinations of spherical harmonics up to some fixed degree; let $\theta \in V$ be such a function $S^2 \to \mathbb{R}$. Let $G = SO(3)$, acting on the sphere via 3-dimensional rotation; this induces an action on V via $(g \cdot \theta)(x) = \theta(g^{-1} \cdot x)$. Thus we observe many noisy copies of a fixed function on the sphere, each rotated randomly.

1.4 Method of moments

From the samples we can estimate the following moments:

Definition 1.2 (moment tensor)**.** The *order-d moment tensor* is $T_d(\theta) := \mathbb{E}_g[(g \cdot \theta)^{\otimes d}]$ where $g \sim \mathrm{Haar}(G)$.

We can estimate $T_d(\theta)$ from the samples by computing $\frac{1}{n}\sum_{i=1}^n y_i^{\otimes d}$ plus a correction term to cancel bias from the noise terms (see Section 7.1 for details). The moments $T_d(\theta)$ are related to polynomials that are invariant under the group action:

Definition 1.3 (invariant ring)**.** Let $\mathbf{x} = (x_1, \dots, x_p)$ be indeterminants corresponding to a basis for V. We have an action of G on $\mathbb{R}[\mathbf{x}] := \mathbb{R}[x_1, \dots, x_p]$ given by $(g \cdot f)(\mathbf{x}) = f(g^{-1} \cdot \mathbf{x})$. The *invariant ring* $\mathbb{R}[\mathbf{x}]^G \subseteq \mathbb{R}[\mathbf{x}]$ is the ring consisting of polynomials f that satisfy $g \cdot f = f$ for all $g \in G$. An element of the invariant ring is called an *invariant polynomial* (or simply an *invariant*). Invariant polynomials can be equivalently characterized as polynomials of the form $\mathbb{E}_g[f(g \cdot \mathbf{x})]$ where $f \in \mathbb{R}[\mathbf{x}]$ is any polynomial and $g \sim \mathrm{Haar}(G)$.

The two objects above are equivalent in the following sense. The moment tensor $T_d(\theta)$ contains the same information as the set of evaluations $f(\theta)$ for all $f \in \mathbb{R}[\mathbf{x}]^G$ that are homogeneous of degree d. In particular, for any such polynomial f, $f(\theta)$ is a linear combination of the entries of $T_d(\theta)$.

Consider for now the simple problem of distinguishing between two fixed hypotheses $\theta = \tau_1$ and $\theta = \tau_2$, where τ_1 and τ_2 are two fixed vectors in V. One method is to find an invariant polynomial f for which $f(\tau_1) \neq f(\tau_2)$ and to estimate $f(\theta)$ using the samples. The sample complexity of this procedure depends on the degree of f because we need $O(\sigma^{2d})$ samples to accurately estimate $f(\theta)$. We will prove the following (see Section 3).

Theorem 1.4 (distinguishing upper bound)**.** *Fix $\tau_1, \tau_2 \in V$. If there exists a degree-d invariant polynomial $f \in \mathbb{R}[\mathbf{x}]^G$ with $f(\theta) \neq f(\tau)$ then, using $O(\sigma^{2d})$ samples, it is possible to distinguish between $\theta = \tau_1$ and $\theta = \tau_2$ with type-I and type-II error probabilities each at most* $1/3$*.*

Here, $O(\cdots)$ hides factors that depend on G (and its action on V), τ_1, and τ_2, but not σ; we are most interested in how the sample complexity scales as σ becomes large (with everything else held fixed). The error probability $1/3$ is arbitrary and can be boosted by taking more samples (see Theorem 3.2).

Furthermore, we have a matching lower bound to show that the method of moments is optimal: the sample complexity is driven by the minimum degree of an invariant polynomial that separates τ_1 and τ_2:

Theorem 1.5 (distinguishing lower bound)**.** *Fix $\tau_1, \tau_2 \in V$. Let d^* be the smallest positive integer d for which $T_d(\tau_1) \neq T_d(\tau_2)$. Then $\Omega(\sigma^{2d^*})$ samples are required to distinguish between $\theta = \tau_1$ and $\theta = \tau_2$ with type-I and type-II error probabilities each at most* $1/3$*.*

See Section 3 for more details.

1.5 Recovery

We now address the problem of recovering the signal θ from the samples. Our goal is to recover the orbit of θ, defined as follows.

Definition 1.6. For $\theta_1, \theta_2 \in V$, define an equivalence relation $\overset{G}{\sim}$ by letting $\theta_1 \overset{G}{\sim} \theta_2$ if there exists $g \in G$ such that $g \cdot \theta_1 = \theta_2$.

Definition 1.7. The *orbit* of θ (under the action of G) is the equivalence class of θ under $\overset{G}{\sim}$, i.e. the set $\{g \cdot \theta \ : \ g \in G\}$.

There are two decisions to be made in terms of our recovery criteria:

1. Do we assume that θ is a *generic* signal, or do we allow for a *worst-case* signal? (Here *generic* means that there is a measure-zero set of disallowed signals.)

2. Do we want to output a θ' such that $\theta' \overset{G}{\sim} \theta$ (*unique recovery*), or simply a finite list $\theta_1, \ldots, \theta_s$ of candidates such that one of them satisfies $\theta_i \overset{G}{\sim} \theta$ (*list recovery*)?

The terminology "list recovery" is borrowed from the idea of *list decoding* in the theory of error-correcting codes [Eli57]. By taking all combinations of the two options above, there are four different recovery criteria. We will illustrate by examples in Section 1.6 that the different criteria can have very different behavior.

Our main focus will be on the *generic list recovery* case, as it is algebraically the most tractable to analyze. For the following reasons we also argue that it is perhaps the most practically relevant case. Clearly real-world signals are generic. Also, unique recovery is actually impossible in some practical applications; for instance, in cryo-EM it is impossible to determine the chirality of the molecule. Furthermore, one could hope to use application-specific clues to pick the true signal out from a finite list; for instance, in cryo-EM we might hope that the spurious solutions in our finite list do not look like "reasonable" molecules and can be thrown out.

We need the following definitions to capture the notion of *approximately* recovering the orbit of θ.

Definition 1.8. Let V/G denote the set of orbits of V, that is, the equivalence classes of V modulo the relation $\overset{G}{\sim}$.

Definition 1.9. For $\theta_1, \theta_2 \in V$, let

$$d_G(\theta_1, \theta_2) = \min_{g \in G} \|\theta_1 - g \cdot \theta_2\|_2.$$

This pseudometric induces a metric on the quotient space V/G in the obvious way, so we can write $d_G(\mathfrak{o}_1, \mathfrak{o}_2)$ for $\mathfrak{o}_1, \mathfrak{o}_2 \in V/G$. By slight abuse of notation, we write $d_G(\theta_1, \mathfrak{o}_2)$ for $d_G(\mathfrak{o}_1, \mathfrak{o}_2)$, where $\mathfrak{o}_1$ is the orbit of θ_1.

We are now ready to (informally) state our main result on recovery (see Section 3 for more details).

Theorem 1.10 (recovery upper bound, informal)**.** *Fix θ. If the moments $T_1(\theta), \cdots, T_d(\theta)$ uniquely determine the orbit of θ, then using $O(\sigma^{2d})$ samples, we can produce an estimator $\widehat{\theta}$ such that $d_G(\theta, \widehat{\theta}) \le \varepsilon$ with high probability.*

The analagous result holds for list recovery (see Section 3): if the moments determine a finite number of possibilities for the orbit of θ then we can output a finite list of estimators, one of which is close to the orbit of θ.

Thus, we have reduced to the algebraic question of determining how many moments are necessary to determine the orbit of θ (either uniquely or in the sense of list recovery). In Section 4 we will use tools from algebraic geometry and invariant theory in order to address these algebraic questions.

1.6 Examples

The following examples illustrate that the four different recovery criteria above can be very different in terms of sample complexity.

1. Multi-reference alignment (MRA): Recall that this is the case $G = \mathbb{Z}/p$ acting on $V = \mathbb{R}^p$ via cyclic shifts. It is known [PWB$^+$17] that if θ is generic then unique recovery is possible at degree 3 (i.e. using the moment tensors up to order 3). However, for a worst-case θ, degree p is required (even for list recovery); as shown in [BRW17], there are some very particular pairs of signals that match on moments up to order $p-1$. This illustrates a large gap in difficulty between the generic and worst-case problems.

2. Let the signal be a list of four vectors in $\mathbb{R}^3$: $\theta = (x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_4, y_4, z_4) \in V = \mathbb{R}^{12}$. Let $g \in SO(3)$ act block-diagonally on V, i.e. g is applied separately to each of the four vectors. Thus this can be thought of as the problem of trying to learn the shape of a tetrahedron (whose vertices are the vectors) up to rotation. There are no degree-1 invariants, but at degree-2 we learn the Gram matrix of the four vectors (i.e., their norms and pairwise inner products). This narrows down the number of possible tetrahedra to 2 (since we cannot distinguish a tetrahedron from its reflection across some axis), allowing for list recovery. Unique recovery is not possible until degree 3, when we can use the sign of the determinant
$$\begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix}$$
to distinguish a tetrahedron from its reflection. This shows a gap between list recovery and unique recovery.

1.7 Extensions: projection and heterogeneity

We now consider some extensions to the basic orbit recovery problem (Problem 1.1), motivated by the application of cryo-EM:

1. **Projection**: In cryo-EM, we do not observe a noisy 3-dimensional model of the rotated molecule; we only observe a 2-dimensional projection of it. We will model this projection by a linear map $\Pi : \mathbb{R}^p \to \mathbb{R}^q$ that maps a 3-dimensional model to its 2-dimensional projection. The samples are then given by $y_i = \Pi(g_i \cdot \theta) + \xi_i$ where $\xi_i \sim \mathcal{N}(0, \sigma^2 I_{q \times q})$.

2. **Heterogeneity**: In cryo-EM we observe images of many different copies of the same molecule, each rotated differently. However, if our sample is not pure, we may have a mixture of different molecules and want to recover the structure of all of them. We will model this by taking K different unknown signals $\theta_1, \ldots, \theta_K$ along with positive mixing weights $w_1, \ldots, w_K$ which sum to 1. Each sample takes the form $y_i = g_i \cdot \theta_{k_i} + \xi_i$ where k_i is chosen at random according to the mixing weights.

In the next section we will formally define a generalization of the orbit recovery problem that allows for either (or both) of the above extensions. All of our results will apply to this general case.

2 General problem statement

Our results will apply not only to the basic orbit recovery problem (Problem 1.1) but to a generalization (Problem 2.4 below) that captures the projection and heterogeneity extensions discussed in Section 1.7.

Definition 2.1 (mixing weights)**.** Let $w = (w_1, \ldots, w_K) \in \Delta_K := \{(z_1, \ldots, z_K) \,:\, z_\ell \ge 0 \;\; \forall \ell, \, \sum_{\ell=1}^K z_\ell = 1\}$. Let $k \overset{w}{\sim} [K]$ indicate that k is sampled from $[K] = \{1, \ldots, K\}$ such that $k = \ell$ with probability w_ℓ.

In a heterogeneous problem with K different signals, we can only hope to recover the signals up to permutation. To formalize this, our compound signal will lie in a larger vector space V and we will seek to recover its orbit under a larger group G.

Definition 2.2 (setup for heterogeneity)**.** Let $\tilde{G}$ be a compact group acting linearly, continuously, and orthogonally on $\tilde{V} = \mathbb{R}^p$. Let $V = \tilde{V}^{\oplus K} \oplus \Delta_K$ so that $\theta \in V$ encodes K different signals along with mixing weights: $\theta = (\theta_1, \ldots, \theta_K, w)$. Let S_K denote the symmetric group (permutations on K symbols). We let an element $(g_1, \ldots, g_K, \pi) \in \tilde{G}^K \times S_K$ act on V as follows: first, each g_k acts on the corresponding θ_k and then π permutes the θ_k along with their mixing weights. Thus we have a compact group $G = \tilde{G}^K \rtimes S_K$ acting on V. (G is simply $\tilde{G}^K \times S_K$ as a set, but we write $\rtimes$ because the multiplication structure makes this a *semidirect product.*)

Of course, by taking $K = 1$ we recover the basic setup (without heterogeneity) as a special case.

Remark 2.3. We will often need V to be a vector space on which G acts linearly. To this end we drop the nonnegativity constraint on w and use the embedding $\Delta_K \hookrightarrow \mathbb{R}^{K-1}$ given by $(w_1, \ldots, w_K) \mapsto (w_1 - 1/K, \ldots, w_{K-1} - 1/K)$. In other words we have $K-1$ unknown real-valued parameters $\bar{w}_1, \ldots, \bar{w}_{K-1}$ and we define $\bar{w}_K = -\sum_{k=1}^{K-1} \bar{w}_k$ and $w_k = \bar{w}_k + 1/K$. Note that G acts linearly on $\bar{w}_1, \ldots, \bar{w}_{K-1}$.

We are now ready to give the general problem statement.

Problem 2.4 (generalized orbit recovery)**.** *Let $\tilde{V} = \mathbb{R}^p$ and $W = \mathbb{R}^q$. Let $\tilde{G}$ be a compact group acting linearly, continuously, and orthogonally on $\tilde{V}$. Let $\Pi : \tilde{V} \to W$ be a linear map. Let $\theta = (\theta_1, \ldots, \theta_K, w) \in V := \tilde{V}^{\oplus K} \oplus \Delta_K$ be an unknown collection of K signals with mixing weights. For $i \in [n] = \{1, 2, \ldots, n\}$ we observe*

$$y_i = \Pi(g_i \cdot \theta_{k_i}) + \xi_i$$

where $g_i \sim \mathrm{Haar}(\tilde{G})$, $k_i \overset{w}{\sim} [K]$, $\xi_i \sim \mathcal{N}(0, \sigma^2 I_{q \times q})$, all independently. The goal is to estimate the orbit of θ under $G = \tilde{G}^K \rtimes S_K$.

Note that this serves as a reduction from the heterogeneous setup to the basic setup in the sense that we are still only concerned with recovering the orbit of a vector θ under the action of some compact group.

As discussed previously, we apply the method of moments. The moments are now defined as follows.

Definition 2.5 (moment tensor)**.** For the generalized orbit recovery problem (Problem 2.4), the *order-d moment tensor* is $T_d(\theta) := \mathbb{E}_{g,k}[(\Pi(g \cdot \theta_k))^{\otimes d}]$ where $g \sim \mathrm{Haar}(\tilde{G})$ and $k \overset{w}{\sim} [K]$. Equivalently, $T_d(\theta) = \sum_{k=1}^K w_k \, \mathbb{E}_g[(\Pi(g \cdot \theta_k))^{\otimes d}]$.

The invariant ring is defined as in Definition 1.3 but now for the larger group G acting on the larger V:

Definition 2.6 (invariant ring)**.** Note that $\dim(V) = Kp + K - 1$ and let $\mathbf{x} = (x_1, \ldots, x_{\dim(V)})$ be indeterminants corresponding to a basis of V; here we are thinking of V as a vector space so that the last $K-1$ variables correspond to $\bar{w}_1, \ldots, \bar{w}_{K-1}$ (see Remark 2.3). We then let $\mathbb{R}[\mathbf{x}]^G \subseteq \mathbb{R}[\mathbf{x}]$ be the polynomials in $\mathbf{x}$ that are invariant under the action of G (as in Definition 1.3).

Recall that in the basic orbit recovery problem, $T_d(\theta)$ corresponds precisely to the homogeneous invariant polynomials of degree d; now $T_d(\theta)$ corresponds to a subspace of the homogeneous invariant polynomials of degree d. Specifically, the method of moments gives us access to the following polynomials (evaluated at θ):

Definition 2.7. Let U_d^T be the subspace (over $\mathbb{R}$) of the invariant ring $\mathbb{R}[\mathbf{x}]^G$ consisting of all $\mathbb{R}$-linear combinations of entries of $T_d(x)$. Let $U_{\le d}^G = U_1^G \oplus \cdots \oplus R_d^G \subseteq \mathbb{R}[\mathbf{x}]^G$.

We will be interested in whether the subspace $U_{\le d}^G$ contains enough information to uniquely determine the orbit of θ (or determine a finite list of possible orbits) in the following sense.

