Appendix

A Proofs

We begin by stating and proving a result mentioned in the main text: once we construct an invariant separator, we can obtain a universal model by composing the separation with a standard fully connected neural network:

Theorem A.1 (Separation Implies Universality)**.** *Let $f : \mathbb{R}^{d\times n} \to \mathbb{R}$ be a G-invariant continuous function. If $F : \mathbb{R}^{d\times n} \to \mathbb{R}^M$ is an invariant separator, then for any compact set $K \subset \mathbb{R}^{d\times n}$, and any $\epsilon > 0$, there exists a neural network $\mathcal{N}^\epsilon : \mathbb{R}^M \to \mathbb{R}$ such that $\sup\limits_{x\in K} \mid f(x) - \mathcal{N}^\epsilon \circ F(x) \mid < \epsilon$.*

Proof. Let $\epsilon > 0$ and $K \subseteq \mathbb{R}^{d\times n}$ be a compact set. Using Proposition 1.3 in Dym and Gortler [2023], there exists a continuous f^ϵ such that

$$|f(x) - f^\epsilon \circ F(x)| < \frac{\epsilon}{2}$$

The image of a compact set under a continuous function is a compact set, see Munkres [2000], then $S \coloneqq \mathrm{Im}\,(F)$ is compact. By the Universal Approximation Theorem Cybenko [1989], we can approximate f^ϵ with a fully-connected Neural Network with arbitrary precision, i.e. there exists a Neural Network Function $\mathcal{N}^\epsilon$ such that for all $x \in S$,

$$|f^\epsilon(x) - \mathcal{N}^\epsilon(x)| < \frac{\epsilon}{2}$$

By the Triangle Inequality, for all $x \in K$,

$$|f(x) - \mathcal{N}^\epsilon \circ F(x)| \leq |f(x) - f^\epsilon \circ F(x)| + |f^\epsilon \circ F(x) - \mathcal{N}^\epsilon \circ F(x)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

□

Theorem 2.1. *Two iterations of the* 1*-EWL test assign two point clouds $\mathcal{X}, Y \in \mathbb{R}^{3\times n}_{distinct}$ the same value, if and only if $X \underset{\mathcal{O}[3,n]}{=} Y$.*

Proof. Assume we initialize the hidden states with null information. After a single iteration, we have

$$h_i^{(1)} = \{\!\{ \|x_i - x_j\| \ \mid \ j \neq i \}\!\} = d(i, X) \tag{8}$$

By assumption, all $h_i^{(1)}$, i = 1,. . ., n, are distinct. Thus at the next iteration,

$$h_i^{(2)} = \Big(h_i^{(1)}, \{\!\{ (h_j^{(1)}, \|x_i - x_j\|) \ \mid \ j \neq i \}\!\} \Big) \tag{9}$$

$$\{\!\{ h_i^{(2)} \ \mid \ i \in [n] \}\!\} \tag{10}$$

we know each node's *ordered* distances from the other nodes, as the $i-th$ node is uniquely (intra-point-cloud) determined by $h_i^{(1)}$. Thus we can recover the distance matrix

$$\begin{array}{cc}
\begin{matrix} h_1^{(1)} & \quad h_2^{(1)} & \quad . & . & \quad h_n^{(1)} \end{matrix} & \\
\begin{pmatrix}
\|x_1 - x_1\| & \|x_2 - x_1\| & . & . & \|x_n - x_1\| \\
\|x_1 - x_2\| & 0 & . & . & . \\
. & . & 0 & . & . \\
. & . & . & 0 & . \\
\|x_1 - x_n\| & \|x_2 - x_n\| & . & . & \|x_n - x_n\|
\end{pmatrix} &
\begin{matrix} h_1^{(1)} \\ h_2^{(1)} \\ . \\ . \\ h_n^{(1)} \end{matrix}
\end{array}$$

Thus, we can fully recover the point cloud up to Euclidean motion, see Victor Garcia Satorras [2021]. In conclusion, if $X, Y \in \mathbb{R}^{d\times n}$ are assigned the same value by 1-EWL, then they are identical up to permutation and Euclidean motion, i.e. an $\mathcal{O}[3, n]$ transformation. □

By Theorem 2.1, 1-EWL is complete on $\mathbb{R}^{3\times n}_{distinct}$. We now show that 1-EWL is incomplete at most on a (non-trivial) measure-zero set, thus by definition, it is complete almost everywhere on the space of point clouds endowed with permutation, rotation, and reflection symmetries.

Theorem A.2 (1-EWL Separates Almost All Complete Euclidean Graphs)**.** *Let μ be the Lebesgue measure on $\mathbb{R}^{3\times n}$ where $n \geq 3$. Then, $\mu(\mathbb{R}^{3\times n} \setminus \mathbb{R}^{3\times n}_{distinct}) = 0$.*

Proof. We defined $\mathbb{R}^{3\times n}_{distinct} = \{X \in \mathbb{R}^{3\times n} \mid d(i,X) \neq d(j,X) \;\; \forall i \neq j\}$.

