Equivariant Frames and the Impossibility of Continuous Canonicalization

Nadav Dym [*1] **Hannah Lawrence** [*2] **Jonathan W Siegel** [*3]

Abstract

Canonicalization provides an architecture-agnostic method for enforcing equivariance, with generalizations such as frame-averaging recently gaining prominence as a lightweight and flexible alternative to equivariant architectures. Recent works have found an empirical benefit to using probabilistic frames instead, which learn weighted distributions over group elements. In this work, we provide strong theoretical justification for this phenomenon: for commonly-used groups, there is no efficiently computable choice of frame that preserves continuity of the function being averaged. In other words, unweighted frame-averaging can turn a smooth, non-symmetric function into a discontinuous, symmetric function. To address this fundamental robustness problem, we formally define and construct *weighted* frames, which provably preserve continuity, and demonstrate their utility by constructing efficient and continuous weighted frames for the actions of $SO(d)$, $O(d)$, and S_n on point clouds.

1. Introduction

Equivariance has emerged in recent years as a cornerstone of geometric deep learning, with widespread adoption in domains including biology, chemistry, and graphs (Jumper et al., 2021; Corso et al., 2023; Liao et al., 2023; Satorras et al., 2021; Frasca et al., 2022). This fundamental idea — incorporating known data symmetries into a learning pipeline — often enables improved generalization and sample complexity, both in theory (Petrache & Trivedi, 2024; Mei et al., 2021; Bietti et al., 2021; Elesedy, 2021) and in practice (Batzner et al., 2022; Liao et al., 2023).

[*]Equal contribution [1]Faculty of Mathematics, Faculty of Computer Science, Technion, Israel [2]Department of Electrical Engineering and Computer Science, MIT, MA, USA [3]Department of Mathematics, Texas A&M University, TX, USA. Correspondence to: Nadav Dym <nadavdym@technion.ac.il>.

Proceedings of the 41^{st} International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

The genesis of equivariant learning focused on equivariant *architectures*, i.e. custom parametric function families containing only functions with the desired symmetries. However, equivariant architectures must be custom-designed for each group action, which reduces the transferability of engineering best practices. Moreover, the building blocks of many equivariant architectures (such as tensor products) are computationally intensive (Passaro & Zitnick, 2023).

In light of these difficulties, the more lightweight and modular approach of frame-averaging has received renewed attention. Frame-averaging (Puny et al., 2021) harnesses a generic neural network f to create an equivariant framework by averaging the network's output over input transformations. It is a direct extension of group-averaging (also known as the Reynolds operator), whereby f is made invariant by averaging over all input transformations from a group G:

$$\mathcal{I}_{\text{avg}}[f](v) = \int_G f(g^{-1}v)dg. \tag{1}$$

However, while group averaging scales with $|G|$ (which can be large or infinite), frame-averaging can enjoy computational advantages by averaging over only an input-dependent subset of G. A notable special case of frame-averaging is canonicalization (Figure 1), which "averages" over a *single, canonical* group transformation per point. Intuitively, canonicalization is a standardization of the input data, such as centering a point cloud with respect to translations. Frame-averaging methods are universal (so long as f is universal), in the sense that they can approximate all continuous equivariant functions, and are projections[1] onto the space of invariant functions. Recent approaches include using both fixed (Duval et al., 2023; Du et al., 2022; 2023), and learned (Zhang et al., 2019; Kaba et al., 2022; Luo et al., 2022) frames and canonicalizations. Despite their simplicity, however, it seems that such approaches have generally not yet supplanted popular equivariant architectures in applications.

In this work, we unearth an insidious problem with frames, which may shed light on their slow adoption in applications: they very often induce discontinuity. The absence of a continuous canonicalization for permutations was already observed by Zhang et al. (2020). We prove that there are no continuous canonicalization for rotations, either. More-

[1]Recall that a projection $\mathcal{P}$ need only satisfy $\mathcal{P}(\mathcal{P}f) = \mathcal{P}f$.

over, we show that the only continuity preserving frame for permutations is the $(n!)$-sized Reynolds operator, and that there is no continuity preserving frame with finite size for rotations in two dimensions. In other words, even if the generic network f is continuous, the frame-averaged function may not be! This constitutes a significant lack of robustness, in which one can slightly perturb an input, and have the predicted output entirely change.