Definition 2.8. A subspace $U \subseteq \mathbb{R}[\mathbf{x}]^G$ *resolves* $\theta \in V$ if there is no orbit $\mathfrak{o} \in V/G$ with $\mathfrak{o} \not\ni \theta$ such that for every $f \in U$ we have $f(\theta) = f(\mathfrak{o})$. Similarly, U *list-resolves* θ if there are only finitely many orbits $\mathfrak{o}_1, \dots, \mathfrak{o}_s$ such that $f(\theta) = f(\mathfrak{o}_i)$ for all $f \in U$.

Here we have abused notation by writing $f(\mathfrak{o})$ to denote the (constant) value that f takes on every $\theta \in \mathfrak{o}$. The following question is of central importance.

Question 2.9. *Fix $\theta \in V$. How large must d be in order for $U_{\le d}^T$ to uniquely resolve θ? How large must d be in order for $U_{\le d}^T$ to list resolve θ?*

The answer depends on G and V but also on whether θ is a generic or worst-case signal, and whether we ask for unique recovery or list recovery. Our statistical results in Section 3 will show that the sample complexity of the generalized orbit recovery problem is $\Theta(\sigma^{2d})$ where d is the minimal d from Question 2.9. More specifically, the recovery procedure that obtains this bound is based on estimating the moments $T_1(\theta), \dots, T_d(\theta)$ and solving a system of polynomial equations to (approximately) recover θ. Our algebraic results in Section 4 will give general methods to answer Question 2.9 for any G and V.

3 Statistical results

In this section, we state upper and lower bounds on the performance of optimal estimators for the orbit recovery problem. Proofs are deferred to Section 7.1. Our approach will be the *method of moments* introduced in Section 1.4. We assume for normalization purposes that there exists a constant $c \ge 1$ such that $c^{-1} \le \|\theta\| \le c$, so that σ captures entirely the signal-to-noise ratio of the problem. We denote by Θ the subset of V consisting of vectors satisfying this requirement.

Denote by P_θ the distribution of a sample arising from the generalized orbit recovery problem (Problem 2.4) with parameter θ.

Definition 3.1. Given $\theta \in \Theta$, the *order-d matching set for θ*, $\mathcal{M}_{\theta,d}$, is the set consisting of all $\tau \in V$ such that $f(\tau) = f(\theta)$ for all $f \in U_{\le d}^T$.

We note that $U_{\le d}^T$ resolves θ exactly when $\mathcal{M}_{\theta,d}$ contains a single orbit, and $U_{\le d}^T$ list-resolves θ when $\mathcal{M}_{\theta,d}$ is the union of a finite number of orbits.

We are now ready to state a formal theorem justifying Theorems 1.4 and 1.10, above. The following theorem establishes that we can approximately learn the order-d matching set for θ with probability at least $1 - \delta$ on the basis of $O(\sigma^{2d} \log(1/\delta))$ samples. Denote by $\mathcal{M}_{\theta,d}^\varepsilon$ the *ε-fattening* of $\mathcal{M}_{\theta,d}$, i.e., the set of all $\phi \in \Theta$ such that $\min_{\tau \in \mathcal{M}_{\theta,d}} \|\phi - \tau\| \le \varepsilon$.

Theorem 3.2. *For any positive integer n, noise level $\sigma \ge \max_{\theta \in \Theta} \|\theta\|$, and accuracy parameter $\delta > 0$, there exists an estimator $\widehat{\mathcal{M}}_n = \widehat{\mathcal{M}}_n(y_1, \dots, y_n) \subseteq V$ such that, for any positive constant ε, if $y_1, \dots, y_n \sim \mathrm{P}_\theta$ i.i.d. and $n \ge c_{\theta,\varepsilon,d} \log(1/\delta)\sigma^{2d}$, then with probability at least $1 - \delta$,*

$$\mathcal{M}_{\theta,d} \subseteq \widehat{\mathcal{M}}_n \subseteq \mathcal{M}_{\theta,d}^\varepsilon\,.$$

The constant $c_{\theta,\varepsilon,d}$ in Theorem 3.2 can be replaced by $c_{\theta,d}\varepsilon^{-2}$ in the unique recovery setting if θ is suitably generic, but the dependence on ε can be worse in general. What is key is that $c_{\theta,\varepsilon,d}$ does not depend on σ, so that Theorem 3.2 captures the behavior of the sample complexity in the large σ limit.

Theorem 3.2 essentially follows from the observation that, since the variance of y_i is $O(\sigma^2)$, a degree-d polynomial in the entries of y_i has variance $O(\sigma^{2d})$. This implies that for any $f \in U^T_{\leq d}$, the evaluation $f(\theta)$ can be accurately estimated on the basis of $O(\sigma^{2d})$ samples. By inverting a suitable polynomial system, we can thereby identify $\mathcal{M}_{\theta,d}$, at least approximately. A full proof of Theorem 3.2 appears in Section 7.1.

Theorem 3.2 captures the behavior described in both Theorem 1.4 and Theorem 1.10. Indeed, if τ_1 and τ_2 differ on some $f \in U^T_d$, then $\mathcal{M}_{\tau_1,d}$ is separated from $\mathcal{M}_{\tau_2,d}$, and Theorem 3.2 implies that we can therefore distinguish between the two distributions when ε is sufficiently small. Moreover, as the following corollary shows, Theorem 3.2 implies that the confidence set $\widehat{\mathcal{M}}_n$ allows us to recover or list-recover the orbit of θ.

Corollary 3.3. *Suppose that $\mathcal{M}_{\theta,d}$ is the union of M orbits $\mathfrak{o}_1, \ldots, \mathfrak{o}_M$, where M is finite. There exists an ε_θ such that, for $\varepsilon < \varepsilon_\theta$, if $n \geq c_{\theta,\varepsilon,d}\log(1/\delta)\sigma^{2d}$, then on the basis of n i.i.d. samples from P_θ we can produce M estimators $\widehat{\theta}_1, \ldots \widehat{\theta}_M$ such that, with probability at least $1-\delta$, there exists a permutation $\pi : [M] \to [M]$ satisfying*

$$d_G(\widehat{\theta}_i, \mathfrak{o}_{\pi(i)}) \leq \varepsilon$$

for all $i \in [N]$.

Proof. Since G is a compact group acting continuously on V, the orbits are compact. By assumption $\mathcal{M}_{\theta,d}$, is a union of a finite number of orbits, so there exists an ε_θ such that $d_G(\mathfrak{o}_i, \mathfrak{o}_j) \geq 4\varepsilon_\theta$ for any $i \neq j$. For any $\varepsilon < \varepsilon_\theta$, let $\mathcal{N}$ be an $\varepsilon/2$-net of V/G, and construct $\widehat{\mathcal{M}}_n$ as in Theorem 3.2. Theorem 3.2 implies the existence of a constant $c_{\theta,\varepsilon,d}$ such that as long as $n \geq c_{\theta,\varepsilon,d}\log(1/\delta)\sigma^{2d}$, with probability at least $1-\delta$, $\mathcal{M}_{\theta,d} \subseteq \widehat{\mathcal{M}}_n \subseteq \mathcal{M}^{\varepsilon/2}_{\theta,d}$.

Consider the set $\mathcal{C}$ consisting of $\mathfrak{o} \in \mathcal{N}$ such that $d_G(\mathfrak{o}, \widehat{\mathcal{M}}_n) \leq \varepsilon/2$. With probability $1-\delta$, any element of $\mathcal{C}$ is within ε of $\mathfrak{o}_i$ for some $i \in [M]$, and for each $\mathfrak{o}_i$ there exists an $\mathfrak{o} \in \mathcal{C}$ that is at most ε away. By assumption, distinct orbits in $\mathcal{M}_{\theta,d}$ are separated by more than 4ε, so if any two elements of $\mathcal{C}$ are separated by at most 2ε, then they are close to the same element of $\mathcal{M}_{\theta,d}$. As a result, it is possible to partition $\mathcal{C}$ into M sets $\mathcal{C}_1, \ldots, \mathcal{C}_M$ such that $d_G(\mathfrak{o}, \mathfrak{o}') \leq 2\varepsilon$ if $\mathfrak{o}, \mathfrak{o}'$ are in the same set, and $d_G(\mathfrak{o}, \mathfrak{o}') > 2\varepsilon$ if $\mathfrak{o}$ and $\mathfrak{o}'$ are in different sets. For $i \in [M]$, let $\widehat{\theta}_i$ be any element of V such that the orbit of $\widehat{\theta}_i$ lies in $\mathcal{C}_i$. The claim follows. □

The constant ε_θ in the statement of Corollary 3.3 will not be known in general. However, a weaker statement still holds when $\varepsilon \geq \varepsilon_\theta$. Indeed, consider the image of the set $\widehat{\mathcal{M}}_n$ under the projection $V \mapsto V/G$. There exists a finite partition of the resulting set such that any two orbits in the same cluster are closer than any two orbits in different clusters. (Note that the partition clustering into a single set always satisfies this requirement.) If we are able to choose this partition such that the diameter of each set is at most ε', then by choosing a representative from each cluster, we obtain a finite set of estimators, at least one of which is guaranteed to be ε'-close to θ with high probability. Corollary 3.3 implies that this partition can be taken to consist of at most M clusters for ε' arbitrarily small, as long as $n \geq C_{\varepsilon'}\sigma^{2d}$.

We now prove a lower bound showing that the dependence on σ in Theorem 3.2 is tight. We show that if $U^T_{\leq d-1}$ fails to resolve (or list-resolve) θ, then $\Omega(\sigma^{2d})$ samples are necessary to recover (or list-recover) the orbit of θ. Together with Theorem 3.2, this lower bound implies that if d^* is the smallest positive integer for which $U^T_{d^*}$ resolves (or list-resolve) θ, then $\Theta(\sigma^{2d^*})$ samples are required to recover (or list-recover) the orbit of θ. We make this lower bound precise in Theorem 3.4.

Theorem 3.4. *For any positive integer d, there exists a constant c_d such that if τ_1 and τ_2 are elements in $\mathcal{M}_{\theta,d-1}$ lying in different orbits, then no procedure can distinguish between P_{θ_1} and P_{θ_1} with probability greater than $2/3$ if $n \leq c_d\sigma^{2d}$.*

Note that via Le Cam's method [LeC73], Theorem 3.4 translates into lower bounds for the problem of recovering θ. The proof of Theorem 3.4 relies on a tight bound for the Kullbeck-Leibler divergence between the distributions $\mathrm{P}_{\mathfrak{o}_1}$ and $\mathrm{P}_{\mathfrak{o}_2}$ established in [BRW17]. A proof of Theorem 3.4 appears in Section 7.1

4 Algebraic results

Throughout, we assume the setup defined in Section 2 for the generalized orbit recovery problem. In particular, G is a compact group acting linearly and continuously on a finite-dimensional real vector space V, though we do not require in this section that the action be orthogonal. We have the invariant ring $\mathbb{R}[\mathbf{x}]^G$ corresponding to the action of G on V, and a subspace $U \subseteq \mathbb{R}[\mathbf{x}]^G$ (e.g. $U^T_{\le d}$) of invariants that we have access to. We are interested in whether the values $f(\theta)$ for $f \in U$ determine the orbit of $\theta \in V$ under G. The specific structure of G and $U^T_{\le d}$ (as defined in Section 2) will be largely unimportant and can be abstracted away.

4.1 Invariant theory basics

We will often need the following basic operator that averages a polynomial over the group G.

Definition 4.1 (Reynolds operator)**.** The *Reynolds operator* $\mathcal{R} : \mathbb{R}[\mathbf{x}] \to \mathbb{R}[\mathbf{x}]^G$ is defined by

$$(\mathcal{R}(f))(\mathbf{x}) = \mathop{\mathbb{E}}_{g \sim \mathrm{Haar}(G)} f(g \cdot \mathbf{x}).$$

Note that the Reynolds operator is a linear projection from $\mathbb{R}[\mathbf{x}]$ to $\mathbb{R}[\mathbf{x}]^G$ that preserves the degree of homogeneous polynomials (i.e. a homogeneous polynomial of degree d gets mapped either to a homogeneous polynomial of degree d or to zero).

Observation 4.2. *Let $\mathbb{R}[\mathbf{x}]^G_d$ denote the vector space consisting of homogeneous invariants of degree d. We can obtain a basis for $\mathbb{R}[\mathbf{x}]^G_d$ by applying $\mathcal{R}$ to each monomial in $\mathbb{R}[\mathbf{x}]$ of degree d. (This yields a spanning set which can be pruned to a basis if desired.)*

In our setting, we have the following basic fact from invariant theory.

Theorem 4.3 (e.g. [Kač94] Theorem 4.1-3)**.** *The invariant ring $\mathbb{R}[\mathbf{x}]^G$ is finitely generated as an $\mathbb{R}$-algebra. In other words, there exist generators $f_1, \ldots, f_m \in \mathbb{R}[\mathbf{x}]^G$ such that $\mathbb{R}[f_1, \ldots, f_m] = \mathbb{R}[\mathbf{x}]^G$.*

Furthermore, there is an algorithm to find a generating set; see Section 7.3.1. Another basic fact from invariant theory implies that the entire invariant ring is sufficient to determine the orbit of θ. (This is not always true for non-compact groups; see Example 2.3.1 in [DK15].)

Theorem 4.4 ([Kač94] Theorem 6-2.2)**.** *The full invariant ring $\mathbb{R}[\mathbf{x}]^G$ resolves every $\theta \in V$.*

Proof. Let $\mathfrak{o}_1, \mathfrak{o}_2 \in V/G$ be distinct (and therefore disjoint) orbits. Since G is compact and acts continuously, $\mathfrak{o}_1$ and $\mathfrak{o}_2$ are compact subsets of V. Thus by Urysohn's lemma there exists a continuous function $\tilde{f} : V \to \mathbb{R}$ such that $\tilde{f}(\tau) = 0 \;\; \forall \tau \in \mathfrak{o}_1$ and $\tilde{f}(\tau) = 1 \;\; \forall \tau \in \mathfrak{o}_2$. The Stone–Weierstrass theorem states that a continuous function on a compact domain can be uniformly approximated to arbitrary accuracy by a polynomial. This means there is a polynomial $f \in \mathbb{R}[\mathbf{x}]$ with $f(\tau) \le 1/3 \;\; \forall \tau \in \mathfrak{o}_1$ and $f(\tau) \ge 2/3 \;\; \forall \tau \in \mathfrak{o}_2$. It follows that $h = \mathcal{R}(f)$ is an invariant polynomial that separates the two orbits: $h(\mathfrak{o}_1) \le 1/3$ and $h(\mathfrak{o}_2) \ge 2/3$. □

Thus, in order to determine the orbit of θ it is sufficient to determine the values of all invariant polynomials. (This condition is clearly also necessary in the sense that if the orbit is uniquely determined then so are the values of all invariants.)

Remark 4.5. In what follows we will be discussing algorithms that take the problem setup as input (including $\tilde{G}$ and its action on $\tilde{V}$, along with Π, K) and decide whether or not $U^T_{\le d}$ (see Definition 2.7) is capable of a particular recovery task (e.g. list recovery every $\theta \in V$). We will always assume that these algorithms have a procedure to compute a basis for U^T_d (for any d) in exact symbolic arithmetic. This is non-trivial in some cases because $T_d(\mathbf{x})$ (and thus U^T_d) involves integration over the group, but we will not worry about these details here.

Remark 4.6. We will draw from various references for algorithmic aspects of invariant theory. The case of finite groups is treated by [Stu08]. Although the invariant ring is sometimes taken to be $\mathbb{C}[\mathbf{x}]^G$ instead of $\mathbb{R}[\mathbf{x}]^G$, this is unimportant in our setting because the two are essentially the same: since our group action is real, a basis for $\mathbb{R}[\mathbf{x}]^G$ (over $\mathbb{R}$) is a basis for $\mathbb{C}[\mathbf{x}]^G$ (over $\mathbb{C}$). The case of infinite groups is covered by [DK15]. Here the group is assumed to be a *reductive* group over $\mathbb{C}$ (or another algebraically-closed field). This means in particular that the group is a subset of complex-valued matrices defined by polynomial constraints. Although compact groups such as $SO(3)$ do not satisfy this, the key property of a reductive group is the existence of a Reynolds operator satisfying certain properties; since this exists for compact groups (Definition 4.1), some (but not all) results still hold in our setting.