Then

$$\begin{aligned}\mathbb{R}^{3\times n} \setminus \mathbb{R}^{3\times n}_{distinct} &= \{X \in \mathbb{R}^{3\times n} \mid \exists i \neq j \in [n] \; s.t. \;\; d(j,X) = d(i,X)\} \\ &= \{X \in \mathbb{R}^{3\times n} \mid \exists i \neq j \in [n] \; s.t. \;\; \|\psi_{pow}(d(j,X)) - \psi_{pow}(d(i,X))\|^2 = 0\} \qquad (11)\end{aligned}$$

where ψ_{pow} is the power-sum polynomials defined as $\psi_{pow}(\vec{x}) = (\sum_{i=1}^n x_i, \ldots, \sum_{i=1}^n x_i^n)$, which is known to be injective on multisets with n elements.

Equation 11 defines an algebraic manifold with a non-trivial polynomial equality constraint, thus is of dimension $\leq 3n-1$. If an algebraic manifold embedded in $\mathbb{R}^{3\times n}$ has dimension $\leq 3n-1$, then it has measure zero Basu et al. [2006a].

$\square$

Theorem 3.2. *For every $X, Y \in \mathbb{R}^{3\times n}$, a single iteration of the 2-SEWL test assigns X and Y the same value if and only if $X \underset{\mathcal{SO}[3,n]}{=} Y$.*

Proof. Let X, Y $\in \mathbb{R}^{3\times n}$. Recall that a single iteration of the 2-SEWL assigns to each index pair i,j an initial coloring $\mathbf{C}_{(0)}[i,j] = \mathbf{C}_{(0)}[i,j](X)$ corresponding to the 2×2 Gram matrix of the points x_i, x_j. The coloring is then refined via

$$\mathbf{C}_{(\mathbf{1})}(i,j) = \mathbf{Embed}^{(0)}\left(\mathbf{C}_{(\mathbf{0})}(i,j), \{\!\!\{\left(\mathbf{C}_{(\mathbf{0})}(k,j), \mathbf{C}_{(\mathbf{0})}(i,k), \langle x_i \times x_j, x_k\rangle\right)\}\!\!\}_{k=1}^n\right).$$

and then a final global coloring is obtained from

$$\mathbf{C}_{\mathcal{G}} = \mathbf{Embed}^{(1)}\left(\{\!\!\{\mathbf{C}_{(1)}(i,j) | \, (i,j) \in [n]^2\}\!\!\}\right)$$

Let us denote the global feature $\mathbf{C}_{\mathcal{G}}$ obtained from X by $\mathbf{C}_{\mathcal{G}}(X)$, and the global feature obtained from Y by $\mathbf{C}_{\mathcal{G}}(Y)$.

By construction, if $X \underset{\mathcal{SO}[3,n]}{=} Y$ then $\mathbf{C}_{\mathcal{G}}(X) = \mathbf{C}_{\mathcal{G}}(Y)$. We need to prove that if $\mathbf{C}_{\mathcal{G}}(X) = \mathbf{C}_{\mathcal{G}}(Y)$ then $X \underset{\mathcal{O}[3,n]}{=} Y$.

To make the proof more readable, we introduce the following notation (we will later describe its significance):

$$X_{[i,j]} = [x_i, x_j, x_i \times x_j] \in \mathbb{R}^{3\times 3} \qquad (12)$$

$$P_{[i,j,k]} = X_{[i,j]}^T x_k \qquad (13)$$

$$G_{[i,j]}(X) = X_{[i,j]}^T X_{[i,j]} \qquad (14)$$

$$h_{[i,j]}(X) = \mathbf{Embed}_\alpha \{\!\!\{P_{[i,j,k]} \mid k \in [n], k \neq i,j\}\!\!\} \qquad (15)$$

$$m_{[i,j]}(X) = \left(G_{[i,j]}(X), h_{[i,j]}(X)\right) \qquad (16)$$

We now show that the multiset $C_{\mathcal{G}}(X) = \{\!\!\{\mathbf{C}_{(\mathbf{1})}(\mathbf{i},\mathbf{j}) \mid i,j \in [n]\}\!\!\}$ allows recovering the multiset $h_X := \{\!\!\{m_{[i,j]} \mid i,j \in [n], i,j \in [n]\}\!\!\}$.

It is enough to show that we can recover $m_{[i,j]}$ from its corresponding $\mathbf{C}_{(1)}[i,j]$ for every $i,j \in [n]$ and then the multiset equivalence follows immediately.Note that $G_{[i,j]}$ is the 3×3 Gram matrix of the vectors $x_i, x_j, x_i \times x_j$. It can be recovered from $\mathbf{C}_{(\mathbf{0})}(\mathbf{i},\mathbf{j})$, which is the 2×2 Gram matrix of x_i, x_j, because $\langle x_i \times x_j, x_k\rangle = \langle x_i \times x_j, x_i\rangle = 0$ and

$$\|x_i \times x_j\| = \|x_i\|\|x_j\||\sin\theta| = \|x_i\|\|x_j\|\sqrt{1-\cos\theta^2} = \|x_i\|\|x_j\|\sqrt{1-\left(\frac{\langle x_i, x_j\rangle}{\|x_i\|\|x_j\|}\right)^2},$$

where θ is the angle between x_i and x_j. The quantity on the RHS of the equation can be extracted from $\mathbf{C_{(0)}(i,j)}$.