To address this issue, we generalize Puny et al. (2021)'s definition of frames in two ways. First, we define *weighted frames*, where group elements are assigned non-uniform, input dependent weights. Second, we observe that their definition of frame equivariance necessitates very large frames at points with large stabilizer. To avoid this, we define weak equivariance, which relaxes the notion of equivariance at points with non-trivial stabilizers. Finally, we define a natural notion of continuity for these generalized frames, and name the resulting frames **robust frames**. We show that invariant projection operators induced by robust frames *always preserve continuity*. Finally, we show that robust frames of moderate size can be constructed for group actions of interest, including the action of permutations (where small unweighted frames cannot preserve continuity).

Serendipitously, our results coincide with several quite recent works, which demonstrate a significant benefit to using probabilistic or weighted frames quite similar to ours (Kim et al., 2023; Mondal et al., 2023; Pozdnyakov & Ceriotti, 2023). Thus, our results provide both a theoretical framework and strong justification for empirically successful approaches which learn distributions over the group, while also suggesting future avenues in practice.

We visualize canonicalization and frame-averaging on the left of Figure 1, while the right shows the relation between canonicalization, averaging, frames, and weighted frames.

1.1. Our Contributions

To summarize, our main contributions are:

1. **Limitations of canonicalization:** Using tools from algebraic topology, we show that *a continuous canonicalization does not exist* for S_n, $SO(d)$, and $O(d)$.

2. **Limitations of frames:** We show that for finite groups acting freely on a connected space, the only frame which preserves continuity is the Reynolds operator. In particular, this is the case for the group of permutations S_n acting on $d \geq 2$ dimensional point clouds. For the infinite group $SO(2)$ we show that there is no continuity preserving frame of finite cardinality.

3. **Robust frames:** We define weighted frames as an alternative to standard (unweighted) frames, and introduce notions of weak equivariance and continuity. We call weighted frames satisfying both criteria *robust frames*. Unlike unweighted frames, the invariant projection operators induced by robust frames *always* preserve continuity. With some care, this can be generalized to the equivariant case as well.

4. **Examples of robust frames:** We give constructions of robust frames of moderate, polynomial size for S_n, $SO(d)$, and $O(d)$. For S_n, we also provide complementary lower bounds on the size of any robust frame, which is directly related to the computational efficiency of implementation.

To give a more concrete picture, we summarize our results restricted to a large family of examples: the various group actions of "geometric groups" on point clouds. A point cloud is a matrix $X \in \mathbb{R}^{d \times n}$, with columns denoted $(x_1, \ldots, x_n)$. Point clouds arise in many applications, ranging from graphics and computer vision (as representations of objects and scenes in the physical world), to chemistry and biology (as representations of molecular systems). Group actions of interest here include the application of translation, rotations, orthogonal transformations, or permutations.

Our results for point clouds and these group actions are described in Table 1. The action of translations has trivial continuous canonicalizations. When $n, d > 1$, the action of permutations admits no continuous canonicalization, nor any unweighted frame which preserves continuity besides the full Reynolds frame. In contrast, robust frames whose complexity is linear in the input dimension can be constructed, at least for the subset $\mathbb{R}^{d \times n}_{\text{distinct}}$ (on which the permutation group acts freely). When considering robust frames defined on the whole space, we require a larger cardinality, but it is still significantly smaller than the full $n!$ cardinality required by unweighted frames.

For the action of rotations on $\mathbb{R}^{d \times n}$, we prove that there is no continuous canonicalization when $n \geq d$. Moreover, when $d = 2$, we show that no continuity preserving frame of finite cardinality exists, either. In contrast, robust frames exist with a cardinality which (when d is small and $n \to \infty$) scales like $\sim n^{d-1}$.

1.2. Paper structure

In Section 2, we introduce useful mathematical preliminaries, including criteria for an equivariant projection operator to preserve continuity. We then establish impossibility results for continuous canonicalizations and frames. As a solution, Section 3 defines weighted frames, and establishes criteria under which they are robust (i.e. preserve continuity). The following two sections, Section 4 and Section 5, give explicit continuity-preserving robust frame constructions for invariance under S_n and $SO(d)$, respectively. Finally, Section 6 discusses extensions to equivariance.