4.2 Generic list recovery

We will see that the case of list recovery a generic signal is governed by the notion of algebraic independence.

Definition 4.7. Polynomials $f_1, \dots, f_m \in \mathbb{R}[\mathbf{x}]$ are *algebraically dependent* if there exists a nonzero polynomial $P \in \mathbb{R}[y_1, \dots, y_m]$ such that $P(f_1, \dots, f_m) = 0$ (i.e. $P(f_1(\mathbf{x}), \dots, f_m(\mathbf{x}))$ is equal to the zero polynomial). Otherwise, they are *algebraically independent*.

Definition 4.8. The *transcendence degree* of a subspace $U \subseteq \mathbb{R}[\mathbf{x}]$, denoted $\operatorname{trdeg}(U)$ is the maximum value of m for which there exist algebraically independent $f_1, \dots, f_m \in U$. A set of $\operatorname{trdeg}(U)$ such polynomials is called a *transcendence basis* of U.

We now present our algebraic characterization of the generic list recovery problem.

Theorem 4.9 (generic list recovery)**.** *Let $U \subseteq \mathbb{R}[\mathbf{x}]^G$ be a finite-dimensional subspace. If* $\operatorname{trdeg}(U) = \operatorname{trdeg}(\mathbb{R}[\mathbf{x}]^G)$ *then there exists a set $S \subseteq V$ of full measure such that if $\theta \in S$ then U list-resolves θ. Conversely, if* $\operatorname{trdeg}(U) < \operatorname{trdeg}(\mathbb{R}[\mathbf{x}]^G)$ *then there exists a set $S \subseteq V$ of full measure such that if $\theta \in S$ then U does not list resolve θ.*

The proof is deferred to Sections 7.3.2 and 7.3.3. A set has *full measure* if its complement has measure zero. The intuition behind Theorem 4.9 is that $\operatorname{trdeg}(\mathbb{R}[\mathbf{x}]^G)$ is the number of degrees of freedom that need to be pinned down in order to learn the orbit of θ, and so we need this many algebraically independent constraints (invariant polynomials). Note that we have not yet given any bound on how large the finite list might be; we will address this in Section 4.3.

In order for Theorem 4.9 to be useful, we need a way to compute the transcendence degree of both $\mathbb{R}[\mathbf{x}]^G$ and U. In what follows, we will discuss methods for both of these: in Section 4.2.1 we show how to compute $\operatorname{trdeg}(\mathbb{R}[\mathbf{x}]^G)$ analytically, and in Section 4.2.2 we give an efficient algorithm to compute $\operatorname{trdeg}(U)$ for a subspace U. By taking $U = U^T_{\le d}$ this yields an efficient algorithm to determine the smallest degree d at which $U^T_{\le d}$ list-resolves a generic θ (thereby answering Question 2.9 for the case of generic list recovery).

4.2.1 Computing the transcendence degree of $\mathbb{R}[\mathbf{x}]^G$.

Intuitively, the transcendence degree of $\mathbb{R}[\mathbf{x}]^G$ is the number of parameters required to describe an orbit of G. For finite groups, this is simply the dimension of V:

Proposition 4.10 ([Stu08] Proposition 2.1.1)**.** *If G is a finite group,* $\operatorname{trdeg}(\mathbb{R}[\mathbf{x}]^G) = \dim(V)$.

For infinite groups, the intuition is $\operatorname{trdeg}(\mathbb{R}[\mathbf{x}]^G) = p - p'$ where $p = \dim(V)$ and p' is the dimension of a generic orbit. For instance, if $SO(3)$ acts on $V = \mathbb{R}^3$ in the standard way (rotations in 3 dimensions) we have $p' = 2$ because each orbit is a sphere. This means there is only one parameter to learn, namely the 2-norm. On the other hand, if $SO(3)$ acts on (a sufficiently rich class of) functions $S^2 \to \mathbb{R}$ (as in the S^2 registration problem; see Section 5.2) we have $p' = 3$ because each orbit resembles a copy of $SO(3)$ which has dimension 3.

Formally, we can compute the transcendence degree of $\mathbb{R}[\mathbf{x}]^G$ using a central object in invariant theory: the *Hilbert series* (see e.g. [DK15]).

Definition 4.11. Let $\mathbb{R}[\mathbf{x}]^G_d$ be the subspace (over $\mathbb{R}$) of $\mathbb{R}[\mathbf{x}]^G$ consisting of homogeneous invariants of degree d. The *Hilbert series* of $\mathbb{R}[\mathbf{x}]^G$ is the formal power series

$$H(t) := \sum_{d=0}^{\infty} \dim(\mathbb{R}[\mathbf{x}]^G_d)\, t^d.$$

For a given G acting on V, there is an explicit formula (*Molien's formula*) for the Hilbert series:

Proposition 4.12 ([Kač94] Remark 3-1.8)**.** *Let $\rho : G \to \mathrm{GL}(V)$ be the representation by which G acts on V. Then for $|t| < 1$, $H(t)$ converges and we have*

$$H(t) = \mathop{\mathbb{E}}_{g \sim \mathrm{Haar}(G)} \det(I - t\,\rho(g))^{-1}.$$

This formula is tractable to compute, even for complicated groups; see Section 5.2 for details in the case of $SO(3)$. Once we have the Hilbert series, it is easy to extract $\mathrm{trdeg}(\mathbb{R}[\mathbf{x}]^G)$ as follows.

Proposition 4.13. *The order of the pole at $t = 1$ of $H(t)$ is equal to* $\mathrm{trdeg}(\mathbb{R}[\mathbf{x}]^G)$*.*

See Section 7.3.4 for a proof.

For heterogeneous problems ($K > 1$), the transcendence degree can be computed easily from the transcendence degree of the corresponding homogeneous ($K = 1$) problem.

Proposition 4.14. *Let $\tilde{G}$ be a compact group acting linearly and continuously on $\tilde{V}$, and let $G = \tilde{G}^K \rtimes S_K$ act on $V = \tilde{V}^{\oplus K} \oplus \Delta_K$ as in Definition 2.2. Let $\mathbb{R}[\mathbf{x}]^G$ be the invariant ring corresponding to the action of G on V, and let $\mathbb{R}[\tilde{\mathbf{x}}]^{\tilde{G}}$ be the invariant ring corresponding to the action of $\tilde{G}$ on $\tilde{V}$ (i.e. the $K = 1$ problem). Then* $\mathrm{trdeg}(\mathbb{R}[\mathbf{x}]^G) = K \cdot \mathrm{trdeg}(\mathbb{R}[\tilde{\mathbf{x}}]^{\tilde{G}}) + K - 1$.

The proof can be found in Section 7.3.5. Note, however, that the result is intuitively obvious by counting parameters. We know $\mathrm{trdeg}(\mathbb{R}[\tilde{\mathbf{x}}]^{\tilde{G}})$ is the number of parameters required to describe an orbit of $\tilde{G}$ acting on $\tilde{V}$. Thus, in the heterogeneity problem we have $\mathrm{trdeg}(\mathbb{R}[\tilde{\mathbf{x}}]^{\tilde{G}})$ parameters for each of the K signals, plus an additional $K - 1$ parameters for the K mixing weights (since they sum to 1).

4.2.2 Algorithm for transcendence basis of U.

In this section we prove the following.

Theorem 4.15. *There is an efficient algorithm to perform the following task. Given a basis $\{u_1, \ldots, u_s\}$ for a finite-dimensional subspace $U \subseteq \mathbb{R}[\mathbf{x}]$, output a transcendence basis for U.*

The first ingredient we need is the following simple classical test for algebraic independence (see, e.g., [ER93, BMS13] for a proof).

Definition 4.16 (Jacobian)**.** Given polynomials $f_1, \ldots, f_m \in \mathbb{R}[\mathbf{x}] = \mathbb{R}[x_1, \ldots, x_p]$, we define the *Jacobian matrix* $J_{\mathbf{x}}(f_1, \ldots, f_m) \in (\mathbb{R}[\mathbf{x}])^{m \times p}$ by $(J_{\mathbf{x}}(f_1, \ldots, f_m))_{ij} = \partial_{x_j} f_i$ where ∂_{x_j} denotes formal partial derivative with respect to x_j.

Proposition 4.17 (Jacobian criterion for algebraic independence)**.** *Polynomials $\mathbf{f} = (f_1, \ldots, f_m)$ are algebraically independent if and only if the Jacobian matrix $J_{\mathbf{x}}(\mathbf{f})$ has full row rank (over the field $\mathbb{R}(\mathbf{x})$).*

It suffices to test the rank of the Jacobian at a generic point $\mathbf{x}$.

Corollary 4.18. *Fix $\mathbf{f} = (f_1, \ldots, f_m)$. Let $z \sim \mathcal{N}(0, I_{p \times p})$. If $\mathbf{f}$ is algebraically dependent then $J_{\mathbf{x}}(\mathbf{f})|_{\mathbf{x}=z}$ does not have full row rank. If $\mathbf{f}$ is algebraically independent then $J_{\mathbf{x}}(\mathbf{f})|_{\mathbf{x}=z}$ has full row rank with probability 1.*

Proof. An $m \times p$ matrix has deficient row rank if and only if either $m > p$ or every maximal square submatrix has determinant zero. Every such determinant of $J_{\mathbf{x}}(\mathbf{f})$ is a polynomial in $\mathbf{x}$; if this polynomial is not identically zero then plugging in generic values for $\mathbf{x}$ will not cause it to vanish. □

In practice we may choose to plug in random *rational* values for $\mathbf{x}$ so that the rank computation can be done in exact symbolic arithmetic. The Jacobian test will still succeed with overwhelming probability (provided we use a fine enough mesh of rational numbers). Also note that if we find *any* value of $\mathbf{x}$ for which the Jacobian has full row rank, this constitutes a proof of algebraic independence.

Curiously, although the above gives an efficient test for algebraic dependence, it is much harder ($\#P$-hard) to actually find the algebraic dependence (i.e., the polynomial relation) when one exists [Kay09].

It is well-known that subsets of $\mathbb{R}[\mathbf{x}]$ that are algebraically independent form a *matroid*; this is called an *algebraic matroid* (see e.g. [Sch03]). In particular, we have the following exchange property:

Proposition 4.19. *Let I, J be finite subsets of $\mathbb{R}[\mathbf{x}]$, each algebraically independent. If $|I| < |J|$ then there exists $f \in J \smallsetminus I$ such that $I \cup \{f\}$ is algebraically independent.*

We next note that in the task from Theorem 4.15, a transcendence basis can always be taken from the basis $\{u_1, \dots, u_s\}$ itself.

Lemma 4.20. *Let U be a finite-dimensional subspace of $\mathbb{R}[\mathbf{x}]$ with basis $B = \{u_1, \dots, u_s\}$. If U contains r algebraically independent elements, then so does B.*

Proof. Let $B' \subseteq B$ be a maximal set of algebraically independent elements of B. If $|B| < r$ then by the exchange property (Proposition 4.19) there exists $v \in U \smallsetminus B'$ such that $B' \cup \{v\}$ is algebraically independent. Write $v = \sum_{i=1}^{s} \alpha_i u_i$. Since B' is maximal, we have from the Jacobian criterion (Proposition 4.17) that for all $1 \le i \le s$, the row vector $J_{\mathbf{x}}(u_i)$ lies in the $\mathbb{R}(\mathbf{x})$-span of $\mathcal{B} := \{J_{\mathbf{x}}(b)\}_{b \in B'}$. But this means that $J_{\mathbf{x}}(v) = \sum_{i=1}^{s} \alpha_i J_{\mathbf{x}}(u_i)$ lies in the $\mathbb{R}(\mathbf{x})$-span of $\mathcal{B}$. By the Jacobian criterion this contradicts the fact that $B' \cup \{v\}$ is algebraically independent. □

The algebraic matroid property implies that a simple greedy algorithm suffices to prove Theorem 4.15.

Proof of Theorem 4.15.
We employ the following algorithm. As input, receive a list of polynomials $\{u_1, \dots, u_s\}$. Initialize $I = \emptyset$. For $i = 1, \dots, s$, add $\{u_i\}$ to I if $I \cup \{u_i\}$ is algebraically independent, and do nothing otherwise. (Note that this condition can be efficiently tested by Corollary 4.18). Output the resulting set I.

We now show correctness. Let I_i be the set after item u_i has been considered (and possibly added), and set $I_0 = \emptyset$. It suffices to show that for each $i \in \{0, \dots, s\}$, I_i is a maximal independent subset of $\{u_1, \dots, u_i\}$. We proceed by induction. The claim is vacuously true when $i = 0$. Assume it holds for $i - 1$. If I_i is not a maximal independent subset of $\{u_1, \dots, u_i\}$, then there exists an independent set $J \subseteq \{u_1, \dots, u_i\}$ with $|J| > |I|$, so by the exchange property (Proposition 4.19) there exists a u_j with $j \le i$ such that $u_j \notin I_i$ and $I_i \cup \{u_j\}$ is independent. In particular, the subset $I_{j-1} \cup \{u_j\}$ of $I_i \cup \{u_j\}$ is independent. But the fact that u_j was not added at the $(j-1)$th step implies that $I_{j-1} \cup \{u_j\}$ is not independent, a contradiction. So I_i is indeed maximal.

We obtain that $I = I_s$ is a maximal independent subset of $\{u_1, \dots, u_s\}$, and hence by Lemma 4.20 a transcendence basis of U. □

If one wants only the transcendence degree of U (and not a transcendence basis), an alternative algorithm is to simply take the row rank of $J_{\mathbf{x}}(u_1, \dots, u_s)$ where $\{u_1, \dots, u_s\}$ is a basis (or spanning set) for U.

4.3 Generic unique recovery

For list recovery problems, the following gives an explicit upper bound on the size of the list.

Theorem 4.21. *Let U be a subspace of the invariant ring $\mathbb{R}[\mathbf{x}]^G$. Let F_G be the field of fractions of $\mathbb{R}[\mathbf{x}]^G$. If $[F_G : \mathbb{R}(U)] = D < \infty$ then there exists a set $S \subseteq V$ of full measure such that for any $\theta \in S$, U list-resolves θ with a list of size $\le D$.*

The proof is deferred to Section 7.3.2. Here $\mathbb{R}(U)$ is the smallest subfield of F_G containing both $\mathbb{R}$ and U, and $[F_G : \mathbb{R}(U)]$ denotes the degree of a field extension; see Section 7.3.2 for more details. Since $[F_G : F_U] = 1$ is equivalent to $\mathbb{R}(U) = F_G$, we have the following criterion for unique recovery.

Corollary 4.22 (generic unique recovery)**.** *If $\mathbb{R}(U) = F_G$ then there exists a set $S \subseteq V$ of full measure such that if $\theta \in S$ then U resolves θ.*

The intuition here is that we want to be able to learn every invariant polynomial by adding, multiplying, and dividing polynomials from U (and scalars from $\mathbb{R}$). We need θ to be generic so that we never divide by zero in the process.

Theorem 4.23. *For a finite-dimensional subspace $U \subseteq \mathbb{R}[\mathbf{x}]^G$, there is an algorithm to compute the degree of the field extension from Theorem 4.21. As input, the algorithm requires a basis for U and the ability to compute the Reynolds operator (Definition 4.1).*

We give the algorithm and the proof in Section 7.3.6. The algorithm uses Gröbner bases and is unfortunately inefficient to run in practice.

4.4 Worst-case unique recovery

We give a sufficient algebraic condition for worst-case unique recovery:

Theorem 4.24 (worst-case unique recovery)**.** *Let $U \subseteq \mathbb{R}[\mathbf{x}]^G$ be a finite-dimensional subspace with basis $\{f_1, \ldots, f_m\}$. If U generates $\mathbb{R}[\mathbf{x}]^G$ as an $\mathbb{R}$-algebra (i.e. $\mathbb{R}[f_1, \ldots, f_m] = \mathbb{R}[\mathbf{x}]^G$) then U resolves every $\theta \in V$.*

Proof. Every element of $\mathbb{R}[\mathbf{x}]^G$ can be written as a polynomial in the f_i (with coefficients in $\mathbb{R}$). This means the values $f_1(\theta), \ldots, f_m(\theta)$ uniquely determine all the values $f(\theta)$ for $f \in \mathbb{R}[\mathbf{x}]^G$ and so the result follows because $\mathbb{R}[\mathbf{x}]^G$ resolves every $\theta \in V$ (Theorem 4.4). □

Theorem 4.25. *There is an algorithm to test whether or not U generates $\mathbb{R}[\mathbf{x}]^G$ as n $\mathbb{R}$-algebra. As input, the algorithm requires a basis for U and the ability to compute the Reynolds operator (Definition 4.1).*

We give the algorithm and the proof in Section 7.3.6. The algorithm uses Gröbner bases and is unfortunately inefficient to run in practice.