As for $h_{i,j}$, we can recover it as a multiset since:

$$P_{[i,j,k]} = (\langle x_i, x_k\rangle, \langle x_j, x_k\rangle, \langle x_i \times x_j, x_k\rangle) = \left(\mathbf{C}_{(0)}(i,k)[1,2], \mathbf{C}_{(0)}(k,j)[1,2], \langle x_i \times x_j, x_k\rangle\right) \tag{17}$$

We saw that $h_X = \mathbf{Embed}\{\!\{m_{[i,j]}(X)|i,j \in [n]\}\!\}$ can be recovered from $\mathbf{C}_{\mathcal{G}}(X)$ and thus in particular our assumption that $\mathbf{C}_{\mathcal{G}}(X) = \mathbf{C}_{\mathcal{G}}(Y)$ implies that $h_X = h_Y$. We will now use this to show that $X \underset{\mathcal{SO}[3,n]}{=} Y$.

We first deal with the degenerate case where all points in X are identical, that is $x_1 = x_2 = \ldots = x_n$. In this case, all Gram matrices $G_{i,j}(X)$, and all entries in each of the matrices, will be identical, and thus by assumption also all Gram matrices $G_{i,j}(Y)$, and all their entries will be identical. This implies that Y also consists of a single point with the same norm as the one point in X, and therefore $X \underset{\mathcal{SO}[3,n]}{=} Y$.

We can now assume that not all points in X are the same. Define $r(X) = \mathrm{rank(X)} = \max_{i,j\in[n]} \mathrm{rank}(G_{[i,j]}(X))$ (note that we assume that $n \geq 3$). By assumption we have $\{\!\{G_{[i,j]}(X)|i,j \in [n]\}\!\} = \{\!\{G_{[i,j]}(Y)|i,j \in [n]\}\!\}$, thus $r(X) = r(Y)$ and there exist $i \neq j$ and $s,t \in [n]$ such that $G_{[i,j]}(X) = G_{[s,t]}(Y)$, ($\star$) they both have rank r, and $x_i \neq x_j$. Due to (*) it follows that $y_s \neq y_t$ and in particular $s \neq t$.

By Kraft and Procesi [1996], the equality of Gram matrices implies that there exists an orthogonal transformation, $T \in \mathcal{O}(3)$, such that

$$T(x_i) = y_s,\ T(x_j) = y_t,\ T(x_i \times x_j) = y_s \times y_t. \tag{18}$$

If $x_i \times x_s \neq 0$ we see that T preserves orientation and therefore $T \in \mathcal{SO}(3)$ (see Remark A.3). If not, and if T is a reflection, we can modify T to be a rotation which still satisfies (18) by composing it with a relfection that fixes the ≤ 1 dimensional subspace spanned by x_i, x_j. Thus in any case we can assume that $T \in \mathcal{SO}(3)$. By assumption and ($\star$), we have $\{\!\{P_{[i,j,k]}(X)\, k \neq i,j\}\!\} = \{\!\{P_{[s,t,k]}(Y),\, k \neq s,t\}\!\}$.
This implies that there exists some permutation $\sigma \in S_n$ such that $\sigma(i) = s.\sigma(j) = t$ and $P_{[i,j,k]}(X) = P_{[s,t,\sigma(k)]}(Y)$ for all $k \neq i,j$. We deduce that for all $k \neq i,j$

$$\langle y_s, Tx_k\rangle = \langle Tx_i, Tx_k\rangle = \langle x_i, x_k\rangle = \langle y_s, y_{\sigma(k)}\rangle$$

and similarly

$$\langle y_t, Tx_k\rangle = \langle y_t, y_{\sigma(k)}\rangle$$
$$\langle y_s \times y_t, Tx_k\rangle = \langle y_s \times y_t, y_{\sigma(k)}\rangle$$

Now note that each y_k is in the span of $y_s, y_t, y_s \times y_t$ (even when $r < 3$), and similarly every x_k is in the span of $x_i, x_j, x_i \times x_j$ and so Tx_k is also in the span of $y_s, y_t, y_s \times y_t$. It follows that $Tx_k - y_{\sigma(k)} = 0$, and thus we showed that X and Y are related by a $\mathcal{SO}[3,n]$ transformation.