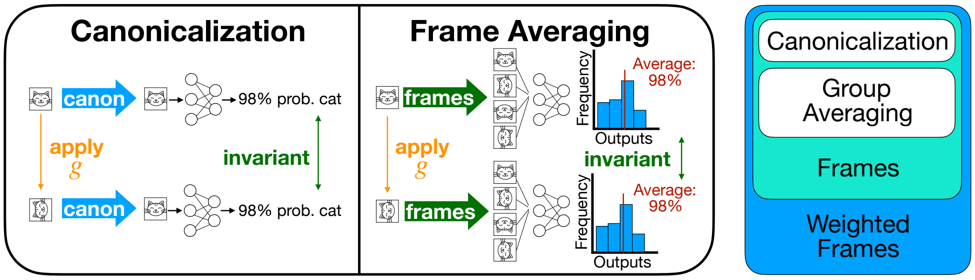

Figure 1. **Left:** Canonicalization and its generalization to frame averaging. Under group transformation of the input, both its canonicalization and the set of inputs transformed by the frame are invariant. **Right:** Group averaging and group canonicalization are special cases of frames of maximal and minimal size, respectively. Frames, in turn, are a special case of weighted frames.

Action	Translation	Permutation	Permutation	$SO(d)$	$O(d)$
Domain	$\mathbb{R}^{d\times n}$	$\mathbb{R}^{d\times n}_{\text{distinct}}$	$\mathbb{R}^{d\times n}$	$\mathbb{R}^{d\times n}$	$\mathbb{R}^{d\times n}$
Canonicalization	yes	no	no	no if $n \geq d$	no if $n > d$
Frame	1	$N = n!$	$N = n!$	$N = \infty$ if $n \geq d$	$N > 1$ if $n > d$
Weighted frame	1	$N \leq (n-1)d$	$\frac{n}{2} < N \leq n^{2(d-1)}$	$N \leq (d-1)!\binom{n}{d-1}$	$N \leq 2(d-1)!\binom{n}{d-1}$

Table 1. Summary of main results. For various group actions, we show lower and upper bounds on the minimal cardinality N for which a continuity-preserving frame or weighted frame exists, and whether a continuous canonicalization exists (in which case $N = 1$). The $N = \infty$ result for unweighted $SO(d)$ frames is proven only when $d = 2$ (although we conjecture it holds for general $d \geq 2$ as well).

2. Projections, canonicalization, and frames

Preliminaries Unless stated otherwise, throughout the paper we consider compact groups G acting linearly and continuously on (typically) finite dimensional real vector spaces V and W, or else on subsets of V and W closed under the action of G (see Appendix A for a formal definition) . We will work with the groups S_n of n-dimensional permutations, $O(d)$ of orthogonal matrices ($M \in \mathbb{R}^{d\times d}$ $s.t.$ $M^T M = I$), and $SO(d)$ of rotations ($M \in O(d)$ $s.t.$ $\det(M) = 1$).

A function $f : V \to W$ is equivariant if $f(gv) = gf(v)\ \forall v \in V, g \in G$, and invariant (a special case where G acts trivially on W) if $f(gv) = f(v) \forall v \in V, g \in G$. A compact group G admits a unique Borel probability measure which is both left and right invariant. As in (1), we denote integration according to this measure by $\int dg$. The orbit of $v \in V$ is $[v] := \{gv :\ g \in G\}$, and the stabilizer is $G_v = \{g \in G |\, gv = v\}$. We say that v has a trivial stabilizer if $G_v = \{e\}$, where $e \in G$ denotes the identity element. G acts on V *freely* if $G_v = \{e\}\ \forall v \in V$. For H a subgroup of G, G/H denotes the left cosets of H in G.

For G acting on spaces V and W as described, let $F(V, W)$ denote the space of functions from V to W, and let $C(V, W)$ denote the subset of these functions which are also continuous. Let $F_{\text{equi}}(V, W)$ (and $C_{\text{equi}}(V, W)$) denote (continuous) functions which are also *equivariant*. Notions of denseness and boundedness in $C(V, W)$, described in this paper, are with respect to the topology of uniform convergence on compact subsets of V. Full proofs of all claims in the paper are given in the Appendix.