If G is a finite group, it is known that $\mathbb{R}[\mathbf{x}]^G$ has a generating set for which all elements have degree $\le |G|$ (this is *Noether's degree bound*; see Theorem 2.1.4 in [Stu08]). It follows that $\mathbb{R}[\mathbf{x}]^G_{\le |G|}$ resolves every $\theta \in V$. Recall (from Section 1.6) that this is tight for MRA: degree $|G|$ is necessary for worst-case recovery.

A precise characterization of when U resolves every $\theta \in V$ is (by definition) that U should be a *separating set* (see [DK15] Section 2.4). The notions of generating and separating sets do not always coincide, as illustrated by Example 2.4.2 in [DK15]. However, we are not aware of an example where a generating set requires higher degree than a separating set.

4.5 Worst-case list recovery

We give a sufficient algebraic condition for worst-case list recovery:

Theorem 4.26 (worst-case list recovery)**.** *Let $U \subseteq \mathbb{R}[\mathbf{x}]^G$ be a finite-dimensional subspace with basis $\{f_1, \ldots, f_m\}$. If $\mathbb{R}[\mathbf{x}]^G$ is finitely generated as a $\mathbb{R}[f_1, \ldots, f_m]$-module, then U list-resolves every $\theta \in V$.*

In other words, this condition says that there exists a basis $g_1, \ldots, g_s \in \mathbb{R}[\mathbf{x}]^G$ such that every element of $\mathbb{R}[\mathbf{x}]^G$ can be written as a linear combination of $g_1, \ldots, g_s$ with coefficients from $\mathbb{R}[f_1, \ldots, f_m]$. It is sufficient to take U to be a set of *primary invariants* from a *Hironaka decomposition* (see Section 7.3.4).

Proof. Since $\mathbb{R}[\mathbf{x}]^G$ finitely generated as an $\mathbb{R}$-algebra (Theorem 4.3), if $\mathbb{R}[\mathbf{x}]^G$ is finitely generated as a $\mathbb{R}[f_1, \ldots, f_m]$-module then it follows that (see [Sha94] Section 5.3) every $h \in \mathbb{R}[\mathbf{x}]^G$ satisfies a monic polynomial

$$h^k + c_{k-1}h^{k-1} + \cdots + c_1 h + c_0 = 0$$

with $c_i \in \mathbb{R}[f_1, \ldots, f_m]$. Letting $h_1, \ldots, h_s$ be generators for $\mathbb{R}[\mathbf{x}]^G$ (as an $\mathbb{R}$-algebra), we have that the values $f_1(\theta), \ldots, f_m(\theta)$ determine a finite set of possible values for $h_1(\theta), \ldots, h_s(\theta)$, each of which determines (at most) one orbit for θ. □

5 Examples

In this section we work out some specific examples, determining the degree at which generic list recovery is possible using the methods of Section 4.2. (We focus on generic list recovery because our algorithms for the other recovery criteria are unfortunately too slow even for quite small examples.)

The results we give in this section are stated as *conjectures* rather than theorems for a few reasons. First, each setting we consider is really an infinite sequence of larger and larger problems, and we can only run our algorithm for a finite number of these. A clear pattern will emerge, and we conjecture that it continues indefinitely, but this is not a rigorous proof. (One could potentially prove this by analytically analyzing the rank of the associated Jacobian, but we do not attempt to do this here.) Another reason that our conjectures are non-rigorous is that for $SO(3)$ (in Section 5.2) the Jacobian that we want to show is non-singular has irrational values and so it is slow to compute the rank symbolically; we thus compute the smallest singular value numerically and check that it is reasonably far from zero. Despite all of the above, we are quite confident that the conjectures presented in this section are correct.

The following themes emerge in the examples studied in this section. First, we see that many problems are possible at degree 3, which is promising from a practical standpoint. Second, we do not encounter any unexpected algebraic dependencies, and so we are able to show that heuristic parameter-counting arguments are correct.

5.1 Multi-reference alignment (MRA)

Recall that this is the case of $G = \mathbb{Z}/p$ acting on $V = \mathbb{R}^p$ via cyclic shifts. It is already known that for the basic MRA problem (without projection or heterogeneity), generic unique recovery is possible at degree 3 for any p [BRW17]. The algorithm of Section 4.2 confirms that generic *list* recovery is possible at degree 3 (at least for the values of p that we tested).

We can also prove that for $p \geq 3$, generic list recovery is impossible at degree 2. This follows from Theorem 4.9 because $\operatorname{trdeg}(\mathbb{R}[\mathbf{x}]^G) = p$ (since G is finite) but the number of algebraically independent invariants of degree ≤ 2 is at most $\lfloor n/2 \rfloor + 1$. We can see this as follows. A basis for the invariants of degree ≤ 2 is $\{\mathcal{R}(x_1), \mathcal{R}(x_1^2), \mathcal{R}(x_1x_2), \mathcal{R}(x_1x_3), \ldots, \mathcal{R}(x_1x_s)\}$ with $s = \lfloor n/2 \rfloor + 1$. Here $\mathcal{R}$ denotes the Reynolds operator, which averages over cyclic shifts of the variables. For instance, $\mathcal{R}(x_1x_2) = \frac{1}{p}(x_1x_2 + x_2x_3 + x_3x_4 + \cdots + x_px_1)$. Note that the basis above has size $\lfloor n/2 \rfloor + 2$ but there is an algebraic dependence within it because $\mathcal{R}(x_1)^2$ can be written in terms of the other basis elements. The claim now follows.

Generic list recovery is possible at degree 1 for $p = 1$ and at degree 2 for $p = 2$. (This is true even for worst-case unique recovery; recall from Section 4.4 that degree $|G|$ is always sufficient for this.)

We now move on to variants of the MRA problem that capture features of the cryo-EM problem.

5.1.1 MRA with projection

We now consider MRA with a projection step. We imagine that the coordinates of the signal are arranged in a circle so that G acts by rotating the signal around the circle. We then observe a projection of the circle onto a line so that each observation is the sum of the two entries lying “above” it on the circle. We formally define the setup as follows.

Problem 5.1 (MRA with projection)**.** *Let $p \geq 3$ be odd. Let $V = \mathbb{R}^p$ and $G = \mathbb{Z}/p$ acting on V via cyclic shifts. Let $q = (p-1)/2$ and $W = \mathbb{R}^q$. Let $\Pi : V \to W$ be defined by*

$$\Pi(v_1, \ldots, v_p) = (v_1 + v_p, v_2 + v_{p-1}, \ldots, v_{(p-1)/2} + v_{(p+3)/2}).$$

We call the associated generalized orbit recovery problem (Problem 2.4) MRA with projection*. (We consider the homogeneous case $K = 1$.)*

Note that since p is odd, there is one entry $v_{(p+1)/2}$ which is discarded by Π. The reason we consider the odd-p case rather than the seemingly more elegant even-p case is because generic list recovery is actually impossible in the even-p case. This is because the signals θ and $\theta + (c, -c, c, -c, \ldots)$ cannot be distinguished from the samples, even if there is no noise.

Note that we cannot hope for unique list recovery because it is impossible to tell whether the signal is wrapped clockwise or counterclockwise around the circle. In other words, reversing the signal via $(\theta_1, \ldots, \theta_p) \mapsto (\theta_p, \ldots, \theta_1)$ does not change the distribution of samples. We can still hope for generic list recovery, hopefully with a list of size exactly 2. This degeneracy is analagous to the chirality issue in cryo-EM: it is impossible to determine the chirality of the molecule (i.e. if the molecule is reflected about some 2-dimensional plane, this does not change the distribution of samples).

It appears that, as in the basic MRA problem, generic list recovery is possible at degree 3. We proved this for p up to 21 by running the algorithm of Section 4.2 on a computer, and we conjecture that this trend continues.

Conjecture 5.2. *For MRA with projection, for any odd $p \geq 3$, generic list recovery is possible at degree 3.*

Note that generic list recovery is impossible at degree 2 because the addition of the projection step to basic MRA can only make it harder for $U^T_{\leq d}$ to list-resolve θ.

5.1.2 Heterogeneous MRA

We now consider heterogeneous MRA, i.e. the generalized orbit recovery problem (Problem 2.4) with $\tilde{G} = \mathbb{Z}/p$ acting on $\tilde{V} = \mathbb{R}^p$ via cyclic shifts, $K \geq 2$ heterogeneous components, and no projection (i.e., Π is the identity).

We will see that generic list recovery is possible at degree 3 provided that p is large enough compared to K. First note that the number of degrees of freedom to be recovered is $\mathrm{trdeg}(\mathbb{R}[\mathbf{x}]^G) = Kp + K - 1$ (see Propositions 4.10 and 4.14). Let us now count the number of distinct entries of $T_d(\mathbf{x})$ for $d \leq 3$. Note that $T_d(\mathbf{x})$ is symmetric (under permutations of indices) but we also have additional symmetries given by cyclic shifts, e.g. $(T_3(\mathbf{x}))_{i,j,k} = (T_3(\mathbf{x}))_{i+c,j+c,k+c}$ where c is an integer and the sums $i+c, j+c, k+c$ are computed modulo p. One can compute that $T_1(\mathbf{x})$ has 1 distinct entry, $T_2(\mathbf{x})$ has $\lfloor p/2 \rfloor + 1$ distinct entries, and $T_3(\mathbf{x})$ has $p + \lceil (p-1)(p-2)/6 \rceil$ distinct entries. The total number of distinct entries is

$$\mathcal{U} := p + 2 + \lfloor p/2 \rfloor + \lceil (p-1)(p-2)/6 \rceil.$$

By Theorem 4.9, list recovery is impossible when $\mathcal{U} < Kp + K - 1$. By running the algorithm from Section 4.2 we observe that the converse also appears to hold. (We tested this up to $K = 15$.)

Conjecture 5.3. *For heterogeneous MRA, generic list recovery is possible at degree 3 precisely if $\mathcal{U} \geq Kp + K - 1$. This condition on $\mathcal{U}$ can be stated more explicitly as follows:*

- *$K = 2$ requires $p \geq 1$.*
- *$K = 3$ requires $p \geq 12$.*
- *$K = 4$ requires $p \geq 18$.*
- *Each $K \geq 5$ requires $p \geq 6K - 5$.*

Recent work [BBLS17] also studies the heterogeneous MRA problem. Similarly to the present work, they apply the method of moments and solve a polynomial system of equations in order to recover the signal. To solve the system they use an efficient heuristic method that has no provable guarantees but appears to work well in practice. Their experiments suggest that this method succeeds only when (roughly) $K \le \sqrt{p}$ instead of the condition (roughly) $K \le p/6$ that we see above (and that [BBLS17] also identified based on parameter-counting). Exploring this discrepancy is an interesting direction for future work. One question of particular interest is whether this example evinces a statistical-computational gap, whereby polynomial-time methods fail to succeed once K exceeds $\sqrt{p}$.

5.2 S^2 registration

Let $G = SO(3)$. For each $\ell = 0, 1, 2, \ldots$ there is an irreducible representation V_ℓ of $SO(3)$ of dimension $2\ell+1$. These representations are of real type, i.e. they can be defined over the real numbers so that $V_\ell = \mathbb{R}^{2\ell+1}$. Let $\mathcal{F}$ be a finite subset of $\{0, 1, 2, \ldots\}$ and consider the orbit recovery problem in which G acts on $V = \oplus_{\ell \in \mathcal{F}} V_\ell$.

As intuition for the above setup, V_ℓ is a basis for the degree-ℓ *spherical harmonic* functions $S^2 \to \mathbb{R}$ defined on the surface of the unit sphere $S^2 \subseteq \mathbb{R}^3$. The spherical harmonics are a complete set of orthogonal functions on the sphere and can be used (like a "Fourier series") to represent any function $S^2 \to \mathbb{R}$. Thus the signal $\theta \in V$ can be thought of as a function on the sphere, with $SO(3)$ acting on it by rotating the sphere.

The primary case of interest is $\mathcal{F} = \{1, \ldots, F\}$ for some F (the number of "frequencies"). We will see that generic list decoding is possible at degree 3 so long as $F \ge 10$. Note that we do not include $0 \in \mathcal{F}$, but we now justify why this is without loss of generality. V_0 is the trivial representation, i.e. the 1-dimensional representation on which every group element acts as the identity. In the interpretation of spherical harmonics, the V_0-component is the mean value of the function over the sphere. We claim that the S^2 registration problem with $0 \in \mathcal{F}$ can be easily reduced to the problem with $\mathcal{F}' = \mathcal{F} \smallsetminus \{0\}$. This is because the V_0-component is itself a degree-1 invariant; given the value of this invariant, one can subtract it off and reduce to the case without a V_0-component (i.e. the case where the function on the sphere is zero-mean). Thus we have that e.g. generic list recovery is possible (at a given degree) for $\mathcal{F}$ if and only if it is possible for $\mathcal{F}'$.

Using the methods in Section 4.6 of [DK15], we can give a formula for the Hilbert series of $\mathbb{R}[\mathbf{x}]^G$; see Section 7.4.1 for details of the derivation.)

Proposition 5.4. *Consider S^2 registration with frequencies $\mathcal{F}$. For $|t| < 1$, the Hilbert series of $\mathbb{R}[\mathbf{x}]^G$ is given by*

$$H(t) = \sum_{z \in \mathcal{P}} \operatorname{Res}(f, z)$$

where

$$f(z) = \frac{1 - \frac{1}{2}(z + 1/z)}{z \prod_{\ell \in \mathcal{F}} \prod_{m=-\ell}^{\ell} (1 - tz^m)} = \frac{-z^{N-2}(1-z)^2}{2 \prod_{\ell \in \mathcal{F}} \left[\prod_{m=1}^{\ell} (z^m - t) \prod_{m=0}^{\ell} (1 - tz^m) \right]}$$

with $N = \frac{1}{2} \sum_{\ell \in \mathcal{F}} \ell(\ell+1)$. Here $\operatorname{Res}(f, z)$ denotes the residue (from complex analysis) of the function f at the point z, and $\mathcal{P}$ is the set of poles of $f(z)$ inside the unit circle (in $\mathbb{C}$). Namely, $\mathcal{P}$ contains $t^{1/m} e^{2\pi \mathbf{i} k/m}$ for all $m \in \{1, 2, \ldots, \max_{\ell \in \mathcal{F}} \ell\}$ and for all $k \in \{0, 1, \ldots, m-1\}$. If $N \le 1$, $\mathcal{P}$ also contains 0.

Using the Hilbert series we can extract $\operatorname{trdeg}(\mathbb{R}[\mathbf{x}]^G)$ via Proposition 4.13. We observe that $\operatorname{trdeg}(\mathbb{R}[\mathbf{x}]^G) = p - p'$ where

$$p = \dim(V) = \sum_{\ell \in \mathcal{F}} (2\ell + 1) \tag{1}$$

and

$$p' = \begin{cases} 0 & \ell_{\max} = 0 \\ 2 & \ell_{\max} = 1 \\ 3 & \ell_{\max} \ge 2 \end{cases} \qquad \text{where } \ell_{\max} = \max_{\ell \in \mathcal{F}} \ell. \tag{2}$$

Note that the above is consistent with the intuitive meaning of $\operatorname{trdeg}(\mathbb{R}[\mathbf{x}]^G)$ as the number of parameters that need to be learned in order to determine the orbit of $\theta \in V$. The meaning of p' is the dimension of a generic orbit. V_1 is (isomorphic to) the standard 3-dimension representation of $SO(3)$ whereby $SO(3)$ acts by rotating a vector in 3-dimensional space; thus if $\ell_{\max} = 1$, each orbit resembles a sphere and has dimension 2. When $\ell_{\max} \geq 2$, each orbit resembles a copy of $SO(3)$ and has dimension 3.

