Complete and Efficient Covariants for 3D Point Configurations

Hartmut Maennel HARTMUTM@GOOGLE.COM
Google DeepMind Zürich, Brandschenkestr. 110, 8002 Zürich, Switzerland

Editors: Michael Bleher, Freya Jensen, Levin Maier, Diaaeldin Taha, and Anna Wienhard

Abstract

We investigate the question: "How can we efficiently describe equivalence classes of finite sets of (colored) points in $\mathbb{R}^3$, where (colored) point sets are equivalent if they can be transformed into each other by a rotation?" It sounds very simple, but we will see it leads to some interesting mathematical structures. However, they only become a part of the picture when we have to take into account some application specific constraints: We want to characterize these configurations by features that do not depend on the number of points in the set, and that are fast to evaluate.

1. Introduction: $O(3)$–invariant set functions on $\mathbb{R}^3$

I would like to talk about the mathematical aspects of work done in Quantum Chemistry at Google DeepMind together with Oliver Unke and Klaus–Robert Müller, the main reference is our paper Maennel et al. (2024). This work lead to the question posed in the abstract.

1.1. Application in quantum chemistry

To motivate these constraints I describe a bit of the application, although it is not strictly necessary for the mathematics that follows.
Our practical application is for quantum mechanically accurate simulations of the dynamics of molecular systems. This means: Given the positions and types (i.e. elements) of n atoms (in some molecular system), we want to predict the energy of that system as a function of the positions of the n atoms. In particular, when we take the gradient of this energy with respect to the position of an atom, we get the (negative) force on that atom. And using our approximation of the forces, we then can simulate the evolution of the system by iterating over small time steps, giving us a trajectory for the dynamics of this molecular system.

We use the assumption that the energy can be modeled well from the local interactions, so the main term in our model of the energy is a sum over contributions from each atom, that takes into account only those other atoms whose distance to the central atom is below a threshold. The number of atoms within this relevant radius varies, but as a ballpark estimate we could say the numbers of atoms goes up to 20. (See appendix A.1 for a bit more details.)

With this locality assumption it will be enough to have a lot of training examples of the energy of smallish systems of atoms, and then we will train a neural network that gives a

good approximation to energy contributions around each atom and hope that this generalizes to other molecular systems.
Force fields obtained in this way are called "Machine Learned Force Fields", and they have proven to be a valuable alternative to more costly DFT calculations or to less accurate classical force fields, see e.g. Thomas et al. (2018); Anderson et al. (2019); Fuchs et al. (2020); Satorras et al. (2021); Unke et al. (2021a); Schütt et al. (2021); Unke et al. (2021b); Batzner et al. (2022); Frank et al. (2022); Musaelian et al. (2023); Unke et al. (2024).

To get good neural network models, we want to build in physical knowledge, in particular: The energy contribution E around one atom should be invariant under

- rotations and reflections that leave the central atom invariant,
- permutation of nuclei with the same charge Z_i.

This now leads us to the question which input features and network architectures with built in invariance can guarantee unique characterization, i.e. that for any two input lists that are not related by such a symmetry operation, the input features are able to distinguish between these configurations.

1.2. The mathematical problem for a fixed number of points

Let us first consider the abstract problem: Given a set of "colored points"

$$\{(\boldsymbol{r}_i, Z_i) \,|\, i = 1, ..., n\} \qquad \text{with } \boldsymbol{r}_i \in \mathbb{R}^3, Z_i \in \{c_1, c_2, ..., c_{max}\},$$

come up with useful features that uniquely characterize this set up to rotations and permutations.

To keep this section simple, we specialize to the case of one color (the general case is not substantially different). This means we look for features of $\mathbb{R}^{3n}/(O(3) \times \Sigma)$ where Σ is the permutation group of n elements.
We look at this as $\left(\mathbb{R}^{3n}/O(3)\right)/\Sigma$. For the first quotient, note that the list of $n(n-1)/2$ distances between all pairs of $\boldsymbol{r}_0 = 0, \boldsymbol{r}_1, ..., \boldsymbol{r}_{n-1}$ determines the configuration up to $O(3)$. This list depends on the order of the n points, if we permute the points, we get a corresponding permutation of the $n(n-1)/2$ distances, and we have to determine invariants under this action of Σ on $\mathbb{R}^{n(n-1)/2}$.

This problem is "solved" by Hilbert's basis theorem: We can find a finite list of polynomials that generate the ring of invariant functions

$$\mathbb{R}[D]^{\Sigma} := \mathbb{R}[d_{0,1}^2, d_{0,2}^2, ..., d_{n-2,n-1}^2]^{\Sigma}.$$

While one can make this existence theorem constructive, the number of invariants is just too large to be of practical use (see appendix A.2 for some more details).

1.3. Representation theory arises naturally

Apart from the size of their feature sets, these solutions have one other flaw: In our case, the size n of the point set is *not fixed*, we actually want to learn features that make sense for any n and ideally also generalize across different n!
Can we formulate $O(3)$–invariant functions on finite point sets *independent of size*?

For simplicity, let us restrict our point sets even more and focus on point sets on the unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$. So individual points do not carry $O(3)$–invariant information.

We can indeed give such size agnostic, $O(3) \times \Sigma_n$–invariant features: $F(S) := |\sum_{\boldsymbol{r}\in S} \boldsymbol{r}\,|$ depends only on the set of unit vectors and contains useful information.
Can we generalize this to invariants of the form $F(S) := |\sum_{\boldsymbol{r}\in S} f(\boldsymbol{r})|$ for some interesting functions f?
For this to still make sense, we must have

1. We can add up the $f(\boldsymbol{r})$ and take a "length":
   This works if the $f(\boldsymbol{r})$ are in some Euclidean vector space V.

For this to also be $O(3)$–invariant, we want for all $g \in O(3)$

$$F(gS) = \left|\sum_{\boldsymbol{r}\in S} f(g\boldsymbol{r})\right| \overset{4}{=} \left|\sum_{\boldsymbol{r}\in S} \rho(g) f(\boldsymbol{r})\right| \overset{3}{=} \left|\rho(g) \sum_{\boldsymbol{r}\in S} f(\boldsymbol{r})\right| \overset{2}{=} \left|\sum_{\boldsymbol{r}\in S} f(\boldsymbol{r})\right| = F(S)$$

which is true when these three conditions are fulfilled:

2. $O(3)$ acts on V and conserves lengths: $\rho : O(3) \to O(V)$
3. The action is linear (we need it to be compatible with $+$; with continuity this is enough to imply $\mathbb{R}$–linearity).
4. $f : \mathbb{R}^3 \to V$ is $O(3)$–equivariant.

This means ρ is an orthogonal representation $O(3) \to O(V)$ and $f : \mathbb{R}^3 \to V$ is equivariant with respect to the canonical action of $O(3)$ on $\mathbb{R}^3$ and the action via ρ on V. (For a summary of all representation theory needed for this application, see e.g. Unke and Maennel (2024).)

In fact, we can even argue that ρ should be an *irreducible* representation, because otherwise we waste information: If we have an invariant subspace V_1 of (V, ρ), then the orthogonal complement V_2 is again an invariant subspace, $(V, \rho) = (V_1, \rho_1) \oplus (V_2, \rho_2)$, and we can write the vectors $f(\boldsymbol{r})$ as sum of the two parts $f_1(\boldsymbol{r}) \in V_1$, $f_2(\boldsymbol{r}) \in V_2$. Therefore we have

$$F(S)^2 \;:=\; \left|\sum_{\boldsymbol{r}\in S} \begin{pmatrix} f_1(\boldsymbol{r}) \\ f_2(\boldsymbol{r}) \end{pmatrix}\right|^2 = \left|\sum_{\boldsymbol{r}\in S} f_1(\boldsymbol{r})\right|^2 + \left|\sum_{\boldsymbol{r}\in S} f_2(\boldsymbol{r})\right|^2$$

so we do not get anything new that we did not already get by using the smaller $O(3)$–invariant subspaces V_1, V_2, so we should rather look at these parts separately.