2.1. Projection operators

Group averaging, as well as the cheaper alternatives summarized in Figure 1, achieve equivariant models via *equivariant projection operators*, a notion we now define formally.

Definition 2.1 (BEC operator)**.** Let $\mathcal{E} : F(V, W) \to F(V, W)$ be a linear operator. We say that $\mathcal{E}$ is a

1. **Bounded operator** If for every compact K, there exists a positive constant C_K such that

$$\max_{v\in K} \|\mathcal{E}[f](v)\| \leq C_K \max_{v\in K} \|f(v)\|$$

2. **Equivariant projection operator** for every $f \in F(V, W)$, $\mathcal{E}(f)$ is equivariant, and moreover if f is equivariant then $\mathcal{E}[f] = f$.

3. **Preserves continuity** If $f : V \to W$ is continuous, $\mathcal{E}[f]$ will also continuous.

If $\mathcal{E}$ satisfies all three conditions, we call it a BEC operator. BEC operators can be used to define universal, equivariant models which preserve continuity.

Proposition 2.2. *Let $\mathcal{E} : F(V, W) \to F(V, W)$ be a BEC operator, and $Q \subseteq C(V, W)$ a dense subset . Then $\mathcal{E}(Q) = \{\mathcal{E}(q)|\ q \in Q\}$ contains only continuous equivariant functions, and is dense in $C_{\text{equi}}(V, W)$.*

We will often handle the invariant case separately, denoting $\mathcal{I}$ instead of $\mathcal{E}$ and C_{inv} or F_{inv} instead of C_{equi} or F_{equi}.

We note that equivariant universality can be obtained using other methods, based on approximating all equivariant polynomials (Yarotsky, 2022; Dym & Maron, 2020), or exploiting separating invariants (Dym & Gortler, 2024; Hordan et al., 2024; Villar et al., 2021), see also (Kurlin, 2023; Widdowson & Kurlin, 2023; Cahill et al., 2024). Our focus in this paper is on projection based methods.

Proposition 2.2 motivates our search for BEC operators. Frame-like constructions typically produce bounded and equivariant operators, but preserving continuity is more challenging, and thus is the main focus of this paper.

One simple method to obtain BEC operators is by group averaging. In the invariant setting this was defined in Eq. 1, and we can define an equivariant projection operator $\mathcal{E}_{\mathrm{avg}}$ as $\mathcal{E}_{\mathrm{avg}}[f](x) = \int_G gf(g^{-1}x)dg$. It is not difficult to show that $\mathcal{E}_{\mathrm{avg}}$ is a BEC operator, and therefore preserves continuity. However, the disadvantage of group averaging is the complexity of computing such an operator when the group is large. Thus our goal is designing *efficient* robust frames.

2.2. Canonicalization

Instead of defining projection operators by averaging over *all* members in the orbit of a point v, an *orbit canonicalization* $y : V \to V$ maps all elements in any given orbit to a *unique* orbit element, and "averages" over it.

Definition 2.3 (Orbit canonicalization)**.** A function $y : V \to V$ is an orbit canonicalization if

1. y is G invariant: $y_{gv} = y_v, \quad \forall g \in G, v \in V$
2. y maps v to a member of its orbit: $y_v \in [v], \quad \forall v \in V$

In the case where the action of G on V is *free*, the orbit canonicalization naturally induces a group canonicalization $h : V \to G$. Namely, if y_v is a canonical element in the orbit $[v]$, then the fact that v has a trivial stabilizer implies that there is a unique $h_v \in G$ such that $y_v = h_v^{-1}v$.

Example 2.4. For the (free) action of $\mathbb{R}^d$ on $\mathbb{R}^{d\times n}$ by translation defined above, a simple orbit canonicalization is the map from $X \in \mathbb{R}^{d\times n}$ to the unique Y whose first coordinate is zero, that is $(0, x_2 - x_1, \dots, x_n - x_1)$. The associated group canonicalization is the map $h_X = x_1$.

Canonicalization-based equivariant projection operators. If y_v is a canonicalization, we can define an invariant projection operator on functions $f : V \to \mathbb{R}$ via

$$\mathcal{I}_{\mathrm{can}}[f](v) = f(y_v).$$

One can easily check that $\mathcal{I}_{\mathrm{can}}[f]$ is invariant, and that if f already is invariant then $\mathcal{I}_{\mathrm{can}}[f] = f$.