If one wants to extract a specific coefficient $\dim(\mathbb{R}[\mathbf{x}]^G_d)$ of the Hilbert series, we give an alternative (and somewhat simpler) formula:

Proposition 5.5. *Consider S^2 registration with frequencies $\mathcal{F}$. Let $\chi_d(\phi) : \mathbb{R} \to \mathbb{R}$ be defined recursively by*

$$\chi_0(\phi) = 1,$$

$$\chi_1(\phi) = \sum_{\ell \in \mathcal{F}} \left[1 + 2 \sum_{m=1}^{\ell} \cos(m\,\phi) \right], \text{ and}$$

$$\chi_d(\phi) = \frac{1}{d} \sum_{i=1}^{d} \chi_1(i\phi)\chi_{d-i}(\phi).$$

Then we have

$$\dim(\mathbb{R}[\mathbf{x}]^G_d) = \frac{1}{\pi} \int_0^\pi (1 - \cos\phi)\chi_d(\phi)\,\mathrm{d}\phi.$$

In the following we restrict to the case $0 \notin \mathcal{F}$ for simplicity, but recall that this is without loss of generality. There are no degree-1 invariants, i.e. $\mathbb{R}[\mathbf{x}]^G_1$ is empty. By Theorem 4.9, if $\dim(\mathbb{R}[\mathbf{x}]^G_2) + \dim(\mathbb{R}[\mathbf{x}]^G_3) < \operatorname{trdeg}(\mathbb{R}[\mathbf{x}]^G)$ then generic list recovery is impossible at degree 3. We observe (by running the algorithm of Section 4.2) that the converse also appears to hold.

Conjecture 5.6. *Consider the S^2 registration problem with $0 \notin \mathcal{F}$ (without loss of generality). We have the following.*

- $\operatorname{trdeg}(\mathbb{R}[\mathbf{x}]^G) = p - p'$ *with p, p' defined by (1) and (2).*
- *Generic list decoding is possible at degree 3 if and only if* $\dim(\mathbb{R}[\mathbf{x}]^G_2) + \dim(\mathbb{R}[\mathbf{x}]^G_3) \geq \operatorname{trdeg}(\mathbb{R}[\mathbf{x}]^G)$.
- *In particular, if $\mathcal{F} = \{1, 2, \ldots, F\}$ then generic list decoding is possible at degree 3 if and only if $F \geq 10$.*

6 Open questions

We leave the following as directions for future work.

1. Our methods require testing the rank of the Jacobian on a computer for each problem size. It would be desirable to have analytic results for e.g. (variants of) MRA in any dimension p.

2. We have given an efficient test for whether generic *list* recovery is possible, but have not given a similarly efficient algorithm for generic *unique* recovery. In cases where unique recovery is impossible, it would be nice to give a tight bound on the size of the list; for instance, for MRA with projection, we conjecture that the list has size exactly 2 (due to "chirality"), but we lack a proof for this fact. Our algorithms are based on Gröbner bases, the calculation of which is known to be computationally hard in the worst case [Huỳ86]. Unfortunately, the algorithms we have proposed are also extremely slow in practice, though a faster implementation may be possible.

3. Our procedure for recovering θ from the samples involves solving a polynomial system of equations. While solving polynomial systems is NP-hard in general, the fact that the polynomials used in the orbit recovery problem have special structure leaves open the possibility of finding an efficient (polynomial time) method with rigorous guarantees. Heuristic methods that work well in practice have been investigated by [BBLS17].

4. We have addressed the statistical limits of orbit recovery problems. However, prior work has indicated the possible presence of statistical-to-computational gaps in related problems [PWBM16a], and we conjecture the existence of such gaps for general orbit recovery problems. As discussed in Section 5.1.2, the results of [BBLS17] suggest a possible gap of this kind for heterogeneous MRA.

7 Proofs

7.1 Proofs for Section 3: statistical results

We first prove Theorem 3.2. This theorem in fact holds for more general mixture problems, not merely those arising from the orbit recovery problems defined in Problem 2.4. For convenience, we will state and prove the theorem in its general form.

Problem 7.1 (subgaussian mixture recovery)**.** *Let $V = \mathbb{R}^p$, and let $\Theta \subset V$ be compact. For $\theta \in \Theta$, let μ_θ be a measure on $\mathbb{R}^p$ whose support is contained in the unit ball, and assume the map $\theta \mapsto \mu_\theta$ is continuous with respect to the weak topology. Let $\mathcal{D}$ be a known* 1*-subgaussian distribution on $\mathbb{R}$ for some $\sigma \geq 1$. For $i \in [n] = \{1, 2, \ldots, n\}$, we observe*

$$y_i = x_i + \sigma \xi_i \,,$$

where $x_i \sim \mu_\theta$ and the entries of ξ_i are independently drawn from $\mathcal{D}$. The goal is to estimate θ.

Write P_θ for the distribution arising from the parameter θ, and let $\mathbb{E}_\theta$ be expectation with respect to this distribution. We denote by $\mathbb{E}_\theta^n$ the expectation taken with respect to n i.i.d. samples from P_θ. Where there is no confusion, we also write $\mathbb{E}_\theta$ for expectation with respect to the distribution μ_θ.

We require that μ_θ have bounded support; the requirement that it be supported in the unit ball is for normalization purposes only. We assume throughout that $\sigma \geq 1$. The following definiton gives the generalization of Definition 3.1 to the subgaussian mixture recovery problem.

Definition 7.2. Given a positive integer d and $\theta \in V$, the *order-d matching set $\mathcal{M}_{\theta,d}$* is the set consisting of all $\phi \in V$ such $\mathbb{E}_\theta[x^{\otimes k}] = \mathbb{E}_\phi[x^{\otimes k}]$ for $k = 1, \ldots, d$, where $\mathbb{E}_\zeta$ represents expectation with respect to $x \sim \mu_\zeta$.

Problem 7.1 generalizes Problem 2.4 by allowing the random vector x_i to arise from more general mixtures than those arising from group actions. Note that when the mixtures do arise from a generalized orbit recovery problem, i.e., when $x_i = \Pi(g_i \cdot \theta_{k_i})$, where g_i and k_i are distributed as in Problem 2.4, then Definition 7.2 reduces to Definition 3.1.

Having made these definitions, our goal in this section is to show that Theorem 3.2 holds word-for-word in the setting of subgaussian mixture recovery. To do so, we show that entries of the moment tensors $\mathbb{E}_\theta[x^{\otimes k}]$ can be estimated on the basis of $O(\sigma^{2k})$ samples from P_θ .

7.1.1 Estimation of moments

Our estimators will be based on a system of orthogonal univariate polynomials under the measure corresponding to $\mathcal{D}$. Let $H_0(x) = 1$, and for $k \geq 1$, define

$$H_k(x) = x^k - \sum_{j=0}^{k-1} \frac{\mathbb{E}_{\xi\sim\mathcal{D}}[\xi^k H_j(\xi)]}{\mathbb{E}_{\xi\sim\mathcal{D}}[H_j(\xi)^2]} H_j(x) \,.$$

It is easy to check that these polynomials are orthogonal under the inner product given by $\langle f, g\rangle = \mathbb{E}_{\xi\sim\mathcal{D}}[f(\xi)g(\xi)]$ and that the polynomials $H_0, \ldots, H_k$ form a basis for the space of polynomials of degree at most k. We denote them by H_k because they coincide with the classic Hermite polynomials when $\mathcal{D}$ is Gaussian.

Like the Hermite polynomials, they satisfy the identity

$$\mathbb{E}_{\xi\sim\mathcal{D}}[H_k(x + \xi)] = x^k \,. \tag{3}$$

Indeed, we can expand $H_k(x+\xi)$ in the basis of the orthogonal polynomials as

$$H_k(x+\xi) = \sum_{j=0}^{k} \alpha_j(x) H_j(\xi)\,,$$

where $\alpha_j(x)$ is a polynomial in x of degree at most $k-j$. Since $H_j(\xi)$ has zero mean for $j \geq 1$ by construction, we obtain

$$\mathbb{E}[H_k(x+\xi)] = \alpha_0(x)\,,$$

and since H_k is a monic polynomial of degree k, we must have $\alpha_0(x) = x^k$.

We briefly review multi-index notation.

Definition 7.3. A p-dimensional *multi-index* is a tuple $\alpha = (\alpha_1, \dots, \alpha_p)$ of nonnegative integers. For $x \in \mathbb{R}^p$, let $x^\alpha = \prod_{j=1}^p x_j^{\alpha_j}$.

For any multi-index α, we write $|\alpha| = \sum_{j=1}^p \alpha_j$. Given independent samples $y_1, \dots, y_n$ from P_θ, consider the estimate for $\mathbb{E}_{x \sim \mu_\theta}[x^\alpha]$ given by

$$\widetilde{x^\alpha} := \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{p} \sigma^{\alpha_j} H_{\alpha_j}(\sigma^{-1} y_i)\,.$$

We first show that $\widetilde{x^\alpha}$ is unbiased.

Lemma 7.4. *For all $\theta \in \Theta$, $\mathbb{E}_\theta^n[\widetilde{x^\alpha}] = \mathbb{E}_\theta[x^\alpha]$.*

Proof. Since $\widetilde{x^\alpha}$ is a sum of i.i.d. terms, it suffices to prove the claim for a single sample. By (3),

$$\begin{aligned}
\mathbb{E}_\theta\Big[\prod_{j=1}^{p} \sigma^{\alpha_j} H_{\alpha_j}(\sigma^{-1} y_i)\Big] &= \mathop{\mathbb{E}}_{x \sim \mu_\theta}\Big[\mathop{\mathbb{E}}_{\xi_1, \dots, \xi_p \sim \mathcal{D}}\big[\prod_{j=1}^{p} \sigma^{\alpha_j} H_{\alpha_j}(\sigma^{-1}(x_j + \sigma\xi_j))\big|x\big]\Big] \\
&= \mathop{\mathbb{E}}_{x \sim \mu_\theta}\Big[\prod_{j=1}^{p} \mathop{\mathbb{E}}_{\xi_j \sim \mathcal{D}}\big[\sigma^{\alpha_j} H_{\alpha_j}(\sigma^{-1} x_j + \xi_j)\big|x\big]\Big] \\
&= \mathop{\mathbb{E}}_{x \sim \mu_\theta}\Big[\prod_{j=1}^{p} \sigma^{\alpha_j} (\sigma^{-1} x_j)^{\alpha_j}\Big] = \mathbb{E}_\theta[x^\alpha]\,.
\end{aligned}$$

□

It remains to bound the variance.

Proposition 7.5. *For any multi-index α, there exists a constant c_α such that for all $\theta \in \Theta$, $\mathrm{Var}_\theta[\widetilde{x^\alpha}] \leq c_\alpha n^{-1} \sigma^{2|\alpha|}$.*

Proof. Since $\widetilde{x^\alpha}$ is a sum of i.i.d. terms, it suffices to prove the claim for $n = 1$. Given a multi-index α, let $c_\alpha = \prod_{j=1}^p \sup_{x_j \in [-1,1]} \mathbb{E}_{\xi_j \sim \mathcal{D}}[H_{\alpha_j}(x_j + \xi_j)^2]$, and note that c_α is independent of σ. We obtain

$$\begin{aligned}
\mathrm{Var}_\theta[\widetilde{x^\alpha}] &\leq \mathop{\mathbb{E}}_{x \sim \mu_\theta}\Big[\prod_{j=1}^{p} \mathop{\mathbb{E}}_{\xi_j \sim \mathcal{D}}\big[\sigma^{2\alpha_j} H_{\alpha_j}(\sigma^{-1} x_j + \xi_j)^2\big|x\big]\Big] \\
&\leq \sup_{x : \|x\| \leq 1} \prod_{j=1}^{p} \mathop{\mathbb{E}}_{\xi_j \sim \mathcal{D}}\big[\sigma^{2\alpha_j} H_{\alpha_j}(\sigma^{-1} x_j + \xi_j)^2\big] \\
&\leq c_\alpha \sigma^{2|\alpha|}\,,
\end{aligned}$$

as claimed. □

Finally, we apply the "median-of-means" trick [NY83] to show that we can combine the estimators defined above to obtain estimates for the moment tensors $\mathbb{E}_\theta[x^{\otimes k}]$ for $k \leq d$ which are close to their expectation with high probability.

Proposition 7.6. *Let $y_1, \ldots, y_n$ be i.i.d. samples from P_θ. For any degree d and accuracy parameter δ, there exist estimators $\widehat{x^\alpha} = \widehat{x^\alpha}(y_1, \ldots, y_n)$ for all α with $|\alpha| \leq d$ such that with probability at least $1 - \delta$,*

$$\max_{\alpha : |\alpha| \leq d} |\mathbb{E}_\theta[x^\alpha] - \widehat{x^\alpha}| \leq c_d \sigma^d \sqrt{\frac{\log(p/\delta)}{n}} \,,$$

for some constant c_d.

Proof. Split the samples into m subsamples of equal size, for some m to be specified, and for each α construct the m estimators $\widetilde{x_1^\alpha}, \ldots, \widetilde{x_m^\alpha}$ on the basis of the m subsamples. (We assume for convenience that m divides n.) Let $\widehat{x^\alpha}$ be the median of $\widetilde{x_1^\alpha}, \ldots, \widetilde{x_m^\alpha}$.

Chebyshev's inequality together with Proposition 7.5 implies that there exists a constant c_d such that, for each $j = 1, \ldots, m$ and multi-index α,

$$\Pr\left[|\widetilde{x_j^\alpha} - \mathbb{E}_\theta[x^\alpha]| > c_d \sigma^d \sqrt{\frac{m}{n}}\right] \leq 1/4 \,,$$

and since the estimators $\widetilde{x_1^\alpha}, \ldots, \widetilde{x_m^\alpha}$ are independent, a standard concentration argument shows that

$$\Pr\left[|\widehat{x^\alpha} - \mathbb{E}_\theta[x^\alpha]| > c_d \sigma^d \sqrt{\frac{m}{n}}\right] \leq e^{-m/4} \,.$$

By a "stars-and-bars" counting argument [Fel68], there are $\binom{p+d}{d}$ multi-indices α satisfying $|\alpha| \leq d$, so taking a union bound over all choices of α yields

$$\max_{\alpha : |\alpha| \leq d} |\mathbb{E}_\theta[x^\alpha] - \widehat{x^\alpha}| \leq c_d \sigma^d \sqrt{\frac{m}{n}}$$

with probability at least $1 - \binom{p+d}{d} e^{-m/4}$. Choosing $m = 4 \log(\binom{p+d}{d}/\delta)$ and taking c_d sufficiently large in the statement of the theorem yields the claim. □

Note that the constant c_d in the statement of Proposition 7.6 can be made explicit, given knowledge of the distribution $\mathcal{D}$.

7.1.2 Robust solutions to polynomial systems

We now show that approximate knowledge of the moment tensors $\mathbb{E}_\theta[x^{\otimes k}]$ for $k = 1, \ldots, d$ suffices to approximately recover θ.

Lemma 7.7. *For all $\theta \in \Theta$ and $\varepsilon > 0$, there exists a $\varepsilon' > 0$ such that, if $\phi \in \Theta$ satisfies $\max_{k \leq d} \|\mathbb{E}_\theta[x^{\otimes k}] - \mathbb{E}_\phi[x^{\otimes k}]\|_\infty < \varepsilon'$, then there exists a $\tau \in \mathcal{M}_{\theta,d}$ such that $\|\phi - \tau\| < \varepsilon$.*

Proof. We employ a simple compactness argument. Consider the set $F = \{\phi \in \Theta : \forall \, \tau \in \mathcal{M}_{\theta,d} \; \|\phi - \tau\| \geq \varepsilon\}$. Since Θ is compact, so is F. Set

$$\varepsilon' = \min_{\phi \in F} \max_{k \leq d} \|\mathbb{E}_\theta[x^{\otimes k}] - \mathbb{E}_\phi[x^{\otimes k}]\|_\infty \,.$$

Clearly if $\max_{k \leq d} \|\mathbb{E}_\theta[x^{\otimes k}] - \mathbb{E}_\phi[x^{\otimes k}]\|_\infty < \varepsilon'$ for some $\phi \in \Theta$, then there exists a $\tau \in \mathcal{M}_{\theta,d}$ such that $\|\phi - \tau\| < \varepsilon$, so it remains to check that $\varepsilon' > 0$.