Irreducible representations of $O(3)$:
To catalog all invariants we get in this way, we start with an inventory of irreducible representations.
Up to isomorphism there exists exactly one irreducible representation of $SO(3)$ with dimension $2L+1$ for $L = 0, 1, 2, ...$; we denote it by $\mathcal{H}^{(L)}$, and these are all the irreducible representations that exist for $SO(3)$. They can be described as the vector space of homogeneous harmonic polynomials of degree L in three variables.
To extend these to irreducible representations of $O(3)$ we have two possibilities: We set

$$\rho(-Id_3) = Id_V \qquad \text{or} \qquad \rho(-Id_3) = -Id_V.$$

We call ρ *even* in the first case and *odd* in the second case.

Equivariant maps $\mathbb{S}^2 \to \mathbb{R}^{2L+1}$
There is only one map (up to a multiplicative constant and the choice of an orthonormal basis of $\mathbb{R}^{2L+1}$)

$$Y_L : \mathbb{S}^2 \to \mathbb{R}^{2L+1}$$

that is equivariant under $SO(3)$. It is also $O(3)$ equivariant if we extend the representation ρ_L such that it is even for even L or odd for odd L.
These equivariant maps $Y_L : \mathbb{S}^2 \to \mathbb{R}^{2L+1}$ are the (real valued) spherical harmonics, they can be given by homogeneous polynomials of deg L.

This gives *infinitely many* different $O(3)$–invariant features $I_L(S) := |\sum_{\boldsymbol{r}\in S} Y_L(\boldsymbol{r})|$.
So with all these features... does that mean our problem is solved?

1.4. The uniqueness problem

It is easy to see that all these invariants depend only on the *set* of angles, see appendix A.4. This means we can run into a problem if we can find different configurations that have the same set of angles. Unfortunately, in Pozdnyakov et al. (2020) the authors proved that this potential problem is in fact real! (See appendix A.5 for some counterexamples.) Pozdnyakov et al. (2020) call invariants that only depend on the set of angles invariants of "body order 3", because they only depend on the set of triples $A, 0, B$ up to isometries for A, B in our point set on $\mathbb{S}^2$ and 0 the origin.

So we have to come up with invariants of higher body order... and this is indeed possible, for example:

$$\sum_{\boldsymbol{r}_1,\boldsymbol{r}_2,\boldsymbol{r}_3\in S} \langle \boldsymbol{r}_1 \times \boldsymbol{r}_2, \boldsymbol{r}_3\rangle^2 = \sum_{\boldsymbol{r}_1,\boldsymbol{r}_2,\boldsymbol{r}_3\in S} \det(\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3)^2$$

This is an invariant of "body order 4", it depends on the *set of triples* $(\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3)$.
But this alone is also not enough, Pozdnyakov et al. (2020) prove that *even body order 4 is not enough* to distinguish all inequivalent configurations!

So now the stage is set for the fundamental question: How do we get enough invariants (that work for all number of points and can be evaluated quickly) to uniquely identify all configurations?

1.5. Mathematical framework for invariants of finite point sets

We look at functions on "colored point sets", i.e. on finite point sets $S = \{\boldsymbol{r}_1, ..., \boldsymbol{r}_n\} \subset \mathbb{R}^3$ with colors $\gamma_1, ..., \gamma_n$ associated to them in a fixed finite set $\mathcal{C}$ of possible colors. Since we want to learn on and generalize to different n, we keep n *variable*. This has the consequence that our configuration space has to be an *infinite–dimensional* "algebraic variety". We define it as the disjoint union of all

$$Conf(n) := \bigsqcup_{\gamma:\{1,2,...,n\}\to\mathcal{C}} \mathbb{R}^{3n}/\Sigma_\gamma \qquad \text{for } n = 1, 2, ...$$

which are quotients of $\mathbb{R}^{3n}$ by the groups Σ_γ which contain the permutations $\sigma : \{1, 2, ..., n\} \to \{1, 2, ..., n\}$ that only permute points of the same color, i.e. $\gamma(\sigma(i)) = \gamma(i)$, and the disjoint sum is over those color assignments $\gamma : \{1, 2, ..., n\} \to \mathcal{C}$ that satisfy $\gamma(1) \le \gamma(2) \le ... \le \gamma(n)$ for some (arbitrary, but fixed) total order on the color set $\mathcal{C}$.

We define **polynomial point set descriptors (PPSDs)**:
For a function f of k points in $\mathbb{R}^3$ we get a function on configurations of any size by:

$$\sum_{\boldsymbol{r}_1 \in S_{\gamma_1}} ... \sum_{\boldsymbol{r}_k \in S_{\gamma_k}} f(\boldsymbol{r}_1, ..., \boldsymbol{r}_k) \qquad \text{with} \quad S_{\gamma_i} := \{\boldsymbol{r} \in S \,|\, \text{color is } \gamma_i\}$$

We call these functions of *body order* $k+1$. We will in particular consider *polynomial* point set descriptors (PPSDs), which we define as the linear combinations of such functions for *polynomials* f. (Later we will generalize this to allow analytic radial functions.)

We can ask how this differs from just considering different polynomials for each size n.
This is answered by proposition 17 in Maennel et al. (2024):

- The values on $Conf(n)$ for *finitely many* n can be prescribed by arbitrary polynomials of n variables that are symmetric under Σ_γ, but...
- ...from some N on all values on $Conf(n)$ for $n > N$ can be inferred from the values on $Conf(N)$ by a certain rule.

This is exactly what we need to "generalize to higher n".

Fundamental features:
One can show that the PPSDs are closed under multiplication, and this ring of functions is generated by the $\sum_{\boldsymbol{r}\in S_\gamma} P(\boldsymbol{r})$ with polynomial P. We call these the *fundamental features*, they can be computed in time $O(n)$ for sets of n points.

This means to compute any PPSDs we can:
1) Compute some necessary fundamental features $\sum_{\boldsymbol{r}\in S_\gamma} P(\boldsymbol{r})$,
2) Compute the appropriate polynomial in these fundamental features.
With this algorithm any PPSD can be computed in time *linear* in $n = |S|$. This makes $PPSD$s particularly attractive as features that can be evaluated effectively.
(See appendix A.3 for some more remarks on this definition of PPSDs.)

2. Polynomial functions: Completeness / finiteness

2.1. Topological Completeness

We now summarize the results of Maennel et al. (2024) about the characterization of finite point sets by PPSDs. The first theorem reassures us that in principle PPSDs are a "complete" feature set, i.e. they are enough to uniquely characterize point sets.

Theorem 1: Let G be $O(3)$ or $SO(3)$.

1. Any two different configurations can be distinguished by their fundamental features.
2. Any two configurations that are not *equivalent under* G can be distinguished by G*–invariant* PPSDs.

See appendix A.6 for a version for covariant functions and the proof ideas.

2.2. Finiteness

For a given upper bound on the number of points, we can get from the theoretical existence of the previous section to specific *finite* sets of invariants.

Theorem 2: Let G be $O(3)$ or $SO(3)$.

1. Given $n_1, ..., n_c \in \mathbb{N}$, there are *finitely many* G–invariant PPSDs that distinguish all G–equivalence classes of configurations with at most $n_1, ..., n_c$ points of colors $1, ..., c$.
2. Given such a *finite set* of G–invariant PPSDs that distinguishes all G–equivalence classes of colored point sets with at most $n_1, ..., n_c$ points of colors $1, ..., c$, and $n := n_1 + ... + n_c$, there are $6n-5$ *linear combinations* of them that already distinguish all equivalence classes.

Theorem 1 only said that any pair of non–equivalent configurations can be separated, here we say this can be done for all pairs of n–point configurations simultaneously with finitely many invariant functions.
This is what we really need for Machine Learning: We will only have finitely many features!

Proof idea for part 1:
Use the lemma on the moments as in 1. above.