Similarly, if G acts on V freely so that h_v is well-defined, then we can also write $\mathcal{I}_{\mathrm{can}}[f](v) = f(h_v^{-1}v)$, which shows how canonicalizations are frames of cardinality one. Moreover, we can define an equivariant projection operator via $\mathcal{E}_{\mathrm{can}}[f](v) = h_v f(h_v^{-1}v)$. In general, it is not difficult to see that $\mathcal{I}_{\mathrm{can}}$ and $\mathcal{E}_{\mathrm{can}}$ are bounded projection operators in the sense of Definition 2.1. However, below we will show that even the invariant projection operator often cannot achieve continuity preservation.

From the definition of $\mathcal{I}_{\mathrm{can}}$, it is clear that if the canonicalization h_v is continuous, then $\mathcal{I}_{\mathrm{can}}$ preserves continuity. The following shows that these notions are in fact equivalent.

Proposition 2.5. *Let $y : V \to V$ be a canonicalization. Then $\mathcal{I}_{\mathrm{can}} : F(V, \mathbb{R}) \to F(V, \mathbb{R})$ preserves continuity if and only if y is continuous.*

For some relatively simple examples, continuous canonicalizations are available:

1. For the action of $O(d)$ on $\mathbb{R}^d$, $y(x) = \|x\|e_1$.
2. For the action of S_n on $\mathbb{R}^n$, $y(x) = sort(x)$.

In contrast, the following are "natural" orbit canonicalizations, but are not continuous:

1. The action of S_n on $\mathbb{R}^{d\times n}$ with $d > 1$ has a canonicalization defined by lexicographical sorting. As shown in Figure 2, this is not continuous.
2. The action of $SO(2) \cong S^1$ on $\mathbb{C}^n$ has a natural orbit canonicalization by rotating $z \in \mathbb{C}^n$ so that its first coordinate z_1 is a real positive number. However, this is not uniquely defined when $z_1 = 0$, which induces a discontinuity (see also Figure 2).

We will show that, not only are the canonicalizations above not continuous, but there do not exist *any* continuous canonicalizations for these actions (Theorem 2.8).

To give a sense for the practical utility of this fact, consider that learned canonicalization methods for $SO(3)$ (Kaba et al., 2022; Kim et al., 2023) involve orthogonalizing learned features via the Gram-Schmidt process. A priori, depending on the learned network, this may or may not be discontinuous (if the learned features are linearly dependent or nearly zero). Our results answer definitively that no continuous canonicalization is possible.

Examples where there is no continuous canonicalization. We now detail a strategy for proving that a continuous canonicalization does not exist, and apply this method to numerous examples of interest.