□

Remark A.3. In the proof above we said that if (18) holds, and $y_s \times y_t$ is not zero, then T is in $SO(3)$. This follows from the fact that for general orthogonal transformations and vectors a, b

$$(Ta) \times (Tb) = \det(T) T(a \times b).$$

Setting $a = x_i, b = x_j$ and using (18) we obtain that

$$y_s \times y_t = \det(T)(y_s \times y_t)$$

and so if $y_s \times y_t$ is not zero, then $\det(T) = 1$.

Theorem 4.2. *Let F_ϕ denote the parametric function simulating the 2-SEWL test. Then for Lebesgue almost every ϕ the function $F_\phi : \mathbb{R}^{3\times n} \to \mathbb{R}^{6n+1}$ is separating with respect to the action of $\mathcal{SO}[3,n]$.*

Proof. We recall that F_ϕ is defined to simulate a single iteration of the 2-SEWL test using sort-based injective multiset functions. In more detail, recall that the initial coloring corresponding to an index pair (i,j) and a point cloud $X \in \mathbb{R}^{3\times n}$ is given by the Gram matrix of x_i, x_j, and denoted by $\mathbf{C}_{(0)}(i,j) = \mathbf{C}_{(0)}(i,j)(X)$. We then define

$$\mathbf{C}_{(\mathbf{1})}(i,j) = \left(\mathbf{C}_{(\mathbf{0})}(i,j), \mathbf{Embed}_\alpha\left(\left[\mathbf{C}_{(\mathbf{0})}(k,j), \mathbf{C}_{(\mathbf{0})}(i,k), \langle x_i \times x_j, x_k\rangle\right]_{k=1}^n\right)\right).$$

and

$$\mathbf{C}_{\mathcal{G}} = \mathbf{Embed}_\beta\left(\mathbf{C}_{(1)}(i,j) |\, (i,j) \in [n]^2\right)$$

where $\mathbf{Embed}_\theta(y_1, \ldots, y_n)$ is permutation invariant (=multiset function), continuous in θ and y_i, and defined by

$$\mathbf{Embed}_\theta(y_1, \ldots, y_n) = \langle b_j,\ \Psi\left(a_j^T y_1 \ldots, a_j^T y_n\right)\rangle,\ j = 1, \ldots, 6n+1. \tag{19}$$

with $\Psi = \mathrm{sort}$ (or alternatively, Ψ could be the power sum polynomials), and θ denoting the concatenation of all the mapping parameters a_i and b_j.

We denote by $\phi = (\alpha, \beta)$ the concatenation of the two parameter vectors of the **Embed** mappings in the constructions, and $F_\phi(X)$ denoted the output $C_\mathcal{G} = C_\mathcal{G}(X;\phi)$ obtained by this construction.

Since we already showed that the 2-SEWL test is complete, it is sufficient to show that for Lebesgue almost every (α, β), the mapping $\mathbf{Embed}_\alpha$ is permutation invariant and separating on $\mathbb{R}^{3\times n}$, and the mapping $\mathbf{Embed}_\beta$ is permutation invariant and separating on the image of the mapping f_α which we define as

$$f_\alpha(X) = \left(\mathbf{C}_{(1)}(i,j)(X;\alpha) |\, (i,j) \in [n]^2\right).$$

By Theorem 4.1 we know that $\mathbf{Embed}_\alpha$ is separating for Lebesgue almost every α. For fixed α, we know that f_α is a semi-algebraic mapping, since it is a composition of polynomials and the piecewise linear sort function, which are semi-algebraic mappings, and as compositions of semi-algebraic mappings are semi-algebraic mappings. The dimension of the image of a semi-algebraic mapping is never larger than the dimension of the domain, and so $f_\alpha(\mathbb{R}^{3\times n})$ is a semi-algebraic set of dimension $\leq \dim(\mathbb{R}^{3\times n}) = 3n$ (see Basu et al. [2006b] for the necessary real algebraic geometry statements re composition and dimension). To apply Theorem 3.2 we need to work with a permutation invariant domain, so we artifically enlarge the domain of $\mathbf{Embed}_\beta$ to be

$$\bigcup_{\sigma \in S_{n^2}} \sigma(f_\alpha(\mathbb{R}^{3\times n}))$$

which is a finite union of sets of dimension $\leq 3n$ and hence also has dimension $\leq 3n$. It follows that for almost every β the function $\mathbf{Embed}_\beta$ is separating on this permutation invariant set, with embedding dimension of $6n+1$ as we defined in (19). Using Fubini's theorem, this implies that for almost every (α, β) the functions $\mathbf{Embed}_\alpha$ and $\mathbf{Embed}_\beta$ are both separating, and this proves the theorem. $\square$

Complexity We conclude by discussing the complexity of computing F_ϕ. Calculating each $\mathbf{C}_{(1)}(i,j)$ using sort-based embeddings $\mathbf{Embed}_\alpha$ requires $\mathcal{O}(n^2\log(n))$ operations.