Proof idea for part 2:
We mimic the proof for the weak Whitney embedding theorem:
Let $\mathcal{X}$ be the set of G–equivalence classes of colored point sets with $|S_1|, ..., |S_c|$ points of colors $1, 2, ..., c$. Let $\mathcal{F}$ be a finite dimensional vector space of functions on $\mathcal{X}$ that separate points.

1. $\mathcal{X}$ has dimension $3n - 3$.

2. Take functions $f_1, ..., f_N \in \mathcal{F}$ that separate points of $\mathcal{X}$. This defines an embedding $f : \mathcal{X} \hookrightarrow \mathbb{R}^N$.
3. Show that the set $D \subseteq \mathbb{P}^{N-1}$ of directions in $\mathbb{R}^N$ given by two different points $f(\boldsymbol{x}_1), f(\boldsymbol{x}_2)$ has a closure $\overline{D}$ of dimension $\dim(\overline{D}) \leq 2 \cdot \dim(\mathcal{X})$.
4. If $N - 1 > 2 \cdot \dim(\mathcal{X})$, then there is a linear map $\mathbb{R}^N \to \mathbb{R}^{N-1}$ that is injective on $\mathcal{X}$ (and in fact only decreases distances by a factor of at most C for some constant $C > 0$).

Then start with the embedding from 1, use 2 to decrease n as long as $n > 2 \cdot \dim(\mathcal{X}) + 1$.

To make this proof idea precise, we need to find appropriate notions of functions / sets / dimensions!
Here is a sketch of the notions / facts used for a precise proof:

- $\mathcal{X}$ is an orbifold (locally a manifold divided by a finite group).
- $\mathcal{X}, \mathcal{X} \times \mathcal{X} \setminus \Delta \to D, D, \overline{D}$ are semi–algebraic: Given by finitely many algebraic conditions $f(\boldsymbol{x}) = 0$ and $g(\boldsymbol{x}) \geq 0$ (combined by Boolean operations). This uses in particular that the image of a semi–algebraic map is semi–algebraic, this is also known as "Quantifier elimination for $(\mathbb{R}, +, \cdot)$".
- Semi–algebraic sets have dimensions which agree with the dimension of manifolds / orbifolds.
- Use Theorem: $\dim\left(\overline{f(M)}\right) \leq \dim(M)$ for M, f semi–algebraic.

Why do we need "semi–algebraic", isn't this just a topological argument?
No, in particular $\dim(f(X)) \leq \dim(X)$ is wrong for arbitrary continuous functions, as is witnessed by space–filling curves like the Hilbert curve, a continuous surjective function $[0,1] \to [0,1] \times [0,1]$.
Even for smooth maps (but not for polynomial or more generally semi-algebraic maps!) $\dim\left(\overline{f(X)}\right) \leq \dim(X)$ can fail: We can map a line to a torus such that the image is dense: For $f : \mathbb{R} \to \mathbb{R}^4,\ t \mapsto (\sin(t), \cos(t), \sin(\alpha t), \cos(\alpha t))$ we have $\dim(\mathbb{R}) = 1,\ \dim\left(\overline{f(X)}\right) = 2$ for $\alpha \notin \mathbb{Q}$.

3. Polynomial functions: $SO(3)$ harmonic analysis

So far we looked at theoretical results about the existence of features. We now move towards a particular set of invariant / covariant features that can be computed efficiently, which generalize our use of the representation theory of $SO(3)$ from the introduction.

The usual "$SO(3)$–harmonic analysis" can be seen as the study of the representation of $SO(3)$ on the vector space of polynomials on the sphere $\mathbb{S}^2$. We want to go from functions of points to functions of point sets, i.e. we describe here the representation of $SO(3)$ on the vector space of PPSDs on $\mathbb{S}^2$.

3.1. Harmonic analysis of fundamental features

A result of the usual $SO(3)$–harmonic analysis is that any homogeneous polynomial of degree L on $\mathbb{R}^3$ can be written uniquely as

$$P(\boldsymbol{r}) = h_L(\boldsymbol{r}) + |\boldsymbol{r}|^2 \cdot h_{L-2}(\boldsymbol{r}) + \ldots + \begin{cases} |\boldsymbol{r}|^{L-1} \cdot h_1(\boldsymbol{r}) & \text{for } L \text{ odd,} \\ |\boldsymbol{r}|^L \cdot h_0(\boldsymbol{r}) & \text{for } L \text{ even.} \end{cases}$$

with h_l homogeneous harmonic polynomials of degree l, and the spherical harmonics Y_l^m are a basis of the vector space of homogeneous harmonic polynomials of degree l.

We can transfer this to fundamental features given by a polynomial P and a color γ as $\sum_{\boldsymbol{r}\in S_\gamma} P(\boldsymbol{r})$: The fundamental features have a basis consisting of

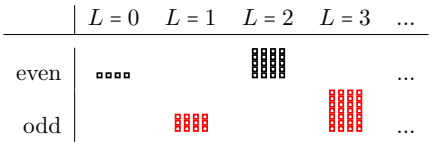
$$f_{\gamma,k,l,m}(\boldsymbol{r}) := \sum_{\boldsymbol{r}\in S_\gamma} |\boldsymbol{r}|^{2k} \cdot Y_l^m(\boldsymbol{r})$$

and for fixed l these span the $\mathcal{H}^{(l)}$–isotypical component of the vector space generated by the fundamental features.

We can extend this to $O(3)$ representations: We have $O(3) = SO(3) \cup (-Id) \cdot SO(3)$ and irreducible representations $\rho : O(3) \to GL(V)$ are irreducible representations of $SO(3)$ on V + "parity":

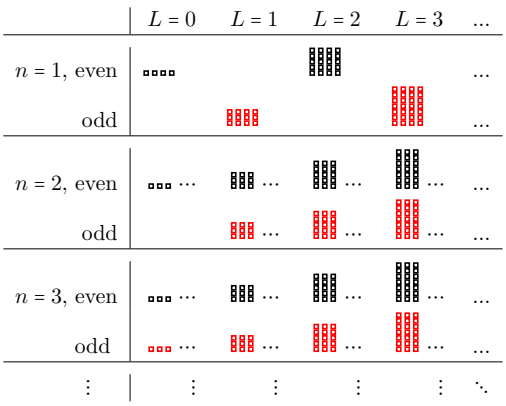
$$\textbf{even}\text{: } \rho(-Id) = +Id_V \ , \qquad \textbf{odd}\text{: } \rho(-Id) = -Id_V .$$

Denote the irreducible $O(3)$ representations by $\mathcal{H}^{(l,\epsilon)}$ with $\epsilon = +1/-1$ for even / odd representations. Our fundamental features of degree L are already $O(3)$–representations, even / odd when L is even / odd.



3.2. Harmonic analysis of PPSDs

General PPSDs are polynomials in the fundamental features. We can decompose these into homogeneous polynomials of degree n in fundamental features, and then decompose them further by isotypical components of $\mathcal{H}^{(L)}$, this gives the following fundamental diagram. Each small square is a basis function, groups of $(2L+1)$ scalar basis functions in $\mathcal{H}^{(L)}$ combine to one $\mathcal{H}^{(L)}$–valued function.