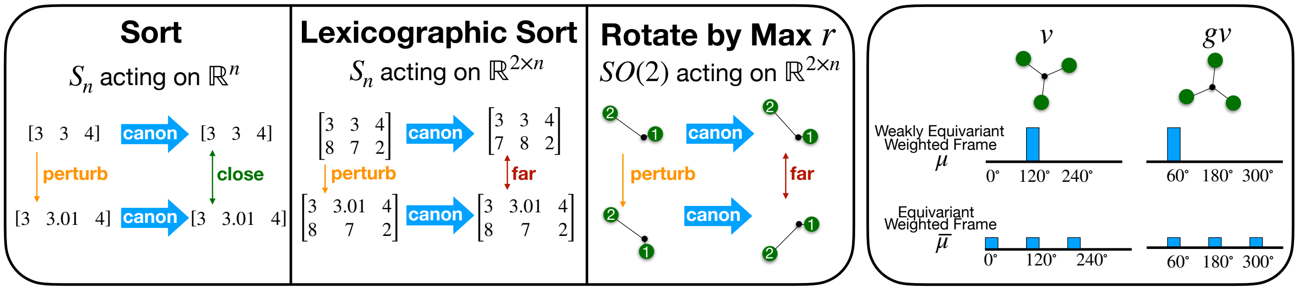

Figure 2. **Left:** Examples of natural canonicalizations. Canonicalizing $\mathbb{R}^n$ with respect to S_n via sorting is continuous, but the lexicographic generalization to $\mathbb{R}^{2\times n}$ is not. Similarly, one can canonicalize $2D$ ordered point clouds with respect to $SO(2)$ by applying the rotation that aligns the first node with the positive x-axis, but this is discontinuous. **Right:** Visualization of a weakly equivariant frame μ for $SO(2)$ acting on an unordered point cloud v with $120°$ self-symmetry, evaluated both at v and at gv, a $60°$ rotation of v. μ_{gv} is not a $60°$ rotation of μ_v, but $\overline{\mu}$ is equivariant by definition, so $\overline{\mu}_{gv}$ *is* a $60°$ rotation of $\overline{\mu}_v$. Thanks to the self-symmetry of v, v is exactly the same whether rotated by $0°$, $120°$, or $240°$. Thus, invariant projection by μ or $\overline{\mu}$ has the same result; μ is simply $3x$ more efficient.

Our general proof strategy is as follows: we begin by noting that a continuous canonicalization $y : V \to V$ really defines a continuous map on the *quotient*, $\tilde{y} : V/G \to V$, which is a right inverse of the quotient map $q : V \to V/G$. Specifically, continuity and G-invariance of y imply that y induces a continuous map on the quotient space V/G. Further, if y maps each v to a member of its orbit, it follows that $q \circ \tilde{y} = I$ is the identity map on the quotient space V/G.

A canonicalization on V will also be a canonicalization when restricted to a G-stable subset $Y \subseteq V$. Thus, for any G-stable subset Y, a continuous canonicalization y yields a right inverse of the quotient map $Y \to Y/G$. Our strategy will be to show that for some G-stable set Y, the quotient map $Y \to Y/G$ cannot have a right inverse. We will prove this by finding topological invariants, specifically homotopy and homology groups (see for instance Hatcher (2002)), which contradict the existence of such a right inverse $\tilde{y}$.

We now apply this method to prove the impossibility of continuous canonicalization for several groups. We begin with a well known example which will be helpful for illustrating our methodology.

Proposition 2.6. *The action of $\mathbb{Z}$ on $\mathbb{R}$ by addition does not have a continuous canonicalization.*

Proof. For this example, we use the fundamental group as an obstruction. The fundamental group is an invariant of topological spaces which essentially consists of loops in the space (see Hatcher (2002), Chapter 1). Maps between spaces induce corresponding maps between their fundamental groups, so the fundamental group can be used as an obstruction to show that certain maps cannot exist.

The quotient $\mathbb{R}/\mathbb{Z}$ is isomorphic to S^1, whose fundamental group is $\mathbb{Z}$, while the fundamental group of $\mathbb{R}$ is trivial. A continuous canonicalization would give a map $\tilde{y} : S^1 \to \mathbb{R}$ such that $S^1 \xrightarrow{\tilde{y}} \mathbb{R} \xrightarrow{q} S^1$ is the identity on S^1. Thus, the induced maps on the fundamental group would satisfy $\mathbb{Z} \xrightarrow{\tilde{y}^*} \{0\} \xrightarrow{q^*} \mathbb{Z}$ is the identity on $\mathbb{Z}$, which is impossible. □

Proposition 2.7. *Consider $O(d)$ and $SO(d)$ acting on $\mathbb{R}^{d\times n}$. If $n > d \geq 1$ for $O(d)$, or $n \geq d \geq 2$ for $SO(d)$, then there is no continuous canonicalization.*

Proof idea. For $SO(d)$, we reduce to $n = d = 2$ and use an algebraic topology argument similar to the one before, but with homology groups. The proof for $O(d)$ is similar. □

We remark that this result is sharp, i.e. a continuous canonicalization exists for $O(d)$ when $n \leq d$ and for $SO(d)$ when $n < d$. We construct such canonicalizations in Appendix B.

2.3. Frames

Puny et al. (2021) generalize canonicalization to allow for *averaging* over an equivariant set of points instead. Concretely, they define a frame $\mathcal{F}$ as $\mathcal{F} : V \to 2^G \backslash \emptyset$, which is equivariant: $\mathcal{F}(gv) = g\mathcal{F}(v)$, where the equality is between sets. The invariant projection operator $\mathcal{I}_{\text{frame}}$ induced by the frame is defined for every $f : V \to \mathbb{R}$ as

$$\mathcal{I}_{\text{frame}}[f](v) := \frac{1}{|\mathcal{F}(v)|} \sum_{g \in \mathcal{F}(v)} f(g^{-1}v).$$

The equivariant projection operator $\mathcal{E}_{\text{frame}}$ is similar, but with summand $gf(g^{-1}v)$ instead.