Since there are $\mathcal{O}(n^2)$ such $\mathbf{C}_{(1)}(i,j)$ the total complexity of computing all of them is $\mathcal{O}(n^4\log(n))$. In the second step we compute $\mathbf{Embed}_\beta$ on multisets of size $D \times N$ where $D = \mathcal{O}(n), N = \mathcal{O}(n^2)$, and with embedding dimension of $\mathcal{O}(n)$. This requires $\mathcal{O}(n^4 + n^3\log(n))$ operations , so the total complexity is $\mathcal{O}(n^4\log(n))$

In Appendix D we extend our results to arbitrary d. In this case, we get a complexity of $\mathcal{O}(n^{d+1}\log(n))$ (where for simplicity we consider the limit $n \to \infty$ with d fixed, to cancel out some mixed terms in n, d which are negligible in this limit.).

B Experiment Details

As mentioned, we exemplified the viability of the theory presented by testing separation on challenging point cloud pairs. We wished to address the following scenario: given a pair of point clouds, each labeled distinctly, what would be the accuracy score of $\mathcal{SO}[3,n]$ (or $\mathcal{O}[3,n]$) invariant architectures in this classification task following training on a labeled dataset of these examples? This setup partly informs us of the *separation* capability of these architectures, i.e. how well do these models distinguish similar (for instance, 1-EWL equivalent), yet non-isomorphic, input?

For implementation, we used code by Joshi et al. [2022] that implements several contemporary $\mathcal{SO}[3,n]$ invariant architectures and evaluated them as described below. This framework has several additional tests for geometric graphs, but they were irrelevant to our setting because they are redundant for the fully-connected geometric graphs we focus on. We modified the implementation of Joshi et al. [2022] by implementing our novel invariant architectures, 2-SEWLnet, implementing EGNN Victor Garcia Satorras [2021], and testing counterexample point cloud pairs from Pozdnyakov et al. [2020], Pozdnyakov and Ceriotti [2022]. We used implementation by Joshi et al. [2022] of MACE Batatia et al. [2022], TFN Thomas et al. [2018] and GVPGNN Jing et al. [2021]. The $\mathcal{SO}[3,n]$ invariant architectures are trained on replicas of each counterexample pair and then testing is performed on the same pair.

B.1 Technical Details

Hyperparameters	2-SEWLnet	EGNN	MACE	TFN	GVPGNN
Learning rate	0.0001	0.0001	0.0001	0.0001	0.0001
Hidden Dimension	2 (Pair-wise)	1	2	64	64
Number of Layers	1	2	3	3	3
Batch size	1	1	1	1	1
Correlation	NA	NA	3	3	NA

Table 2: GNN implementations and code pipeline based on Joshi et al. [2022].

We trained the various invariant models on an NVIDIA A40 GPU implemented in PyTorch Paszke et al. [2019]. The hyperparameters were a learning rate of 0.0001 with Adam optimizer Kingma and Ba [2017], with the learning rate scheduler ReduceOnPLateau that reduces the learning rate once the loss stopped diminishing. We trained each model on a dataset of 50 copies of each pair for 100 epochs, while injecting permutation and rotation to each point cloud during training. The test and validation datasets are each a pair of plain (no permutation or rotation injected) point clouds. Thus each epoch has ternary accuracy results, of 0%, 50%, and 100%. We then average the accuracy of the last 20% of epochs to obtain the overall accuracy. This was done to allow the model to converge while allowing for a sufficiently large test measurements to obtain statistical significance.

B.2 2-SEWLnet

2-SEWLnet is an implementation of a simulation of a single iteration of 2-SEWL, see 4. We implemented the architecture by directly embedding the multiset function F_ϕ, using the low-dimensional invariant embeddings from Section 4. We chose the sort function as the one-dimensional permutation separating invariant, as it constitutes an isometry from $\mathbb{R}^n/S_n$ to $\mathbb{R}^n$, then composing it with linear mappings, yields a Bi-Lipschitz mapping Balan et al. [2022], Dym and Gortler [2023]. We used differentiable sorting from PyTorch Paszke et al. [2019] to enable backpropagation. We note that the model is able to learn the classification task with backpropagation only using the sort vector-wise activation, i.e. without a fully-connected neural network composed on it. This is not a trivial result that is immediately implied by the completeness of 2-SEWL, as we aim to minimize the softmax cross-entropy loss and in practice reach almost zero loss, thus not only yielding two distinct embeddings corresponding to each distinct point cloud (as guaranteed by Theorem 3.2) but learning them to be (approximately) one-hot encoded vectors using our injective continuous **Embed** functions.

B.3 EGNN

We used the EGNN version used for classification problems as described in Victor Garcia Satorras [2021], which does not use an equivariant coordinate update step. This version's expressive power is bounded by 1-EWL. Interestingly, we found that using the original implementation did not yield sufficient separation for the classification task for the examples Hard1-3 in Table 1, yet when using sort and non-linear point-wise activations, rather than sums of neural network functions applied point-wise, led to perfect classification. The results in the table are reported with this implementation.