3.3. Algebraic completeness (Clebsch–Gordan version)

The abstract Clebsch–Gordan theorem for $SO(3)$ is the isomorphism

$$\mathcal{H}^{(l_1)} \otimes \mathcal{H}^{(l_2)} \quad \simeq \quad \mathcal{H}^{(|l_1-l_2|)} \oplus \mathcal{H}^{(|l_1-l_2|+1)} \oplus \ldots \oplus \mathcal{H}^{(l_1+l_2)} \qquad \text{for} \quad |l_1 - l_2| \leq l_3 \leq l_1 + l_2.$$

We can extended it to $O(3)$ for parities $\epsilon_1, \epsilon_2 \in \{\pm 1\}$:

$$\mathcal{H}^{(l_1,\epsilon_1)} \otimes \mathcal{H}^{(l_2,\epsilon_2)} \quad \simeq \quad \bigoplus_{|l_1-l_2| \,\leq\, l_3 \,\leq\, l_1+l_2} \mathcal{H}^{(l_3,\,\epsilon_1\epsilon_2)}$$

Choosing a basis in $\mathcal{H}^{(l_1)}, \mathcal{H}^{(l_2)}, \mathcal{H}^{(l_3)}$ gives a "Clebsch–Gordan product"

$$\mathbb{R}^{2l_1+1,\epsilon_1} \otimes \mathbb{R}^{2l_2+1,\epsilon 2} \to \mathbb{R}^{2l_3+1,\epsilon_1\epsilon_2} \qquad \text{for} \quad |l_1 - l_2| \leq l_3 \leq l_1 + l_2.$$

Using the Clebsch–Gordan products, we can build up the above schema by starting from the fundamental features for $n = 1$ as above, and then fill in row $n+1$ by all Clebsch–Gordan products between row 1 and row n.

Theorem 3:

1. All scalar PPSDs are some linear combination of fields in this schema.
2. Any $SO(3)$*–covariant* PPSD with values in $\mathcal{H}^{(l)}$ is a linear combination of vectors in the l–th column.
3. Any $O(3)$*–covariant* PPSD with values in $\mathcal{H}^{(l,\epsilon)}$ is a linear combination of vectors in the l–th column of the appropriate parity.

This "Clebsch–Gordan version of algebraic completeness" corresponds to the most common approach of constructing invariants / covariants. There also is an older approach to constructing invariants which uses full tensor products and contractions instead of Clebsch–Gordan operations, see Shapeev (2016). In this paper, classical invariant theory was used to prove a completeness theorem. In Maennel et al. (2024) we also give a streamlined version of this flavor of algebraic completeness using our PPSDs.

4. Analytic radial basis functions

Points $\boldsymbol{r} \in \mathbb{R}^3$ can be written in "polar coordinates" as $\boldsymbol{r} = |\boldsymbol{r}| \cdot \boldsymbol{u}$ for $\boldsymbol{u} := \boldsymbol{r}/|\boldsymbol{r}| \in \mathbb{S}^2$, and we can give special functions f on $\mathbb{R}^3$ by an "angular" function g on $\mathbb{S}^2$ and a "radial" function h on $\mathbb{R}_{\geq 0}$:

$$f(\boldsymbol{r}) := h(|\boldsymbol{r}|) \cdot g(\boldsymbol{r}/|\boldsymbol{r}|).$$

To be well defined at the origin, we should also require that either $g(\boldsymbol{u})$ is constant or $h(0) = 0$. However, in practice this does not matter to us since we will always assume that the distance to the central nucleus is $> \epsilon$ for any other nucleus, so we will only use values $r > \epsilon$ for $h(r)$.

When $h(r)$ is a polynomial in r^2 and $g(\boldsymbol{u})$ is a polynomial, then this will again be a polynomial on $\mathbb{R}^3$, and any other polynomial can be approximated by such polynomials (uniformly on compact subsets of $\mathbb{R}^3$).

While in theory polynomial functions are enough to supply all features, in practice they are not optimal for our purposes: We expect that when atoms move far away from the central atom, they will contribute less and less to the energy contribution assigned to the central atom. So we will get features better aligned with the chemistry application when we use radial functions $h(r)$ that decrease to 0 for $r \to \infty$.

Almost all of the above theory remains valid if we make such a generalization, in particular since for the $O(3)$ operation only the angular part is interesting. So we only comment in the appendix A.9 on two details for which the above statements have to be modified.

5. Speeding up Clebsch–Gordan operations

To efficiently use the Clebsch–Gordan products, we bundle different irreducible representations for $l = 0, 1, ..., L-1$ into one array of $1 + 3 + ... + (2L-1) = L^2$ numbers. Then we can apply all Clebsch–Gordan products simultaneously to get from one bundle of body order $(n+1)$ and one of body order $(m+1)$ to a new one of body order $(n+m+1)$. This uses a tensor of order 3 which contains $(L^2)^3 = L^6$ Clebsch–Gordan coefficients, and takes time $O(L^6)$.

All combinations for $0 \leq l_1, l_2 < L$

$$\mathcal{H}^{(l_1)} \otimes \mathcal{H}^{(l_2)} \to \mathcal{H}^{(l_3)} \qquad \text{for } |l_1 - l_2| \leq l_3 \leq \min(l_1 + l_2, L-1)$$

are used and added up with learnable parameters w_{l_1,l_2,l_3}.

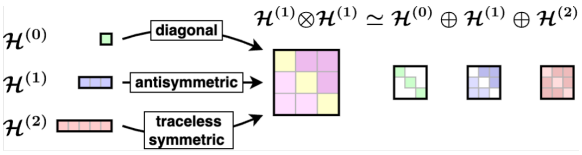
This scaling by $O(L^6)$ prevents the use of larger L in practice, which could give finer angular resolutions. We now describe a way to reduce this $O(L^6)$ scaling to a $O(L^3)$ scaling which makes the use of higher L feasible. The resulting networks are not in a one–to–one equivalence to particular networks using Clebsch–Gordan operations, but can express the same features and satisfy the same algebraic completeness.

5.1. Replacing Clebsch–Gordan operations by matrix multiplication

For even L, the "backwards" Clebsch–Gordan equation

$$\mathcal{H}^{(0)} \oplus \mathcal{H}^{(1)} \oplus ... \oplus \mathcal{H}^{(L)} \simeq \mathcal{H}^{(L/2)} \otimes \mathcal{H}^{(L/2)}$$

shows that we can also encode the bundle $\mathcal{H}^{(0)} \oplus \mathcal{H}^{(1)} \oplus ... \oplus \mathcal{H}^{(L)}$ as one $(L+1) \times (L+1)$ matrix.



Matrix multiplication is covariant, so this gives another way to construct covariants of $l = 0, 1, ..., L$ and body order $(n + m + 1)$ from those of $l = 0, 1, ..., L$ and body order $n + 1$ and $m + 1$.

And matrix multiplication is now $O(L^3)$ instead of $O(L^6)$!
We still have to do the linear transformation from one representation as a vector of length L^2 to another vector of length L^2, but that is still $O(L^4)$ instead of $O(L^6)$.
(We can also get this down to $O(L^3)$ if we use computations with complex numbers.)
Possible downside: We have less parameters, they now only go into the inputs, not in the product.

In the open source software "E3x" (see Unke and Maennel (2024)), this matrix product is implemented as a drop in replacement for the Clebsch–Gordan operations with lower computational complexity and fewer learnable parameters. To achieve that, the list of features in $\mathcal{H}^{(0)} \oplus ... \oplus \mathcal{H}^{(L)}$ is first encoded in a matrix, then the matrix multiplication is computed, and then the product is again decomposed into a list of features in $\mathcal{H}^{(0)} \oplus ... \oplus \mathcal{H}^{(L)}$. This can be used for $SO(3)$–features and also for $O(3)$–features, for which we have to distinguish also between even and odd features.

Experimentally we found that this can speed up significantly the computations on CPUs, but does not seem to benefit computations on GPUs / TPUs unless we go to large degrees L in the representations.
To get a better speed up on GPUs / TPUs, we want to avoid going back and forth between list of features belonging to different $\mathcal{H}^{(l)}$ and their matrix form, but rather keep the matrix products in their matrix form. Also we want to avoid computing many matrix products with small matrices and rather use fewer matrix products with large matrices. To do so we partition a large square matrix into a number of smaller rectangular matrices. This is not directly one–to–one equivalent to a particular computation using many small matrices, but gives the same class of features and the same completeness property holds. (See appendix A.10 for details.)