Since $\theta \mapsto \mu_\theta$ is continuous with respect to the weak topology and μ_θ is supported on a compact set for all $\theta \in \Theta$, the moment map $\theta \mapsto \mathbb{E}_\theta[x^{\otimes k}]$ is also continuous for all $k \leq d$. If $\phi \in F$, then in particular $\phi \notin \mathcal{M}_{\theta,d}$, so there exists a $k \leq d$ for which $\mathbb{E}_\theta[x^{\otimes k}] \neq \mathbb{E}_\phi[x^{\otimes k}]$. Therefore $\varepsilon' > 0$, as desired. □

Lemma 7.7 is simply stating that the function $\phi \mapsto \min_{\tau \in \mathcal{M}_{\theta,d}} \|\phi - \tau\|$ is continuous at θ with respect to the topology induced by the moment maps. Note that, for generic θ when μ_θ arises from an orbit recovery problem, the moment map will be continuously differentiable with a nonsingular Jacobian, so the inverse function theorem implies ε' can be taken to be $\Omega(\varepsilon)$. In general, however, the dependence could be worse.

7.1.3 Proof of Theorem 3.2

Construct the estimators $\widehat{x^\alpha}$ as in Proposition 7.6, and let

$$\widehat{\mathcal{M}}_n = \left\{ \phi \in \Theta : \max_{\alpha : |\alpha| \leq d} |\mathbb{E}_\phi[x^\alpha] - \widehat{x^\alpha}| \leq c_d \sigma^d \sqrt{\frac{\log(p/\delta)}{n}} \right\}$$

Applying Proposiiton 7.6, we have with probability at least $1 - \delta$ that $\mathcal{M}_{\theta,d} \subseteq \widehat{\mathcal{M}}_n$ and that, for all $\phi \in \widehat{\mathcal{M}}_n$,

$$\max_{k \leq d} \|\mathbb{E}_\phi[x^{\otimes k}] - \mathbb{E}_\theta[x^{\otimes k}]\| = \max_{\alpha : |\alpha| \leq d} |\mathbb{E}_\phi[x^\alpha] - \mathbb{E}_\theta[x^\alpha]| \leq 2 c_d \sigma^d \sqrt{\frac{\log(p/\delta)}{n}} .$$

By Lemma 7.7, there exists an $\varepsilon'_{\theta,\varepsilon}$ such that, as long as $2c_d\sigma^d\sqrt{\frac{\log(p/\delta)}{n}} < \varepsilon'_{\theta,\varepsilon}$, then with probability at least $1 - \delta$, we have the desired inclusion $\widehat{\mathcal{M}}_n \subseteq \mathcal{M}^\varepsilon_{\theta,d}$.

Therefore taking $n > (2c_d/\varepsilon'_{\theta,\varepsilon})^2 \log(p/\delta)\sigma^{2d} = c_{\theta,\varepsilon,d}\log(1/\delta)\sigma^{2d}$ suffices. □

7.2 Information geometry of gaussian mixtures

In this section, we establish an upper bound on the Kullbeck-Leibler divergence between different gaussian mixtures, which we denote by $D(\cdot \,\|\, \cdot)$.

The proof follows the outline used in [BRW17], based off a technique developed in [LNS99, CL11].

Proposition 7.8. *Let $\theta, \phi \in \Theta$, let the distribution $\mathcal{D}$ be $\mathcal{N}(0,1)$ for some $\sigma \geq 1$, and let d be any positive integer.*

There exist universal constants C and c such that if $\mathbb{E}_\theta[x^{\otimes k}] = \mathbb{E}_\phi[x^{\otimes k}]$ for $k \leq d-1$, then

$$D(\mathrm{P}_\theta \,\|\, \mathrm{P}_\phi) \leq C\frac{(c\sigma)^{-2d}}{d!} .$$

Proof. We first establish the claim when $d = 1$. Note that the condition on the moment tensors is vacuous in this case. By the convexity of the divergence,

$$D(\mathrm{P}_\theta \,\|\, \mathrm{P}_\phi) \leq \mathop{\mathbb{E}}_{\substack{x \sim \mu_\theta \\ x' \sim \mu_\phi}} D(\mathcal{N}(x, \sigma^2), \mathcal{N}(x', \sigma^2)) = \mathop{\mathbb{E}}_{\substack{x \sim \mu_\theta \\ x' \sim \mu_\phi}} \frac{\|x - x'\|^2}{2\sigma^2} \leq 2\sigma^{-2} ,$$

where in the last step we used the fact that x and x' lie in the unit ball almost surely.

Now, assume $d > 1$, so in particular $\mathbb{E}_\theta[x] = \mathbb{E}_\phi[x]$. Denote their common mean by v. For $\zeta \in \{\theta, \phi\}$ denote by $\bar{\mu}_\zeta$ the distribution of $x - v$ when $x \sim \mu_\zeta$, and let $\overline{\mathrm{P}}_\zeta$ denote distribution of y when $y = x + \xi$ for $x \sim \bar{\mu}_\zeta$ and $\xi \sim \mathcal{N}(0, \sigma^2 I)$. Since this transformation is a deterministic bijection, the data processing inequality implies $D(\mathrm{P}_\theta \,\|\, \mathrm{P}_\phi) = D(\overline{\mathrm{P}}_\theta \,\|\, \overline{\mathrm{P}}_\phi)$.

Note that $\mathbb{E}_{\bar{\mu}_\theta}[x] = \mathbb{E}_{\bar{\mu}_\theta}[x] = 0$ and $\mathbb{E}_{\bar{\mu}_\theta}[x^{\otimes k}] = \mathbb{E}_{\bar{\mu}_\phi}[x^{\otimes k}]$ for $k \leq d-1$. Hence without loss of generality we can reduce to the case where μ_θ and μ_ϕ are both centered and are supported in a ball of radius 2.

We bound the χ^2-divergence between P_θ and P_ϕ. Let f be the density of a standard p-dimensional Gaussian and for $\zeta \in \Theta$, let f_ζ be the density of P_ζ, which can be written explicitly as

$$f_\zeta(y) = \mathbb{E}_{x \sim \mu_\zeta} \sigma^{-p} f(\sigma^{-1}(y - x)) = \sigma^{-p} f(\sigma^{-1} y) \mathbb{E}_{x \sim \mu_\zeta} e^{-\frac{1}{2\sigma^2}(x^2 - 2y^\top x)} .$$

Since $\|x\| \leq 2$ almost surely with respect μ_ζ, Jensen's inequality implies that

$$f_\zeta(y) \geq \sigma^{-p} f(\sigma^{-1} y) e^{-\frac{1}{2\sigma^2}(4 - 2y^\top \mathbb{E}_\zeta x)} = \sigma^{-p} f(\sigma^{-1} y) e^{-\frac{2}{\sigma^2}} \,. \tag{4}$$

Recall that the χ^2 divergence is defined by

$$\chi^2(\mathrm{P}_\theta, \mathrm{P}_\phi) = \int \frac{(f_\theta(y) - f_\phi(y))^2}{f_\theta(y)} \mathrm{d}y \,.$$

Applying (4) to the denominator, expanding the definitions of f_θ and f_ϕ, and applying a change of variables yields

$$\begin{aligned} \chi^2(\mathrm{P}_\theta, \mathrm{P}_\phi) &\leq e^{2/\sigma^2} \int (\mathbb{E}_\theta e^{-\frac{1}{2\sigma^2}(x^2 - 2y^\top x)} - \mathbb{E}_\phi e^{-\frac{1}{2\sigma^2}(x^2 - 2y^\top x)})^2 \sigma^{-p} f(\sigma^{-1} y) \, \mathrm{d}y \\ &= e^{2/\sigma^2} \int (\mathbb{E}_\theta e^{y^\top(\sigma^{-1}x) - \frac{1}{2}\|\sigma^{-1}x\|^2} - \mathbb{E}_\phi e^{y^\top(\sigma^{-1}x) - \frac{1}{2}\|\sigma^{-1}x\|^2})^2 f(y) \, \mathrm{d}y \\ &= e^{2/\sigma^2} \mathbb{E}(\mathbb{E}_\theta e^{g^\top(\sigma^{-1}x) - \frac{1}{2}\|\sigma^{-1}x\|^2} - \mathbb{E}_\phi e^{g^\top(\sigma^{-1}x) - \frac{1}{2}\|\sigma^{-1}x\|^2})^2, \qquad g \sim \mathcal{N}(0, I) \,. \end{aligned} \tag{5}$$

Given $\zeta, \zeta' \in \Theta$, let $x \sim \mu_\zeta$ and $x' \sim \mu_{\zeta'}$ be independent. Then interchanging the order of expectation and using the expression for the moment generating function of a standard Gaussian random variable, we obtain

$$\mathbb{E}_{\zeta,\zeta'} \mathbb{E}_g e^{g^\top(\sigma^{-1}(x + x')) - \frac{1}{2}(\|\sigma^{-1}x\|^2 + \|\sigma^{-1}x'\|^2)} = \mathbb{E}_{\zeta,\zeta'} e^{\frac{x^\top x'}{\sigma^2}} \,.$$

Applying this expression to (5) after expanding the square produces

$$\chi^2(\mathrm{P}_\theta, \mathrm{P}_\phi) \leq e^{2/\sigma^2} \left(\underset{\substack{x \sim \mu_\theta \\ x' \sim \mu_\theta}}{\mathbb{E}} e^{\frac{x^\top x'}{\sigma^2}} - 2 \underset{\substack{x \sim \mu_\theta \\ x' \sim \mu_\phi}}{\mathbb{E}} e^{\frac{x^\top x'}{\sigma^2}} + \underset{\substack{x \sim \mu_\phi \\ x' \sim \mu_\phi}}{\mathbb{E}} e^{\frac{x^\top x'}{\sigma^2}} \right),$$

where in each expectation the random variables x and x' are independent. Since μ_θ and μ_ϕ are compactly supported, Fubini's theorem implies we can expand each term as a power series and interchange expectation and summation to produce

$$\begin{aligned} \chi^2(\mathrm{P}_\theta, \mathrm{P}_\phi) &\leq e^{2/\sigma^2} \sum_{k=0}^{\infty} \frac{\sigma^{-2k}}{k!} \left(\underset{\substack{x \sim \mu_\theta \\ x' \sim \mu_\theta}}{\mathbb{E}} (x^\top x')^k - 2 \underset{\substack{x \sim \mu_\theta \\ x' \sim \mu_\phi}}{\mathbb{E}} (x^\top x')^k + \underset{\substack{x \sim \mu_\phi \\ x' \sim \mu_p hi}}{\mathbb{E}} (x^\top x')^k \right) \\ &= e^{2/\sigma^2} \sum_{k=0}^{\infty} \frac{\sigma^{-2k}}{k!} \left(\langle \mathbb{E}_\theta x^{\otimes k}, \mathbb{E}_\theta x^{\otimes k} \rangle - 2 \langle \mathbb{E}_\theta x^{\otimes k}, \mathbb{E}_{x\phi} x^{\otimes k} \rangle + \langle \mathbb{E}_\phi x^{\otimes k}, \mathbb{E}_\phi x^{\otimes k} \rangle \right) \\ &= e^{2/\sigma^2} \sum_{k=0}^{\infty} \frac{\sigma^{-2k}}{k!} \|\mathbb{E}_\theta[x^{\otimes k}] - \mathbb{E}_\phi[x^{\otimes k}]\|_{\mathrm{HS}}^2 \\ &= e^{2/\sigma^2} \sum_{k=d}^{\infty} \frac{\sigma^{-2k}}{k!} \|\mathbb{E}_\theta[x^{\otimes k}] - \mathbb{E}_\phi[x^{\otimes k}]\|_{\mathrm{HS}}^2 \,, \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ denotes the Frobenius inner product between tensors and $\| \cdot \|_{\mathrm{HS}}$ denotes the Hilbert-Schmidt norm. Since under both μ_θ and μ_ϕ, $\|x\| \leq 2$ almost surely, we have for all $k \geq 2$,

$$\|\mathbb{E}_\theta[x^{\otimes k}] - \mathbb{E}_\phi[x^{\otimes k}]\|_{\mathrm{HS}}^2 \leq 2\|\mathbb{E}_\theta[x^{\otimes k}]\|_{\mathrm{HS}}^2 + 2\|\mathbb{E}_\phi[x^{\otimes k}]\|_{\mathrm{HS}}^2 \leq 4^{k+1} \,.$$

Therefore

$$\chi^2(\mathrm{P}_\theta, \mathrm{P}_\phi) \leq 4 e^{2/\sigma^2} \sum_{k=d}^{\infty} \frac{4^k \sigma^{-2k}}{k!} \leq 4 e^{6/\sigma^2} \frac{4^d \sigma^{-2d}}{d!} \,,$$

and applying the inequality $D(\mathrm{P}_\theta \,\|\, \mathrm{P}_\phi) \leq \chi^2(\mathrm{P}_\theta, \mathrm{P}_\phi)$ [Tsy09] proves the claim. □

7.2.1 Proof of Theorem 3.4

If τ_1 and τ_2 are both in $\mathcal{M}_{\theta,d-1}$, then by Proposition 7.8, the corresponding distributions P_{τ_1} and P_{τ_2} satisfy $D(\mathrm{P}_{\tau_1} \,\|\, \mathrm{P}_{\tau_2}) \leq \frac{C}{(c\sigma)^{2d}d!}$. By the Neyman-Pearson lemma, for any test ψ using n samples,

$$\Pr_{\tau_1}(\psi = 2) + \Pr_{\tau_2}(\psi = 1) \geq 1 - d_{\mathrm{TV}}(\mathrm{P}_{\tau_1}^n, \mathrm{P}_{\tau_2}^n) \geq 1 - \sqrt{\frac{1}{2}D(\mathrm{P}_{\tau_1}^n \,\|\, \mathrm{P}_{\tau_2}^n)} = 1 - \sqrt{\frac{Cn}{2(c\sigma)^{2d}d!}}\,,$$

where we have applied Pinsker's inequality and the chain rule for divergence. Therefore, to achieve an error probability of at most 1/3, we must have $n \geq 2(c\sigma)^{2d}d!/(9C) = c_d\sigma^{2d}$, as claimed. □

7.3 Proofs for Section 4: algebraic results

7.3.1 Algorithm for generators of $\mathbb{R}[\mathbf{x}]^G$

We know that $\mathbb{R}[\mathbf{x}]^G$ is finitely generated as an $\mathbb{R}$-algebra (Theorem 4.3). There are various algorithms to compute a finite set of generators for $\mathbb{R}[\mathbf{x}]^G$ [Stu08, DK15]. However, some require the group to be finite or to be reductive over an algebraically-closed field. One algorithm that certainly works in our context (compact groups) is Algorithm 2.2.5 in [Stu08]. As input it requires the Hilbert series of $\mathbb{R}[\mathbf{x}]^G$ (which can be computed by Proposition 4.12) and a procedure to compute a basis for $\mathbb{R}[\mathbf{x}]^G_d$ (which can be done with the Reynolds operator by Observation 4.2). The idea is as follows. We keep a set of proposed generators $f_1, \ldots, f_m$. At each step we compare the Hilbert series of $\mathbb{R}[\mathbf{x}]^G$ with the Hilbert series of $\mathbb{R}[f_1, \ldots, f_m]$ (which can be computed using Gröbner bases). If these series differ at the t^d term, this means we are missing an invariant at degree d. To remedy this, we create a new homogeneous invariant of degree d using the Reynolds operator, and add it to our set of proposed generators. We repeat until the Hilbert series match.

7.3.2 Bounding the list size for generic signals

In this section we prove Theorem 4.21 and the first part of Theorem 4.9 (see Section 7.3.3 for the second part). Recall the following basic definitions and facts from field theory.

Definition 7.9. If F_2 is a subfield of F_1, we write F_1/F_2 and call this a *field extension*. The *degree* of the extension, denoted $[F_1 : F_2]$, is the dimension of F_1 as a vector space over F_2.