C Background Theory

C.1 Low Dimensional Separating Invariants

In Section 4 We presented a condensed summary of the results from Dym and Gortler [2023] as they pertain to this paper's scope. We devote this appendix to expound on the context of these results. In Subsection 4.1, we discussed the *Power-Sum Symmetric Polynomials*. These polynomials yield a separating invariant with respect to permutations of real-valued vectors in $\mathbb{R}^n$. The invariant learning literature often discussed an extension of this characterization for vector-valued features, the *Multi-Power-Sum Symmetric Polynomials*, defined, for an input $X = (x_1 \dots x_n) \in \mathbb{R}^{d \times n}$ and $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d_{\geq 0}$, as

$$P_\alpha(X) = \sum_{i=1}^{n} x_i^\alpha \tag{20}$$

$$P(X) = \left(P_\alpha(X)\right)_{\alpha \in [n]^d, |\alpha| \leq n} \tag{21}$$

where $x^\alpha := x_1^{\alpha_1} \dots x_d^{\alpha_d}$ and $|\alpha| := \sum_{i=1}^{n} \alpha_i$. These polynomials define a separating invariant for point clouds in $\mathbb{R}^d$ with respect to the permutation of the (n) columns Maron et al. [2019b]. Yet, its embedding dimension is $\binom{n+d}{d}$. The goal of Dym and Gortler [2023] was to reduce the embedding dimension to a complexity linear in the $n \cdot d$ dimension of the input.

As a first example of such a result, Dym and Gortler [2023] show the for Lebesgue almost every $\binom{n+d}{d}$ dimensional vectors $w_1, \dots, w_{2nd+1}$

$$X \underset{S_n}{=} Y \iff \langle w_i, P(X) \rangle = \langle w_i, P(Y) \rangle, i = 1, \dots, 2\dim(\mathcal{M}) + 1 \tag{22}$$

Thus we obtain a separating invariant of dimension $O(d \cdot n)$ rather than $O(n^{2d})$. Yet, we still had to calculate all of the $O(n^{2d})$ polynomial entries. Therefore, computationally speaking, this approach did not yield much.

To remedy this, Dym and Gortler [2023] proposed alternative invariants based on $\mathbb{R}^n$ permutation invariants Ψ. Such a Ψ. and any choice of a vector $a \in \mathbb{R}^d$, induces a permutation invariant on $\mathbb{R}^{d \times n}$ of the form

$$\mathbb{R}^{d \times n} \ni (x_1, \dots, x_n) \mapsto \Psi(a^T x_1, \dots, a^T x_n).$$

This technique of producing high dimensional invariants from low dimensional ones is known as *polarization*. To obtain invariants that are also *separating*, we need (i) to choose Ψ to be invariant and separating on $\mathbb{R}^n$. Two examples of such functions, are the functions Ψ_{sort} and Ψ_{pow} which we defined in (5):

$$\Psi_{sort}(x_1, \dots, x_n) = \operatorname{sort}(x_1, \dots, x_n) \quad \text{and} \quad \Psi_{pow}(x_1, \dots, x_n) = \left(\sum_{i=1}^{n} x_i^t \right)_{t=1}^{n}.$$

Additionally, we need (ii) to choose not a single polarization function defined by a single a, but rather $2nd + 1$ random vectors $a_1, \dots, a_{2nd+1}$. More precisely, Dym and Gortler [2023] showed that for Lebesgue almost every $a_1, \dots, a_{2nd+1} \in \mathbb{R}^d$ and $b_1, \dots, b_{2nd+1} \in \mathbb{R}^n$the function

$$\textbf{Embed}_\theta(x_1, \dots, x_n) = \langle b_j, \ \Psi\left(a_j^T x_1 \dots, a_j^T x_n\right) \rangle, \ j = 1, \dots, 2nd + 1. \tag{23}$$

defined in (6) is permutation invariant and separating. The role of the projection by the b_j is to reduce the embedding dimension to $2nd + 1$ rather than the $(2nd + 1)n$ dimension we would get if these projections were not applied.

We note that the complexity of a single invariant in (23) would be $O(n \log(n))$ (assuming $n > d$) when using sorting or $O(n^2)$ when using power sum polynomials. Accordingly, the complexity of computing $2nd + 1$ invariants would be $O(dn^2 \log(n))$ using sorting or $O(dn^3)$ using power sum polynomials.

Finally, we note that (as mentioned in the main text), if we're interested in separation only on a semi-algebraic permutation invariant subset $\mathcal{X} \subseteq \mathbb{R}^{d \times n}$ with dimension $D_\mathcal{X}$, then the number of separating invariants needed in (23) would be $2D_\mathcal{X} + 1$ rather than $2nd + 1$.

D Extensions

In the main text we described Vanilla k-WL tests which are well-defined for all k and d, and the 2-SEWL test which is well-defined for the case $d = 3$ (since vector products are used). We now explain how to define a $(d-1)$ SEWL test for general d, and then explain how these tests can be easily modified to give a $(d-1)$-EWL test with similar complexity, which is $\mathcal{O}[3,n]$ invariant and separating rather than $\mathcal{SO}[3,n]$ invariant and separating.