6. Summary

We motivated a mathematical question by its application in quantum chemistry: How can we represent finite sets of (colored) points in $\mathbb{R}^3$ up to rotations by (efficiently computable) features that are independent of the number of points? Such features make it possible in principle to use machine learning for rotationally invariant functions on point sets and generalize from smaller to larger point sets. We described a set of functions called "Polynomial Point Set Descriptors (PPSDs)" that can be used for this purpose.
We described the following results:

- Topological completeness: Let $G = SO(3)$ or $G = O(3)$. Any two (not G–equivalent) configurations can be distinguished by a G–invariant PPSD.
- Finiteness, 1: There is *a finite set* of G–invariant PPSDs that can distinguish between the configurations in *any* such pair of configurations with n points.
- Finiteness, 2: Given such a set of PPSDs, we can find $6n - 5$ linear combinations of them that can already distinguish between such configurations of n points.
- Algebraic completeness: We get all G–invariant PPSDs starting from fundamental features (sum of spherical harmonics) by applying Clebsch–Gordan products.
- The same results hold when we replace polynomial functions by products of a polynomial angular part and an analytic radial function.

We also described the following methods for an efficient implementation of such features:

- Certain lists of features can be encoded in matrices, and then Clebsch–Gordan products can be replaced by the more efficient matrix product.
- When using matrix product features in implementations on GPUs/TPUs, the computation is further sped up by combining many small rectangular matrices into one large square matrix.

The fully precise formulations and proofs can be found in Maennel et al. (2024).

References

Brandon Anderson, Truong Son Hy, and Risi Kondor. Cormorant: Covariant molecular neural networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32, pages 1–10. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/03573b32b2746e6e8ca98b9123f2249b-Paper.pdf.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1):2453, 2022.

Thorben Frank, Oliver Unke, and Klaus-Robert Müller. So3krates: Equivariant attention for interactions on arbitrary length-scales in molecular systems. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 29400–29413. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/bcf4ca90a8d405201d29dd47d75ac896-Paper-Conference.pdf.

Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. SE(3)-transformers: 3D roto-translation equivariant attention networks. In *Advances in neural information processing systems*, volume 33, pages 1970–1981, 2020.

Stefan Ganscha, Oliver T. Unke, Daniel Ahlin, Hartmut Maennel, Sergii Kashubin, and Klaus-Robert Müller. The QCML dataset, quantum chemistry reference data from 33.5m DFT and 14.7b semi-empirical calculations. *Scientific Data*, 12(1):406, 2025. doi: 10.1038/s41597-025-04720-7. URL https://doi.org/10.1038/s41597-025-04720-7.

Snir Hordan, Tal Amir, Steven J. Gortler, and Nadav Dym. Complete neural networks for complete euclidean graphs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(11):12482–12490, 3 2024. doi: 10.1609/aaai.v38i11.29141. URL https://ojs.aaai.org/index.php/AAAI/article/view/29141.

Vitaliy Kurlin. Simplexwise distance distributions for finite spaces with metrics and measures. arXiv, 03 2023. URL https://doi.org/10.48550/arXiv.2303.14161.

Hartmut Maennel, Oliver T. Unke, and Klaus-Robert Müller. Complete and efficient covariants for three-dimensional point configurations with application to learning molecular quantum properties. *The Journal of Physical Chemistry Letters*, 15(51):12513–12519, 2024. doi: 10.1021/acs.jpclett.4c02376. URL https://doi.org/10.1021/acs.jpclett.4c02376.

Albert Musaelian, Simon Batzner, Anders Johansson, Lixin Sun, Cameron J Owen, Mordechai Kornbluth, and Boris Kozinsky. Learning local equivariant representations for large-scale atomistic dynamics. *Nature Communications*, 14(1):579, 2023.

L. Pebody, A. J. Radcliffe, and A. D. Scott. Finite subsets of the plane are 18-reconstructible. *SIAM Journal on Discrete Mathematics*, 16(2):262–275, 2003. URL https://doi.org/10.1137/S0895480101391648.

Sergey N Pozdnyakov, Michael J Willatt, Albert P Bartók, Christoph Ortner, Gábor Csányi, and Michele Ceriotti. Incompleteness of atomic structure representations. *Physical Review Letters*, 125(16):166001, 2020.

Valentino Delle Rose, Alexander Kozachinskiy, Cristóbal Rojas, Mircea Petrache, and Pablo Barceló. Three iterations of $(d-1)$-wl test distinguish non isometric clouds of d-dimensional points. arXiv, 2023. URL https://doi.org/10.48550/arxiv.2303.12853.

Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 9323–9332. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/satorras21a.html.

Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *Proceedings of the 38th International Conference on Machine Learning*, pages 9377–9388. PMLR, 2021.

Alexander V Shapeev. Moment tensor potentials: A class of systematically improvable interatomic potentials. *Multiscale Modeling & Simulation*, 14(3):1153–1173, 2016.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. arXiv, 2018. URL https://doi.org/10.48550/arXiv.1802.08219.

Oliver Unke, Mihail Bogojeski, Michael Gastegger, Mario Geiger, Tess Smidt, and Klaus-Robert Müller. SE(3)-equivariant prediction of molecular wavefunctions and electronic densities. In *Advances in Neural Information Processing Systems*, volume 34, pages 14434–14447, 2021a.

Oliver T. Unke and Hartmut Maennel. E3x: $E(3)$-equivariant deep learning made easy. arXiv, 2024. URL https://doi.org/10.48550/arXiv.2401.07595.

Oliver T Unke, Stefan Chmiela, Michael Gastegger, Kristof T Schütt, Huziel E Sauceda, and Klaus-Robert Müller. Spookynet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nature communications*, 12(1):7273, 2021b.

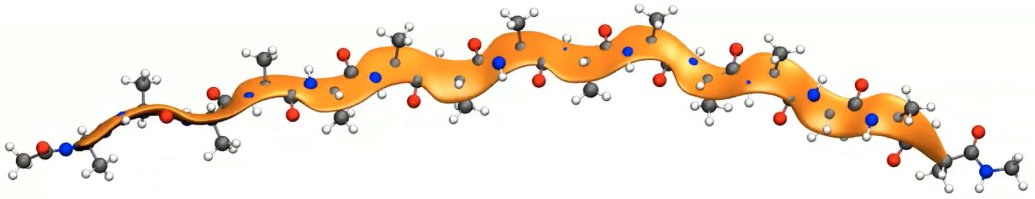
Oliver T. Unke, Martin Stöhr, Stefan Ganscha, Thomas Unterthiner, Hartmut Maennel, Sergii Kashubin, Daniel Ahlin, Michael Gastegger, Leonardo Medrano Sandonas, Joshua T. Berryman, Alexandre Tkatchenko, and Klaus-Robert Müller. Biomolecular dynamics with machine-learned quantum-mechanical force fields trained on diverse chemical fragments. *Science Advances*, 10(14):eadn4397, 2024. doi: 10.1126/sciadv.adn4397. URL https://www.science.org/doi/abs/10.1126/sciadv.adn4397.

Daniel Widdowson and Vitaliy Kurlin. Recognizing rigid patterns of unlabeled point clouds by complete and continuous isometry invariants with no false negatives and no false positives. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1275–1284, June 2023.