Proposition 7.10. *Let $\mathbb{R} \subseteq F_2 \subseteq F_1$ with F_1 finitely generated (as a field) over $\mathbb{R}$. Let r be the transcendence degree of F_1 (over $\mathbb{R}$). The field extension F_1/F_2 has finite degree if and only if F_1 contains r algebraically independent elements.*

Proof. This is a basic fact of field theory. If F_1 contains r algebraically independent elements then the extension F_1/F_2 is algebraic and finitely generated, and therefore has finite degree. Otherwise, the extension is transcendental and has infinite degree. □

In light of the above (and using the fact that $\mathbb{R}[\mathbf{x}]^G$ is finitely generated), Theorem 4.21 implies the first part of Theorem 4.9 and so it remains to prove Theorem 4.21 (i.e. list size is bounded by $D := [F_G : \mathbb{R}(U)]$).

Proof of Theorem 4.21.
Write $F_U := \mathbb{R}(U)$. In characteristic zero, every algebraic extension is separable, so by the primitive element theorem, $F_G = F_U(\alpha)$ for some $\alpha \in F_G$. Since α generates a degree-D extension, α is the root of a degree-D polynomial

$$\alpha^D + b_{D-1}\alpha^{D-1} + \cdots + b_1\alpha + b_0 \tag{6}$$

with coefficients $b_i \in F_U$. Furthermore, every element of F_G can be expressed as

$$c_0 + c_1\alpha + \cdots + c_{D-1}\alpha^{D-1}$$

with $c_i \in F_U$. In particular, let $g_1, \ldots, g_k$ be generators for $\mathbb{R}[\mathbf{x}]^G$ (as an $\mathbb{R}$-algebra) and write

$$g_i = c_0^{(i)} + c_1^{(i)}\alpha + \cdots + c_{D-1}^{(i)}\alpha^{D-1}. \tag{7}$$

Let $S \subseteq V$ be the subset for which α and all the (finitely-many) coefficients $b_i, c_j^{(i)}$ have nonzero denominators; S is a non-empty Zariski-open set and thus has full measure. Now fix $\theta \in S$. Given the values $f(\theta)$ for all $f \in U$, each b_i takes a well-defined value in $\mathbb{R}$ and so from (6) there are at most D possible values that $\alpha(\theta)$ can take. From (7), each value of $\alpha(\theta)$ uniquely determines all the values $g_i(\theta)$ and thus uniquely determines all the values $f(\theta)$ for $f \in \mathbb{R}[\mathbf{x}]^G$. Since $\mathbb{R}[\mathbf{x}]^G$ resolves θ (Theorem 4.4), this completes the proof. □

7.3.3 Generic list recovery converse

In this section we prove the second part of Theorem 4.9 (the converse).

Let $p = \dim(V)$, $\operatorname{trdeg}(U) = q$, and $\operatorname{trdeg}(\mathbb{R}[\mathbf{x}]^G) = r$ so that $q < r \le p$. Let $\mathbf{f} = \{f_1, \ldots, f_m\}$ be a basis for U, and let $\mathbf{g} = \{g_1, \ldots, g_r\}$ be a transcendence basis for $\mathbb{R}[\mathbf{x}]^G$. Let $S \subseteq V$ be the set of points θ for which the Jacobian $J_\mathbf{x}(\mathbf{f})|_{\mathbf{x}=\theta}$ has row rank q and the Jacobian $J_\mathbf{x}(\mathbf{g})|_{\mathbf{x}=\theta}$ has row rank r; by the Jacobian criterion (see Corollary 4.18), S is a non-empty Zariski-open set and thus has full measure.

Fix $\theta \in S$. For a sufficiently small open neighborhood $X \subseteq S$ containing θ we have the following. The Jacobian condition on $\mathbf{f}$ implies that $\{\tau \in X \;:\; \mathbf{f}(\tau) = \mathbf{f}(\theta)\}$ has dimension $p - q$. The Jacobian condition on $\mathbf{g}$ implies that every $z \in \mathbf{g}(X)$ has a preimage $\mathbf{g}^{-1}(z) := \{\tau \in X \;:\; \mathbf{g}(\tau) = z\}$ of dimension $p - r$. Since $p - q > p - r$ it follows that there are infinitely many $\theta_1, \theta_2, \ldots \in X$ such that $\mathbf{f}(\theta_i) = \mathbf{f}(\theta)$ but the values $\mathbf{g}(\theta_1), \mathbf{g}(\theta_2), \ldots$ are all distinct (and thus the θ_i belong to distinct orbits). Therefore U does not list resolve θ.

7.3.4 Hilbert series and Hironaka decomposition

In this section we prove Proposition 4.13 on extracting the transcendence degree from the Hilbert series (as the pole order at $t = 1$). We require the following key structural property of invariant rings, which is called the *Cohen-Macaulay property* or *Hironaka decomposition*.

Theorem 7.11 ([DK15] Section 2.6)**.** *The invariant ring $\mathbb{R}[\mathbf{x}]^G$ has the following structure. There exist homogeneous* primary invariants *$f_1, \ldots, f_r \in \mathbb{R}[\mathbf{x}]^G$ and homogeneous* secondary invariants *$g_1, \ldots, g_s \in \mathbb{R}[\mathbf{x}]^G$ such that*

- *$\{f_1, \ldots, f_r\}$ are algebraically independent, and*
- *any element of $\mathbb{R}[\mathbf{x}]^G$ can be written uniquely as a linear combination of $g_1, \ldots, g_s$ with coefficients from $\mathbb{R}[f_1, \ldots, f_r]$.*

The proof can be found in Section 2.6 of [DK15]; note that the only property of the group that is used is the existence of a Reynolds operator (and so the proof is valid for compact groups).

Proof of Proposition 4.13.
The Hironaka decomposition above implies that the Hilbert series takes the form

$$\frac{\sum_{j=1}^s t^{\deg(g_j)}}{\prod_{i=1}^r (1 - t^{\deg(f_i)})}$$

(this is equation (2.7.3) in [DK15]). It is now clear that the order of the pole at $t = 1$ is precisely r. But we can see as follows that $f_1, \ldots, f_r$ is a transcendence basis for $\mathbb{R}[\mathbf{x}]^G$ and so $r = \operatorname{trdeg}(\mathbb{R}[\mathbf{x}]^G)$. As in the proof of Theorem 4.26, since $\mathbb{R}[\mathbf{x}]^G$ is a finitely generated $\mathbb{R}[f_1, \ldots, f_r]$-module, every $h \in \mathbb{R}[\mathbf{x}]^G$ satisfies a polynomial with coefficients in $\mathbb{R}[f_1, \ldots, f_r]$, which is an algebraic dependence among $\{f_1, \ldots, f_r, h\}$. □

7.3.5 Transcendence degree for heterogeneity

In this section we prove Proposition 4.14. To recall the setup, we have $\tilde{G}$ acting on $\tilde{V}$ with associated variables $\tilde{\mathbf{x}}$. We also have $G = \tilde{G}^K \rtimes S_K$ acting on $V = \tilde{V}^{\oplus K} \oplus \Delta_K$ with associated variables $\mathbf{x}$. Let us also introduce an intermediate group: $G' = \tilde{G}^K$, acting on V (with associated variables $\mathbf{x}$).

Partition the variables $\mathbf{x}$ as follows. For $k = 1, \dots, K$, let $\mathbf{x}^{(k)} = (x_1^{(k)}, \dots, x_p^{(k)})$ be the variables corresponding to signal k. Let $\mathbf{z} = (z_1, \dots, z_{K-1})$ be the variables corresponding to the mixing weights $\overline{w}_1, \dots, \overline{w}_{K-1}$ (see Remark 2.3). Whenever we refer to z_K, this is just shorthand for $-\sum_{k=1}^{K-1} z_k$.

We first prove a simpler version of the result without the action of S_K.

Lemma 7.12. *Let* $\tilde{r} = \mathrm{trdeg}(\mathbb{R}[\tilde{\mathbf{x}}]^{\tilde{G}})$ *and let* $r = K\tilde{r} + K - 1$*. Then*

$$\mathrm{trdeg}(\mathbb{R}[\mathbf{x}]^{G'}) = r.$$

Proof. To show '$\geq$' we need to exhibit r algebraically independent elements of $\mathbb{R}[\mathbf{x}]^{G'}$. Letting $f_1, \dots, f_{\tilde{r}}$ be a transcendence basis for $\mathbb{R}[\tilde{\mathbf{x}}]^{\tilde{G}}$, it suffices to take

$$I := \{f_i(\mathbf{x}^{(k)})\}_{1 \leq i \leq \tilde{r}, 1 \leq k \leq K} \cup \{z_1, \dots, z_{K-1}\}.$$

To show '$\leq$' we first recall that we can obtain a spanning set for the subspace $\mathbb{R}[\mathbf{x}]_d^{G'}$ by applying the Reynolds operator $\mathcal{R}$ (for G') to each degree-d monomial (in the variables $\mathbf{x}$). Such a monomial takes the form

$$m(\mathbf{x}) = M(\mathbf{z}) \prod_{k=1}^{K} m_k(\mathbf{x}^{(k)})$$

where M, m_k are monomials. Applying the Reynolds operator yields

$$\mathcal{R}(m(\mathbf{x})) = \mathop{\mathbb{E}}_{g_1, \dots, g_k \sim \tilde{G}} M(\mathbf{z}) \prod_{k=1}^{K} m_k(g_k \cdot \mathbf{x}^{(k)}) = M(\mathbf{z}) \prod_{k=1}^{K} \mathop{\mathbb{E}}_{g_k \sim \tilde{G}} m_k(g_k \cdot \mathbf{x}^{(k)}).$$

Note that $\mathcal{R}(m(\mathbf{x}))$ is the product of *pure* invariants, i.e. invariants that only involve variables from a single one of the blocks $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(K)}, \mathbf{z}$. It is clear that I (from above) is a maximal set of algebraically independent pure invariants. It is now easy to show using the Jacobian condition (Proposition 4.17) that if any $\mathcal{R}(m(\mathbf{x}))$ is added to I, it will no longer be algebraically independent. The result now follows using basic properties of algebraic independence (Proposition 4.19 and Lemma 4.20). □

Proof of Proposition 4.14.
Since $\mathbb{R}[\mathbf{x}]^G \subseteq \mathbb{R}[\mathbf{x}]^{G'}$, it is clear (in light of the above) that $\mathrm{trdeg}(\mathbb{R}[\mathbf{x}]^G) \leq r$. Thus we need only to show $\mathrm{trdeg}(\mathbb{R}[\mathbf{x}]^G) \geq r$ by demonstrating r algebraically independent invariants. Let $e_1, \dots, e_K$ be the elementary symmetric functions in K variables. With f_i as above, we take the invariants

$$\{e_k(f_i(\mathbf{x}^{(1)}), \dots, f_i(\mathbf{x}^{(K)}))\}_{1 \leq i \leq \tilde{r}, 1 \leq k \leq K} \cup \{e_2(z_1, \dots, z_K), \dots, e_K(z_1, \dots, z_K)\}.$$

Note that $e_1(z_1, \dots, z_K)$ is not included because it is equal to 0. The fact that $e_k(f_i(\mathbf{x}^{(1)}), \dots, f_i(\mathbf{x}^{(K)}))$ are algebraically independent can be seen because $\{e_1, \dots, e_K\}$ is algebraically independent and $\{f_i(\mathbf{x}^{(k)})\}_{i,k}$ is algebraically independent. We can see that $\{e_k(z_1, \dots, z_K)\}_{k \geq 2}$ are algebraically independent as follows. If there were an algebraic dependence then we would have a polynomial P such that $P(e_2(z_2, \dots, z_K), \dots, e_K(z_1, \dots, z_K))$ (now treating z_K as a separate variable) has a root $z_K = -\sum_{k=1}^{K-1} z_k$ and thus has $e_1(z_1, \dots, z_K)$ as factor. But this contradicts the fact that any symmetric polynomial has a *unique* representation in terms of the elementary symmetric polynomials.

□

7.3.6 Gröbner bases

In this section we show how to use Gröbner bases to test various algebraic conditions. In particular, we prove Theorems 4.23 and 4.25. The ideas from this section are mostly standard in the theory of Gröbner bases; see e.g. [CLO07] for a reference.

Definition 7.13. A *monomial order* on $\mathbb{R}[\mathbf{x}]$ is a well-ordering on the set $\mathcal{M}$ of all (monic) monomials, satisfying $M \le N \Leftrightarrow MP \le NP$ for all $M, N, P \in \mathcal{M}$. We will say that a monomial order *favors* a variable x_i if the monomial x_i is larger (with respect to the monomial order) than any monomial not involving x_i. We write $\mathrm{LM}(f)$ to denote the leading monomial of a polynomial f, i.e. the monomial occurring in f that is largest (with respect to the monomial order); $\mathrm{LM}(f)$ does not include the coefficient.

Definition 7.14. A *Gröbner basis* of an ideal $I \subseteq \mathbb{R}[\mathbf{x}]$ is a finite subset $B \subseteq I$ such that for every $f \in I$ there exists $b \in B$ such that $\mathrm{LM}(f)$ is a multiple of $\mathrm{LM}(b)$. We call B a *reduced Gröbner basis* if all its elements are monic and it has the additional property that for every pair of distinct $b, b' \in B$, no monomial occurring in b is a multiple of $\mathrm{LM}(b')$.

The following basic facts about Gröbner bases are proved in [CLO07]. A Gröbner basis is indeed a basis in the sense that it generates the ideal. Every ideal $I \subseteq \mathbb{R}[\mathbf{x}]$ has a Gröbner basis, and has a unique reduced Gröbner basis. Furthermore, there is an algorithm (*Buchberger's algorithm*) to compute the reduced Gröbner basis of an ideal $I = \langle f_1, \ldots, f_m \rangle$, given a list of generators f_i. (It is not a polynomial-time algorithm, however.)

Suppose we are interested in the relations between polynomials $f_1, \ldots, f_m \in \mathbb{R}[\mathbf{x}]$. Introduce additional variables $\mathbf{t} = (t_1, \ldots, t_m)$ and consider the ideal $I := \langle f_1(\mathbf{x}) - t_1, \ldots, f_m(\mathbf{x}) - t_m \rangle \subseteq \mathbb{R}[\mathbf{x}, \mathbf{t}]$. Given $f_1, \ldots, f_m$ there is an algorithm to compute a Gröbner basis for the *elimination ideal*

$$J := \langle f_1(\mathbf{x}) - t_1, \ldots, f_m(\mathbf{x}) - t_m \rangle \cap \mathbb{R}[\mathbf{t}].$$

In fact, the algorithm is simply to compute a Gröbner basis for I using a particular monomial order and then keep only the elements that depend only on $\mathbf{t}$ (see Chapter 3 of [CLO07]). The elimination ideal consists precisely of the polynomial relations among $f_1, \ldots, f_m$:

Lemma 7.15. *For any polynomial $P \in \mathbb{R}[\mathbf{t}]$ we have: $P \in J$ if and only if $P(f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})) \equiv 0$.*

Proof. The direction '$\Rightarrow$' is clear because if we let $t_i = f_i(\mathbf{x})$ for all i then the generators of I vanish and so every element of I vanishes. To show the converse, it suffices to show that for any polynomial $P \in \mathbb{R}[\mathbf{t}]$, $P(f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})) - P(t_1, \ldots, t_m) \in I$. This can be shown inductively using the following key idea:

$$x_1 x_2 - t_1 t_2 = \frac{1}{2}(x_1 - t_1)(x_2 + t_2) + \frac{1}{2}(x_2 - t_2)(x_1 + t_1)$$

and so $x_1 x_2 - t_1 t_2 \in \langle x_1 - t_1, x_2 - t_2 \rangle$. □

Generation as an $\mathbb{R}$-algebra. Suppose we want to know whether $f_m \in \mathbb{R}[f_1, \ldots, f_{m-1}]$. This is equivalent to asking whether there exists $P \in J$ of the form

$$P(\mathbf{t}) = t_m - Q(t_1, \ldots, t_{m-1}) \tag{8}$$

for some $Q \in \mathbb{R}[t_1, \ldots, t_{m-1}]$. Suppose that J contains an element P of the form (8). Compute a Gröbner basis B for J with respect to a monomial order that favors t_m. The leading monomial of P is t_m so by the definition of a Gröbner basis there must be an element $b \in B$ whose leading monomial divides t_m. Since $1 \notin J$ (by Lemma 7.15), the leading monomial of b is exactly t_m and so b takes the form (8). Therefore, $f_m \in \mathbb{R}[f_1, \ldots, f_{m-1}]$ if and only if B contains an element of the form (8).