D.1 SEWL for General Dimension d

The complete 2-SEWL test for $d = 3$ point clouds can be generalized to a $(d-1)$-SEWL test for general d-dimensional point clouds by a generalization of the cross-product operator. This generalization is formally known as the Hodge dual operator Jost [2017].

For fixed $x_{i_1}, \ldots, x_{i_{d-i}} \in \mathbb{R}^d$, we define the following linear functional

$$f : \mathbb{R}^d \to \mathbb{R}$$
$$f(x) = \det\left(x, x_{i_1}, \ldots, x_{i_{d-1}}\right)$$

By Riesz's Representation Theorem Bachman and Narici [2000], every linear functional on $\mathbb{R}^d$ is essentially a function in the form of a dot product against some (unique) vector in $\mathbb{R}^d$. Thus, there exists some vector, denoted by $x^\star = x^\star(x_{i_1}, \ldots, x_{i_{d-1}}) \in \mathbb{R}^d$, such that

$$\det\left(x, x_{i_1}, \ldots, x_{i_{d-1}}\right) = f(x) = \langle x, x^\star \rangle. \tag{24}$$

We see directly from the definition that $x^\star$ is orthogonal to $x_{i_1}, \ldots, x_{i_{d-1}}$ and is non-zero if and only if these $d-1$ vectors are linearly independent. Moreover, this choice of vector is $SO(d)$ equivariant, which means that for any $x_{i_1}, \ldots, x_{i_{d-i}} \in \mathbb{R}^d$ and $R \in SO(d)$ we have

$$x^\star(Rx_{i_1}, \ldots, Rx_{i_{d-1}}) = Rx^\star(x_{i_1}, \ldots, x_{i_{d-1}}). \tag{25}$$

This is because for any $x \in \mathbb{R}^d$ we have

$$\begin{aligned}
\det\left(x, Rx_{i_1}, \ldots, Rx_{i_{d-1}}\right) &= \det\left(RR^T x, Rx_{i_1}, \ldots, Rx_{i_{d-1}}\right) \\
&= \det(R) \det\left(R^T x, x_{i_1}, \ldots, x_{i_{d-1}}\right) \\
&= \langle R^T x, x^\star \rangle \\
&= \langle x, Rx^\star \rangle
\end{aligned}$$

Finally, we note that the coordinates of x^* can be calculated by inserting the unit vectors $e_1, \ldots, e_d$ into (24). That is

$$\left(x^\star\right)_j = \det\bigl(e_j, x_{i_1}, \ldots, x_{i_{d-1}}\bigr)$$

where $\left(x^\star\right)_j$ is the j-th entry of $x^\star$. This requires computing d different determinants of $d \times d$ matrices, and so the total complexity of computing x^* is d^4.

The $(d-1)$-SEWL test We have shown an extension of the definition of the cross-product that respects orientation, thus now we can naturally define a $(d-1)$-SEWL test which will be $\mathcal{SO}[d,n]$ invariant and separating in d dimensions (generalizing the 2-SEWL test for $d = 3$).

We define for each $(d-1)$-tuple $\mathbf{i} \in [n]^{d-1}$ an initial coloring $\mathbf{C}_{(0)}(\mathbf{i}) = \mathbf{C}_{(0)}(\mathbf{i})(X)$ corresponding to the $(d-1) \times (d-1)$ Gram matrix of the points $x_{i_1}, \ldots, x_{i_{d-1}}$. We denote $x^*(x_{i_1}, \ldots, x_{i_{d-1}})$ by $x^*(\mathbf{i})$. The coloring is then refined via

$$\mathbf{C}_{(\mathbf{1})}(\mathbf{i}) = \mathbf{Embed}^{(0)}\left(\mathbf{C}_{(\mathbf{0})}(\mathbf{i}), \{\!\!\{\bigl(\mathbf{C}_{(\mathbf{0})}(\mathbf{i}[k \setminus 1]), \ldots, \mathbf{C}_{(\mathbf{0})}(\mathbf{i}[k \setminus (d-1)]), \langle x^\star(\mathbf{i}), x_k \rangle\bigr)\}\!\!\}_{k=1}^{n}\right).$$

where $\mathbf{i}[j \setminus t]$ is the multi-index $\mathbf{i}$ with its t-th coordinate replaced by j; e.g. for $t = 1$, $\mathbf{i}[j \setminus 1] = (j, i_2, \ldots, i_k)$. Then a final global coloring is obtained from

$$\mathbf{C}_{\mathcal{G}} = \mathbf{Embed}^{(1)}\left(\{\!\!\{\mathbf{C}_{(1)}(\mathbf{i}) | \, \mathbf{i} \in [n]^{d-1}\}\!\!\}\right)$$

The $(d-1)$-SEWL test can be shown to be $\mathcal{SO}[d,n]$ complete, using the same arguments used in the proof of Theorem 3.2.