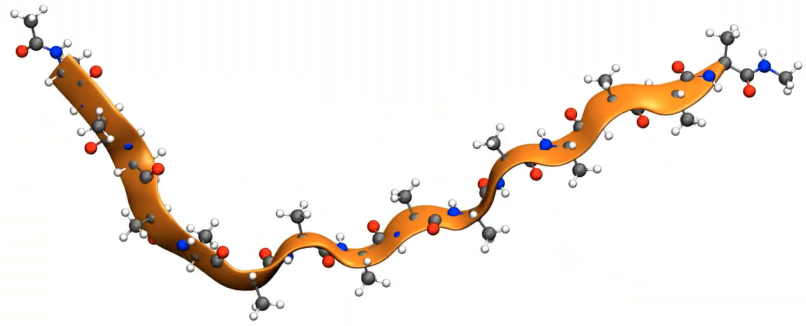
Appendix A. Additional Remarks

A.1. The application to quantum chemistry

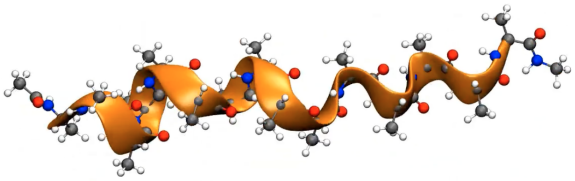
As a visual example, given a larger molecule in some initial position:



we want to see how it will look at later times, e.g. here at 29ps:



and here after 75ps:



The "energy of the system of atoms" that we want to compute (according to quantum mechanics in the Born–Oppenheimer approximation) is a mathematically well defined quantity that can be described easily:

The inputs are: n atomic nuclei with

- positions $\boldsymbol{r}_i \in \mathbb{R}^3$
- nuclear charges $Z_i \in \mathbb{N}$

These data define a simple Hamiltonian operator $\hat{H}(\boldsymbol{r}_1, ..., \boldsymbol{r}_n; Z_1, ..., Z_n)$ (using only the electrostatic forces) on electron wave functions

$$\psi : \mathbb{R}^{3\cdot(Z_1+Z_2+...+Z_n)} \to \mathbb{C}.$$

We assume here for simplicity that the total charge of the molecular system of interest is 0, and don't go into details related to the spin and the antisymmetry of ψ (they do not change

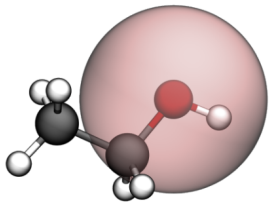
the overall flavor of the problem). Then the (ground state) energy is the lowest eigenvalue E of

$$\hat{H}(\boldsymbol{r}_1, ..., \boldsymbol{r}_n; Z_1, ..., Z_n)\psi = E(\boldsymbol{r}_1, ..., \boldsymbol{r}_n; Z_1, ..., Z_n) \cdot \psi$$

for any wave function ψ, and the force on atom i is $-\nabla_{\boldsymbol{r}_i} E(\boldsymbol{r}_1, ..., \boldsymbol{r}_n; Z_1, ..., Z_n)$.

The operator $\hat{H}(\boldsymbol{r}_1, ..., \boldsymbol{r}_n; Z_1, ..., Z_n)$ is easy to write down explicitly, but the resulting eigenvalue problem is not feasible to solve exactly (in any interesting cases).
However, $E(\boldsymbol{r}_1, ..., \boldsymbol{r}_n; Z_1, ..., Z_n)$ can be approximated well by somewhat costly, but still feasible calculations for small systems (up to 100 atoms), we use a version of Density Functional Theory, see e.g. Ganscha et al. (2025).

We actually want to model much larger systems (e.g. larger molecules in water), and this becomes possible if we make the simplifying assumption that interactions between atoms are mainly local.



In practical applications, we will only consider a limited radius around the atom.
The influence of further out atoms is assumed to be small (or modeled separately).

We also assume that atoms have a distance $> \epsilon$ from each other for some fixed $\epsilon > 0$: If the distance is too small, the repulsion between them becomes very strong and the system has a high energy, but we are only interested in systems that are below some upper bound of the energy.

In practice, we augment our model with an additional term that explicitly models short term repulsion and another part that tries to learn also some long range effects. However, we will not consider these terms here since they are of different flavors.

A.2. THE MATHEMATICAL PROBLEM FOR A FIXED NUMBER OF POINTS

We said it would be enough to find a finite list of polynomials that generate the ring of invariant functions

$$\mathbb{R}[D]^{\Sigma} := \mathbb{R}[d_{0,1}^2, d_{0,2}^2, ..., d_{n-2,n-1}^2]^{\Sigma}.$$

These polynomials exist because of Hilbert's basis theorem:

1. (Hilbert) $\mathbb{R}[D]^{\Sigma}$ is finitely generated.

So this solves the problem... for very pure mathematicians.
But we can even solve this "constructively":

2. (E. Noether) $\mathbb{R}[D]^{\Sigma}$ is generated by polynomials of degree $\leq |\Sigma|$ in each variable in D, $|\Sigma| \leq n!$.

This is a constructive solution (for pure mathematicians).
We can also easily improve on this bound:

3. (Göbel): $\mathbb{R}[D]^{\Sigma}$ generated by polynomials of degree $\leq |D|(|D|-1)/2$ in each variable in D, $|D| = n(n-1)/2$.

But in practice, even this bound is still much too loose to be useful in practice.

If we interpret "features" more loosely, we can also find $O(n^4)$ features that uniquely characterize a point set:
For every triple of 3 distinct points $(\boldsymbol{r}_i, \boldsymbol{r}_j, \boldsymbol{r}_k)$, take the 6 distances between them and the origin 0, and the **set** of all distances to the other points $\{(|\boldsymbol{r}_l|, |\boldsymbol{r}_l - \boldsymbol{r}_i|, |\boldsymbol{r}_l - \boldsymbol{r}_j|, |\boldsymbol{r}_l - \boldsymbol{r}_k|) \mid l = 1, 2, ..., n\}$ Doing this for all triples gives a set that also determines the configuration. However, these "features" are not a list of numbers, but **sets of sets of numbers**. (So they are not fully dividing out the action of the symmetric group, but make the quotient simpler.) See Widdowson and Kurlin (2023); Kurlin (2023); Hordan et al. (2024); Rose et al. (2023) for research in this direction.

A.3. Definition of PPSDs

We mentioned that each PPSD can be computed on point sets of size n in a time of $O(n)$. Note that this is a special property of polynomials, this is *not* true for rational descriptors like

$$\sum_{\{\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3\} \subseteq S} |\boldsymbol{r}_1 - \boldsymbol{r}_2|^{-2} \cdot |\boldsymbol{r}_2 - \boldsymbol{r}_3|^{-2} \cdot |\boldsymbol{r}_3 - \boldsymbol{r}_1|^{-2}$$

(And poles of this type appear e.g. in the Axilrod–Teller potential.)

While our way of constructing "polynomial functions" on point sets of arbitrary size n from polynomials in a fixed number k of variables seems natural, there would also be other options:

- We could use polynomials in the *closest* k points to the origin. However, we would then need to introduce a special operation for the case that several points have the same distance, and even with that, there would be discontinuities. So this does not seem to be useful in practice.

- We could sum only over *different* points:
  This actually leads to the same vector space of functions (Lemma 13 in Maennel et al. (2024)).

- We could use the *average* instead of the *sum* over k–tuples of points:
  This gives the same functions for fixed number of points n, but a different function class / generalization to larger n.
  This would not be the right generalization for us: Far away atoms should not influence the energy contribution of a central atom, but even if we use functions with $f(\boldsymbol{r}) \to 0$ for $|\boldsymbol{r}| \to \infty$, adding points far away would still change the number of points and hence the value of the average.