We can now prove Theorem 4.25: to test whether $\mathbb{R}[f_1, \ldots, f_m] = \mathbb{R}[\mathbf{x}]^G$, compute generators $g_1, \ldots, g_s$ for $\mathbb{R}[\mathbf{x}]^G$ (see Section 7.3.1) and use the above to test whether each g_i is in $\mathbb{R}[f_1, \ldots, f_m]$.

Generation as a field. Suppose we want to know whether $f_m \in \mathbb{R}(f_1, \ldots, f_{m-1})$. This is equivalent to asking whether f_m can be expressed as a rational function of $f_1, \ldots, f_{m-1}$ (with coefficients in $\mathbb{R}$), which is equivalent (by multiplying through by the denominator) to asking whether there exists $P \in J$ of the form

$$P(\mathbf{t}) = t_m Q_1(t_1, \ldots, t_{m-1}) - Q_2(t_1, \ldots, t_{m-1}) \quad \text{with } Q_1 \notin J. \tag{9}$$

Suppose that J contains an element P of the form (9). Compute a *reduced* Gröbner basis B for J with respect to a monomial order that favors t_m. It is a basic property of Gröbner bases that P can be written as

$$P(\mathbf{t}) = \sum_i p_i(\mathbf{t}) b_i(\mathbf{t})$$

where $p_i \in \mathbb{R}[\mathbf{t}]$ and $b_i \in B$ with $\mathrm{LM}(p_i) \le \mathrm{LM}(P)$ and $\mathrm{LM}(b_i) \le \mathrm{LM}(P)$. If no b_i involves the variable t_m then $Q_1 \in J$, a contradiction. Therefore some b_j must have degree 1 in t_m. Since B is a reduced Gröbner basis it cannot contain any element of the form (9) with $Q_1 \in J$. This completes the proof that $f_m \in \mathbb{R}(f_1, \ldots, f_{m-1})$ if and only if B contains an element of the form (9).

Degree of field extension. Consider the setup from Theorem 4.21. We have a finite set $U = \{f_1, \ldots, f_m\} \subseteq \mathbb{R}[\mathbf{x}]^G$ and want to compute $[F_G : F_U]$ where $F_U = \mathbb{R}(U)$ and F_G is the fields of fractions of $\mathbb{R}[\mathbf{x}]^G$. We know there is an algorithm (see Section 7.3.1) to compute a finite list of generators $g_1, \ldots, g_s$ for $\mathbb{R}[\mathbf{x}]^G$ (as an $\mathbb{R}$-algebra). Let $\alpha = \sum_{i=1}^s \lambda_i g_i$ with $\lambda_i \in \mathbb{R}$ chosen generically; for all but finitely-many choices of $\{\lambda_i\}$, $F_G = F_U(\alpha)$, i.e. α generates the field extension of interest (this fact is related to the *primitive element theorem*).

In light of the above, $[F_G : F_U]$ is equal to the smallest positive integer D for which there exists a relation

$$Q_D(f_1, \ldots, f_m)\alpha^D + \cdots + Q_1(f_1, \ldots, f_m)\alpha + Q_0(f_1, \ldots, f_m) \equiv 0$$

for polynomials Q_i with $Q_D(f_1, \ldots, f_m) \not\equiv 0$. This can be tested similarly to field generation. Compute a reduced Gröbner basis B for the elimination ideal $J \subseteq \mathbb{R}[t_1, \ldots, t_m, \tau]$ consisting of the relations among $f_1, \ldots, f_m, \alpha$; use a monomial order that favors τ. Then $[F_G : F_U]$ is equal to the smallest positive integer D for which B contains an element of degree D in τ (or ∞ if B contains no element that involves τ). This proves Theorem 4.23.

Remark 7.16. An alternative to using Gröbner bases for the above tasks is to solve a (very large) linear system in order to find the minimal relation among a set of polynomials. There are bounds on the maximum possible degree of such a relation (if one exists) [Kay09].

7.4 Proofs for S^2 registration

7.4.1 Formula for Hilbert series of $\mathbb{R}[\mathbf{x}]^G$

We can derive the Hilbert series of $\mathbb{R}[\mathbf{x}]^G$ for S^2 registration using the methods in Section 4.6 of [DK15]. Recall Molien's formula (Proposition 4.12):

$$H(t) = \mathop{\mathbb{E}}_{g \sim \mathrm{Haar}(G)} \det(I - t\,\rho(g))^{-1}.$$

Note that $\det(I - t\,\rho(g))$ depends only on the conjugacy class of g. In $SO(3)$, two elements are conjugate if and only if they rotate by the same angle ϕ. When $g \sim \mathrm{Haar}(SO(3))$, the angle $\phi = \phi(g)$ is distributed with density function $\frac{1}{\pi}(1 - \cos\phi)$ on $[0, \pi]$ (see e.g. [Sal79]). If g has angle ϕ, the matrix $\rho_\ell(g)$ by which it acts on the irreducible representation V_ℓ has eigenvalues $e^{-\mathbf{i}\ell\phi}, e^{-\mathbf{i}(\ell-1)\phi}, \ldots, e^{\mathbf{i}\ell\phi}$ (see e.g. [Vve]). The matrix $\rho(g)$ by which g acts on $V = \oplus_{\ell \in \mathcal{F}} V_\ell$ is block diagonal with blocks $\rho_\ell(g)$. Using the above we write an expression for the Hilbert series:

$$H(t) = \frac{1}{\pi} \int_0^\pi \frac{1 - \cos\phi}{\prod_{\ell \in \mathcal{F}} \prod_{m=-\ell}^{\ell} (1 - t e^{\mathbf{i}m\phi})}\, \mathrm{d}\phi = \frac{1}{2\pi} \int_0^{2\pi} \frac{1 - \frac{1}{2}(e^{\mathbf{i}\phi} + e^{-\mathbf{i}\phi})}{\prod_{\ell \in \mathcal{F}} \prod_{m=-\ell}^{\ell} (1 - t e^{\mathbf{i}m\phi})}\, \mathrm{d}\phi.$$

Now write this as a complex contour integral around the unit circle in $\mathbb{C}$ and apply the residue theorem from complex analysis to arrive at the result stated in Proposition 5.4.

7.4.2 Formula for dimension of $\mathbb{R}[\mathbf{x}]_d^G$

The dimension of $\mathbb{R}[\mathbf{x}]^G$ can be extracted as the coefficient of t^d in the Hilbert series from the previous section, but here we give a different formula based on character theory from representation theory. The *character* of a representation $\rho : G \to \mathrm{GL}(V)$ (where V is a finite-dimensional real vector space) is the function $\chi_V : G \to \mathbb{R}$ defined by $\chi_V(g) = \mathrm{tr}(\rho(g))$.

In our case, using the eigenvalues of $\rho_\ell(g)$ from the previous section, we have

$$\chi_{V_\ell}(g) = 1 + 2\sum_{m=1}^{\ell} \cos(m\,\phi(g))$$

where $\phi(g)$ is the angle of rotation of g. For $V = \oplus_{\ell\in\mathcal{F}} V_\ell$ we then have $\chi_V(g) = \sum_{\ell\in\mathcal{F}} \chi_{V_\ell}(g)$.

As a representation of $G = SO(3)$, $\mathbb{R}[\mathbf{x}]_d$ is (isomorphic to) the dth symmetric power of V, denoted $S^d(V)$. There is a recursive formula for the character of $S^d(V)$:

$$\chi_{S^d(V)}(g) = \frac{1}{d}\sum_{i=1}^{d} \chi_V(g^i)\chi_{S^{d-i}(V)}(g).$$

The representation $\mathbb{R}[\mathbf{x}]_d = S^d(V)$ decomposes as the direct sum of irreducible representations V_ℓ. The subspace of $\mathbb{R}[\mathbf{x}]_d$ consisting of all copies of the trivial representation V_0 (the 1-dimensional representation on which every group element acts as the identity) is precisely $\mathbb{R}[\mathbf{x}]_d^G$. Thus, $\dim(\mathbb{R}[\mathbf{x}]_d^G)$ is the number of copies of the trivial representation in the decomposition of $\mathbb{R}[\mathbf{x}]_d$. This can be computed using characters: $\dim(\mathbb{R}[\mathbf{x}]_d^G) = \langle \chi_{S^d(V)}, \chi_{V_0}\rangle = \langle \chi_{S^d(V)}, 1\rangle$ where $\langle f_1, f_2\rangle := \mathbb{E}_{g\sim\mathrm{Haar}(G)}[f_1(g)f_2(g)]$. Since characters are *class functions* (i.e. they are constant on conjugacy classes), we can compute this inner product by integrating over the angle ϕ (as in the previous section). This yields the formula stated in Proposition 5.5.

References

[ADBS16] Cecilia Aguerrebere, Mauricio Delbracio, Alberto Bartesaghi, and Guillermo Sapiro. Fundamental limits in multi-image alignment. *IEEE Trans. Signal Process.*, 64(21):5707–5722, 2016.

[ADLM84] Marc Adrian, Jacques Dubochet, Jean Lepault, and Alasdair W McDowall. Cryo-electron microscopy of viruses. *Nature*, 308(5954):32–36, 1984.

[APS17] Emmanuel Abbe, Joao Pereira, and Amit Singer. Sample complexity of the boolean multireference alignment problem. *arXiv preprint arXiv:1701.07540*, 2017.

[Ban15] Afonso S. Bandeira. *Convex Relaxations for Certain Inverse Problems on Graphs.* PhD thesis, Princeton University, June 2015.

[BBLS17] Nicolas Boumal, Tamir Bendory, Roy R. Lederman, and Amit Singer. Heterogeneous multireference alignment: a single pass approach. *arXiv preprint arXiv:1710.02590*, 2017.

[BBM$^+$17] Tamir Bendory, Nicolas Boumal, Chao Ma, Zhizhen Zhao, and Amit Singer. Bispectrum inversion with application to multireference alignment. *arXiv preprint arXiv:1705.00641*, 2017.

[BCS15] Afonso S. Bandeira, Yutong Chen, and Amit Singer. Non-unique games over compact groups and orientation estimation in cryo-em. *arXiv preprint arXiv:1505.03840*, 2015.

[BCSZ14] Afonso S. Bandeira, Moses Charikar, Amit Singer, and Andy Zhu. Multireference alignment using semidefinite programming. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 459–470. ACM, 2014.

[BMS13] Malte Beecken, Johannes Mittmann, and Nitin Saxena. Algebraic independence and blackbox identity testing. *Information and Computation*, 222:2–19, 2013.

[BRW17] Afonso Bandeira, Philippe Rigollet, and Jonathan Weed. Optimal rates of estimation for multi-reference alignment. *arXiv preprint arXiv:1702.08546*, 2017.

[CL11] T. Tony Cai and Mark G. Low. Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional. *Ann. Statist.*, 39(2):1012–1041, 2011.

[CLO07] David Cox, John Little, and Don O'Shea. *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra.* Springer New York, 2007.

[Dia92] R. Diamond. On the multiple simultaneous superposition of molecular structures by rigid body transformations. *Protein Science*, 1(10):1279–1287, October 1992.

[DK15] Harm Derksen and Gregor Kemper. *Computational invariant theory.* Springer, 2015.

[Eli57] Peter Elias. List decoding for noisy channels. 1957.

[ER93] Richard Ehrenborg and Gian-Carlo Rota. Apolarity and canonical forms for homogeneous polynomials. *European Journal of Combinatorics*, 14(3):157–181, 1993.

[Fel68] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley, 1968.

[Huỳ86] Dũng T. Huỳnh. A superexponential lower bound for Gröbner bases and Church-Rosser commutative Thue systems. *Inform. and Control*, 68(1-3):196–206, 1986.

[Kač94] Victor Kač. Invariant theory [lecture notes]. *https://people.kth.se/~laksov/notes/invariant.pdf*, 1994.

[Kam80] Zvi Kam. The reconstruction of structure from electron micrographs of randomly oriented particles. *Journal of Theoretical Biology*, 82(1):15–39, 1980.

[Kay09] Neeraj Kayal. The complexity of the annihilating polynomial. In *24th Annual IEEE Conference on Computational Complexity (CCC'09)*, pages 184–193. IEEE, 2009.

[LeC73] L. LeCam. Convergence of estimates under dimensionality restrictions. *Ann. Statist.*, 1:38–53, 1973.

[LNS99] O. Lepski, A. Nemirovski, and V. Spokoiny. On estimation of the L_r norm of a regression function. *Probab. Theory Related Fields*, 113(2):221–253, 1999.

[Nog16] Eva Nogales. The development of cryo-em into a mainstream structural biology technique. *Nature methods*, 13(1):24–27, 2016.

[NY83] A. S. Nemirovsky and D. B. and Yudin. *Problem complexity and method efficiency in optimization.* A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1983. Translated from the Russian and with a preface by E. R. Dawson, Wiley-Interscience Series in Discrete Mathematics.

[PWB$^+$17] Amelia Perry, Jonathan Weed, Afonso Bandeira, Philippe Rigollet, and Amit Singer. The sample complexity of multi-reference alignment. *arXiv preprint arXiv:1707.00943*, 2017.

[PWBM16a] Amelia Perry, Alexander S Wein, Afonso S. Bandeira, and Ankur Moitra. Message-passing algorithms for synchronization problems over compact groups. *arXiv preprint arXiv:1610.04583*, 2016.

[PWBM16b] Amelia Perry, Alexander S. Wein, Afonso S. Bandeira, and Ankur Moitra. Optimality and sub-optimality of PCA for spiked random matrices and synchronization. *arXiv:1609.05573*, 2016.

[PZAF05] R. G. Pita, M. R. Zurera, P. J. Amores, and F. L. Ferreras. Using multilayer perceptrons to align high range resolution radar signals. In *Artificial Neural Networks: Formal Models and Their Applications - ICANN 2005*, pages 911–916. Springer Berlin Heidelberg, 2005.

[Sad89] B. M. Sadler. Shift and rotation invariant object reconstruction using the bispectrum. In *Workshop on Higher-Order Spectral Analysis*, pages 106–111, Jun 1989.

[Sal79] Eugene Salamin. Application of quaternions to computation with rotations. Technical report, Working Paper, 1979.

[Sch03] Alexander Schrijver. *Combinatorial optimization: polyhedra and efficiency*. Springer Science & Business Media, 2003.

[SG92] Brian M Sadler and Georgios B Giannakis. Shift-and rotation-invariant object reconstruction using the bispectrum. *JOSA A*, 9(1):57–69, 1992.

[Sha94] Igor R. Shafarevich. *Basic algebraic geometry*. Springer, 1994.

[Sig16] Fred J. Sigworth. Principles of cryo-em single-particle image processing. *Microscopy*, 65(1):57–67, 2016.

[Sin11] Amit Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and computational harmonic analysis*, 30(1):20–36, 2011.

[SS11] Amit Singer and Yoel Shkolnisky. Three-dimensional structure determination from common lines in cryo-EM by eigenvectors and semidefinite programming. *SIAM Journal on Imaging Sciences*, 4(2):543–572, 2011.

[Stu08] Bernd Sturmfels. *Algorithms in invariant theory*. Springer Science & Business Media, 2008.

[TS12] D. L. Theobald and P. A. Steindel. Optimal simultaneous superpositioning of multiple structures with missing data. *Bioinformatics*, 28(15):1972–1979, 2012.

[Tsy09] Alexandre B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

[vdV98] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.

[Vve] Dimitri Vvedensky. *Chapter 8: Irreducible Representations of SO*(2) *and SO*(3). Group theory course [lecture notes]. `http://www.cmth.ph.ic.ac.uk/people/d.vvedensky/groups/Chapter8.pdf`.

[ZvdHGG03] J. P. Zwart, R. van der Heiden, S. Gelsema, and F. Groen. Fast translation invariant classification of hrr range profiles in a zero phase representation. *Radar, Sonar and Navigation, IEE Proceedings*, 150(6):411–418, 2003.