D.2 $(d-1)$-EWL for general dimension d

We now return to the case where reflections are also considered symmetries, and we're looking for complete tests with respect to the group $\mathcal{O}[d,n]$. The Vanilla-d-WL test will be $\mathcal{O}[d,n]$ complete. However, a more efficient test can be obtained by tweaking the $(d-1)$-SEWL test which is not reflection-invariant, to attain a reflection invariant $\mathcal{O}[d,n]$ complete test.

This tweaking is obtained as follows. We fix some reflection R_0 (a reflection is an orthogonal matrix with a negative determinant). We define the $(d-1)$-EWL test for a given $X \in \mathbb{R}^{d\times n}$ by applying the $(d-1)$-SEWL test to both X and R_0X to obtain $\mathbf{C}_{\mathcal{G}}(X)$ and $\mathbf{C}_{\mathcal{G}}(R_0X)$, and then computing a final global feature via

$$C_{\mathcal{G}}^{ref}(X) = \mathbf{Embed}^{(2)}\{\!\{\mathbf{C}_{\mathcal{G}}(X), \mathbf{C}_{\mathcal{G}}(R_0X)\}\!\} \tag{26}$$

In the following theorem, we show how the completeness of the $(d-1)$-SEWL test implies the completeness of the $(d-1)$-EWL test.

Theorem D.1. *For every $X, Y \in \mathbb{R}^{d\times n}$, a single iteration of the $(d-1)$-EWL test assigns X and Y the same value if and only if $X \underset{\mathcal{O}[d,n]}{=} Y$.*

Proof. **Invariance:** We prove that for every $R \in O(d)$ and permutation matrix $P \in S_n$ we have that $C_{\mathcal{G}}^{ref}(RXP) = C_{\mathcal{G}}^{ref}(X)$. We can divide into two cases: If $R \in SO(d)$ then by the $\mathcal{SO}[d,n]$ invariance of $\mathbf{C}_{\mathcal{G}}(X)$ we have that

$$\mathbf{C}_{\mathcal{G}}(RXP) = \mathbf{C}_{\mathcal{G}}(X)$$
$$\mathbf{C}_{\mathcal{G}}(R_0RXP) = \mathbf{C}_{\mathcal{G}}((R_0RR_0^T)R_0XP) = \mathbf{C}_{\mathcal{G}}(R_0X)$$

On the other hand, if R is a reflection, then

$$\mathbf{C}_{\mathcal{G}}(RXP) = \mathbf{C}_{\mathcal{G}}((RR_0^T)R_0XP) = \mathbf{C}_{\mathcal{G}}(R_0X)$$
$$\mathbf{C}_{\mathcal{G}}(R_0RXP) = \mathbf{C}_{\mathcal{G}}(X)$$

and so in both cases, we obtain that

$$C_{\mathcal{G}}^{ref}(RXP) = \mathbf{Embed}^{(2)}\{\!\{\mathbf{C}_{\mathcal{G}}(RXP), \mathbf{C}_{\mathcal{G}}(R_0RXP)\}\!\} = \mathbf{Embed}^{(2)}\{\!\{\mathbf{C}_{\mathcal{G}}(X), \mathbf{C}_{\mathcal{G}}(R_0X)\}\!\} = C_{\mathcal{G}}^{ref}(X)$$

Completeness: We prove that if $X, Y \in \mathbb{R}^{d\times n}$ and $C_{\mathcal{G}}^{ref}(X) = C_{\mathcal{G}}^{ref}(Y)$ then X and Y are related by a permutation and orthogonal transformation.

Since $C_{\mathcal{G}}^{ref}(X) = C_{\mathcal{G}}^{ref}(Y)$ it follows that either $C_{\mathcal{G}}(X) = C_{\mathcal{G}}(Y)$ or $C_{\mathcal{G}}(X) = C_{\mathcal{G}}(R_0Y)$. The completeness of the $(d-1)$-SEWL test (Theorem 3.2) then implies that X is related to either Y or R_0Y by an $\mathcal{SO}[d,n]$ transformation. In either case, this implies that X and Y are related by an $\mathcal{O}[d,n]$ transformation. □

D.3 Continuous Implementation and Computational Complexity

In Section 4 we showed how the 2-SEWL test can be realized by a continuous piecewise differentiable architecture that uses sort-based multi-set injective functions. The complexity of this construction was $O\left(n^4\log(n)\right)$. Similarly, the $(d-1)$-SEWL and $(d-1)$-EWL tests can be computed with complexity of $O\left(n^{d+1}\log(n)\right)$ (in the scenario where d stays constant and $n\to\infty$) . The leading order of the computation complexity stems from computing the n^{d-1} colorings $\left(\mathbf{C}_{(\mathbf{1})}(\mathbf{i}),\ \mathbf{i}\in[n]^{d-1}\right)$ and embedding their corresponding multisets, each one of those requires $\mathcal{O}(n^2\log(n))$ operations.