A.4. Our first invariants $I_L(S)$ depend only on the set of angles

Our first invariants $I_L(S)$ are certain lengths, so they are in particular nonnegative. To see that they depend only on the *set* of angles, we compute their square

$$I_L(S)^2 = \Big\langle \sum_{\boldsymbol{r}\in S} Y_L(\boldsymbol{r}) \,,\, \sum_{\boldsymbol{r}\in S} Y_L(\boldsymbol{r}) \Big\rangle = \sum_{\boldsymbol{r}_1,\boldsymbol{r}_2\in S} \big\langle Y_L(\boldsymbol{r}_1), Y_L(\boldsymbol{r}_2)\big\rangle = \sum_{\boldsymbol{r}_1,\boldsymbol{r}_2\in S} f_L\big(\measuredangle(\boldsymbol{r}_1,\boldsymbol{r}_2)\big)$$

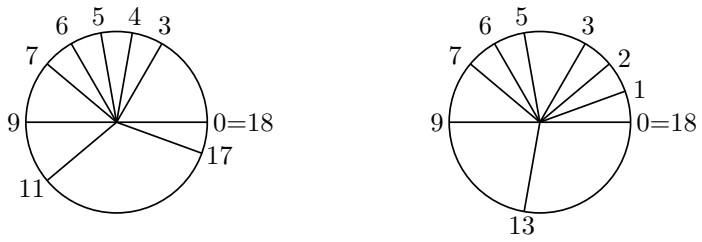
where $f_L(\alpha)$ is some function of the angle α between two vectors; the reason that it depends only on the angle is that any pair of points on $\mathbb{S}^2$ that have the same angle between them can be mapped to some standard pair of points $\boldsymbol{e}_0, \boldsymbol{e}_\alpha$ with the same angle between them by some rotation $g \in SO(3)$: Since Y_L is $SO(3)$–equivariant, we must have

$$\begin{aligned}\big\langle Y_L(\boldsymbol{r}_1), Y_L(\boldsymbol{r}_2)\big\rangle &= \big\langle Y_L(g\boldsymbol{e}_0), Y_L(g\boldsymbol{e}_\alpha)\big\rangle \\ &= \big\langle \rho_L(g) Y_L(\boldsymbol{e}_0), \rho_L(g) Y_L(\boldsymbol{e}_\alpha)\big\rangle = \big\langle Y_L(\boldsymbol{e}_0), Y_L(\boldsymbol{e}_\alpha)\big\rangle =: f_L(\alpha).\end{aligned}$$

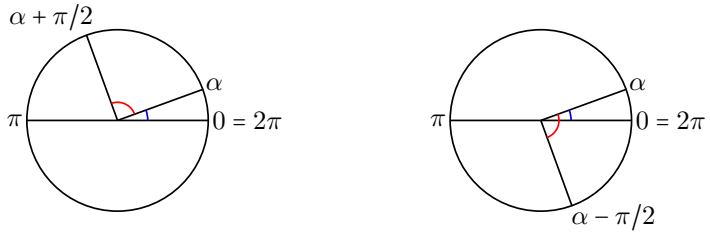
(Actually, the function f_L could be given explicitly by Legendre polynomials, but we don't need this.)

A.5. Counterexamples for completeness

Divide the unit circle into 18 equal angles, and position points at $S_1 := \{0,3,4,5,6,7,9,11,17\}$ or $S_2 := \{0,1,2,3,5,6,7,9,13\}$:



These two configurations have the same multiset of angles, and even the same multisets of "point triples up to rotation / reflection". So we need at least features of body order 5! Furthermore, this is not just an isolated instance, there are even submanifolds of counterexamples, here is the simplest example from Pozdnyakov et al. (2020):
The four points at angles $0, \alpha, \alpha + \frac{\pi}{2}, \pi$ or $0, \alpha, \pi, \alpha - \frac{\pi}{2}$ give the same set of angles between them:



It is not known whether features of body order 5 would suffice, but it is conjectured that there is some finite order that is sufficient, see Pebody et al. (2003).

A.6. Remarks on topological completeness

The "completeness" for *covariant* functions (like e.g. the dipole moment, which is covariant instead of invariant like the energy) cannot be defined by separation of configurations as for *invariant* functions in 2, since the function values $f(config)$ depend on the particular representative $config$ of an equivalence class under the action of G. Therefore we reformulate this condition as a density property in the function space:

3. *G–covariant* PPSDs are uniformly dense in *G–covariant* continuous functions on compact subsets of configurations colored by $\mathcal{C}$.

Note that the formulation "on compact subsets of configurations" includes both a bound on how far points of a configuration can be apart and a bound on the number of points of the configuration.

Proof ideas:

1. The moments $\sum_{\boldsymbol{r}\in S_\gamma} \boldsymbol{r}^{\otimes k}$ for $k = 0, 1, ..., n$ uniquely characterize configurations of n points in $\mathbb{R}^3$ (Lemma 19 in Maennel et al. (2024)). This is a natural generalization of the well known fact that point sets in $\mathbb{R}$ can be characterized by their moments.

2. We prove by standard topological arguments (in particular, Stone–Weierstrass) that we have for general representations V of a compact group G: Any two different G–orbits in the representation space V are separated by invariant polynomials (Proposition 20 in Maennel et al. (2024)). Then we apply this to $V :=$ a direct sum of tensor spaces and use 1.

3. Similar to 2, use Stone–Weierstrass and averaging over G.

A.7. Remarks on $SO(3)$ harmonic analysis

$O(3)$ operates in a canonical way on

- points (in $\mathbb{R}^3$), n–tuples of points, sets of points,
- functions of points, functions of n points, functions of sets of points,
- configurations (on each $\mathbb{R}^{3n}/\Sigma_\gamma$ separately)
- functions of configurations

For $d = 1, 2, 3, ...$ we can also consider "PPSDs f with values in a d–dimensional vector space V": If we fix a basis $\boldsymbol{e}_1, ..., \boldsymbol{e}_d$ of V, these are just given by d PPSDs $f_1, ..., f_d$ as

$$f(config) = f_1(config) \cdot \boldsymbol{e}_1 + ... + f_d(config) \cdot \boldsymbol{e}_d.$$

If we have a representation $\rho : O(3) \to GL(V)$, then we also have an operation on PPSDs with values in V, in particular we can also talk about equivariant PPSDs with values in V:

$$f(config) = \boldsymbol{v} \quad \Longleftrightarrow \quad f\big(\, g(config)\,\big) = \rho(g)\,\boldsymbol{v} \qquad \text{for all } g \in O(3)$$

As we already mentioned in the introduction, there exists up to isomorphism exactly one irreducible representation of $SO(3)$ with dimension $2L+1$ for $L = 0, 1, 2, ...$; we denote it by $\mathcal{H}^{(L)}$.

The (real) spherical harmonics

$$Y_L^m(\boldsymbol{r}) \qquad \text{for} \quad L = 0, 1, 2, ...; \quad m = -L, ..., L-1, L$$

are a basis of the $2L+1$–dimensional vector space of harmonic polynomials of degree L, this transforms as $\mathcal{H}^{(L)}$ under $SO(3)$. To be precise, in the introduction we only defined the Y_L as functions $\mathbb{S}^2 \to \mathbb{R}^{2L+1}$, but in the following, we will also use their simplest extension to all of $\mathbb{R}^3$, which is the polynomial given by

$$Y_L^m(\boldsymbol{r}) = |\boldsymbol{r}|^L \cdot Y_L^m\left(\frac{\boldsymbol{r}}{|\boldsymbol{r}|}\right)$$

With this definition, any homogeneous polynomial of degree L on $\mathbb{R}^3$ can be written uniquely as

$$p(\boldsymbol{r}) = h_L(\boldsymbol{r}) + |\boldsymbol{r}|^2 \cdot h_{L-2}(\boldsymbol{r}) + ... + \begin{cases} |\boldsymbol{r}|^{L-1} \cdot h_1(\boldsymbol{r}) & \text{for } L \text{ odd,} \\ |\boldsymbol{r}|^L \ \cdot \ h_0(\boldsymbol{r}) & \text{for } L \text{ even.} \end{cases}$$

A.8. Finiteness theorems for Clebsch–Gordan features

Here is a recap of what we discussed so far:

- Topological completeness: Any two (not G–equivalent) configurations can be distinguished by a G–invariant PPSD.
- Finiteness, 1: There is *a finite set* of G–invariant PPSDs that can distinguish between *any* such pair of configurations of n points (for any fixed n).
- Finiteness, 2: Given such a set of PPSDs, we can find $6n-5$ linear combinations of them that already can distinguish between any such configurations of n points.
- Algebraic completeness: We get all G–invariant PPSDs from Clebsch–Gordan products of fundamental features.

So we can learn $6n-5$ such functions if we have "sufficient" input features. In our concrete scheme of Clebsch–Gordan product features, can we give a set of input features that would suffice for all n–point configurations?

For n points...

1. ...it is enough to go up to $L \leq n$.
   Proof idea: The first n moments determine the configuration.
2. ...it is enough to use $\leq n$ Clebsch–Gordan factors.
   Proof idea: In the PPSDs we can use sums over *distinct* points, but there are at most n distinct points.

Unfortunately it is not obvious whether we can combine both restrictions.

However, we can get a finite set of generators when we look at the total degree:
The fundamental features $\sum_{\boldsymbol{r}\in S_\gamma} |\boldsymbol{r}|^{2k} \cdot Y_l(\boldsymbol{r})$ are polynomials of degree $2k+l$ in the coordinates, and in each Clebsch–Gordan operation the degrees of the polynomials add up to the new degree.
So we can keep track of the degree of the features, and the dimension of the vectors space of polynomials of given degree is finite. So we get a concrete finite set of polynomials if we have a bound on the degree, e.g. the one from Göbel's theorem.

A.9. Complications for analytic radial functions

Complication 1: Separating n–tuples
We said that polynomial functions $h(r^2)$ can separate points in $\mathbb{R}$, and in fact we can also separate n–tuples of points in $\mathbb{R}$. For a more general class of radial functions, we have to require properties that allow us to reach the same conclusions.
The exact property we need is:

For any $n \in \mathbb{N}$ there is a finite dimensional subspace $\mathcal{F}$ of the radial functions with

$$\text{For all } 0 < r_1 < ... < r_n \text{ exist } f_1, ..., f_n \in \mathcal{F} \text{ with } f_i(r_j) = \delta_{ij}.$$

This is satisfied for polynomials of degree $\leq n$.

We assume the radial basis functions...

1. ... are analytic,
2. ... allow approximation on compact intervals.

From these conditions we can prove the above property.
However, this would *fail for differentiable* instead of analytic in condition 1!

For analytic functions the proof relies on the theorem that the ring $\mathbb{R}\langle\langle x_1, ..., x_n \rangle\rangle$ is Noetherian.

Complication 2: Dimensions The second part of the Finiteness Theorem used semi–algebraic geometry, and we have seen that we need some condition like this.

Luckily, there is a "*sub–analytic geometry*" that is close enough to semi–algebraic geometry that we can use the same proof ideas, if we also *add compactness* conditions.

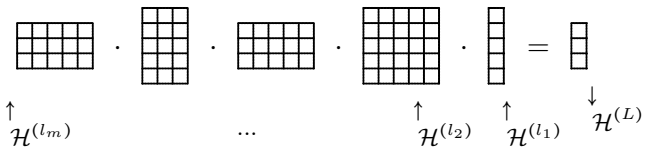
A.10. RECTANGULAR MATRICES AND MATRIX OF MATRICES

Above we encoded a list of vectors in $\mathcal{H}^{(0)} \oplus \mathcal{H}^{(1)} \oplus ... \oplus \mathcal{H}^{(L)}$ in a matrix $\mathcal{H}^{(L/2)} \otimes \mathcal{H}^{(L/2)}$. More generally, for $0 \le a \le b,\ a \equiv b \pmod 2$ we can use

$$\mathcal{H}^{(a)} \oplus ... \oplus \mathcal{H}^{(b)} \simeq \mathcal{H}^{((b+a)/2)} \otimes \mathcal{H}^{((b-a)/2)}$$

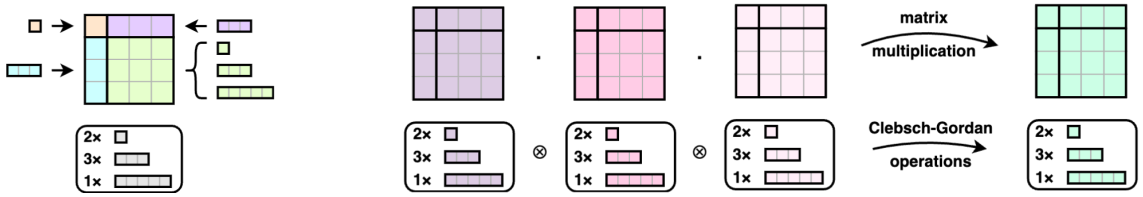
to encode $\mathcal{H}^{(a)} \oplus ... \oplus \mathcal{H}^{(b)}$ into a $(b+a+1) \times (b-a+1)$ matrix.

Matrix multiplication is also covariant for rectangular matrices, and we can directly simulate Clebsch–Gordan products: Such products



are covariant $(2L+1) \times 1$ matrices, i.e. vectors in $\mathcal{H}^{(L)}$, given by polynomials of degree $l_1 + ... + l_m$, and computing them takes $O(m \cdot a^3)$ steps for an upper bound a of matrix sides. Since we can directly simulate Clebsch–Gordan products in this way, the algebraic completeness theorem extends also to these features given by the products of rectangular matrices.

On GPUs / TPUs matrix multiplication is optimized for *large* matrices, so it is much better to multiply a few large matrices than many small matrices! We can do that by using a partition of large square matrices into many smaller rectangular matrices. When we use the same partition across columns and rows, the matrix product of two such large square matrices will again be a square matrix that is partitioned in the same way.



Extraction of features from the large square matrices
To build / apply a model for an invariant / covariant function of a colored point set, we go through these steps:

1. Compute (many) linear combinations with learnable weights of fundamental features

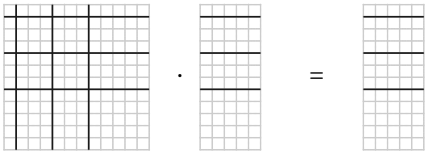
$$L_{\gamma,h,l} := \sum_{\boldsymbol{r} \in S_\gamma} h(|\boldsymbol{r}|) \cdot Y_l(\boldsymbol{r}/|\boldsymbol{r}|) \ \in \mathcal{H}^{(l)}$$

   for the required colors γ, radial basis functions h, and l's. The number of features required for each l is determined by the next steps (their use in rectangular matrices and how they are assembled in the big square matrices).

2. Assemble features from $\mathcal{H}^{(a)}, \mathcal{H}^{(a+1)}, ...\mathcal{H}^{(b)}$ into $(b-a+1)\times(b+a+1)$ or $(b+a+1)\times(b-a+1)$ matrices for the required pairs (a, b).

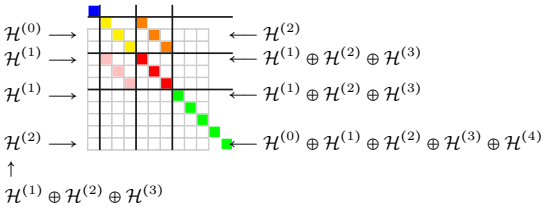
3. Assemble the rectangular matrices into large square matrices.
4. Compute the products of such large square matrices to get matrices containing higher order features.
5. → We are here: Extract features in $\mathcal{H}^{(0)}, \mathcal{H}^{(1)}, ..., \mathcal{H}^{(L)}$ from these product matrices. ←
6. Combine these features with learnable weights to build an invariant / covariant of the required form.

This extraction of features in $\mathcal{H}^{(l)}$ for different l from the large square matrices could be done directly by going to the individual $(b+a+1) \times (b-a+1)$ matrices and doing the Clebsch–Gordan decomposition into $\mathcal{H}^{(a)} \oplus \mathcal{H}^{(a+1)} \oplus ... \oplus \mathcal{H}^{(b)}$, but it is more efficient in the last product to use column vectors instead of matrices to obtain directly column vectors as outputs. By bundling several column vectors together, we get as a result again several column vectors:



If the end result should be scalars (and no covariant vectors), we can take the scalar products of the vectors with other fundamental feature vectors from the first step.

Scalars can also be obtained directly as the traces of square matrices that occur in the partitioned matrices:



Both methods to get scalars and / or vectors can be used to get scalar / vector features from the intermediate product matrices, i.e. the results after the computation of the product of the first $2, 3, ..., m$ large square matrices. This gives a large set of invariants / covariants of different body order, from which the required end result can be constructed, e.g. as linear combination with learnable coefficients.