We note that, when the frame maps to $G \in 2^G$ for all elements of the input space $\mathcal{M}$, then frame-averaging reduces to group-averaging. On the other extreme, a frame where $|\mathcal{F}(v)| = 1 \; \forall v$ is exactly a group canonicalization.

Puny et al. (2021) provide several examples where frames seem a natural alternative to canonicalization. Nonetheless, in many cases (such as PCA), these frames do not preserve continuity. It turns out that this can be unavoidable:

Theorem 2.8. *Let G be a finite group acting continuously on a metric space V, let V_{free} denote the points in V with trivial stabilizer, and let $\mathcal{F}$ be a frame which preserves continuity on V_{free}. If V_{free} is connected, then $\mathcal{F}(v) = G$ for all v in the closure of V_{free}.*

Proof idea. First, we show that for finite groups acting freely on V_{free}, a frame which preserves continuity must be *locally* constant. Next, due to connectivity, this can be extended to a *global* result. Finally, we combine the fact that the frame is globally constant with equivariance ($\mathcal{F}(gv) = g\mathcal{F}(v)$), and extend this to the closure of V_{free}. □

Corollary 2.9. *Let $d, n > 1$, and consider S_n acting on $\mathbb{R}^{d\times n}$. If $\mathcal{F}$ is a continuity-preserving frame, then $\mathcal{F}(X) = S_n$ for all $X \in \mathbb{R}^{d\times n}$.*

The corollary implies that there is no continuity preserving frame of reasonable cardinality for permutations. Interestingly, Theorem 2.8 and Corollary 2.9 imply that the failure of a small S_n frame to preserve continuity is not a result of self-symmetries alone: even continuity only at points with trivial stabilizer implies frame size $n!$.

We next show that, for the infinite group $SO(2)$, there is no unweighted frame of any finite cardinality that preserves continuity under the action on pairs of points.

Theorem 2.10. *Consider $SO(2)$ acting on $\mathbb{R}^{2\times n}$ with $n \geq 2$. If $\mathcal{F}$ is a continuity-preserving frame, then*

$$\sup_{X\in\mathbb{R}^{d\times n}\setminus\mathbf{0}} |\mathcal{F}(X)| = \infty,$$

i.e. there does not exist a finite (unweighted) frame which preseves continuity.

Note that the result above holds at X with trivial stabilizer, since the supremum excludes $X = 0$ (at which any unweighted frame trivially has infinite size).

3. Weighted frames

In some cases, it is therefore impossible to define frames of reasonable size which preserve continuity. To address this issue, we suggest a generalization of frame-averaging to *weighted frames*, and determine conditions under which averaging over a weighted frame preserves continuity. We furthermore obtain advantages in size by defining a weaker notion of equivariant frames for points v with $G_v \neq \{e\}$.

It will first be helpful to recall some measure theory. If μ is a Borel probability measure on a topological group G, we can, for $g \in G$, define a "pushforward" measure $g_*\mu$ which assigns to $A \subseteq G$ the measure $g_*\mu(A) = \mu(g^{-1}(A))$.

For every $v \in V$ and Borel probability measure τ on G, we let $\langle\tau\rangle_v$ be the measure defined for every $A \subseteq G$ by

$$\langle\tau\rangle_v(A) = \int_{s\in G_v} s_*\tau(A)ds,$$

where the integral is over the Haar measure of G_v. This can be thought of as averaging over the stabilizer of v. For a frame $\mu_{[\cdot]}$, we use the shortened notation $\bar{\mu}_v := \langle\mu_v\rangle_v$.

Definition 3.1 (Weighted frames)**.** A weighted frame $\mu_{[\cdot]}$ is a mapping $v \mapsto \mu_v$ from V to the space of Borel probability measures on G. $\mu_{[\cdot]}$ is *equivariant* at $v \in V$ if $\forall g \in G$, $\mu_{gv} = g_*\mu_v$. $\mu_{[\cdot]}$ is *weakly equivariant* at $v \in V$ if $\forall g \in G$,

$$\bar{\mu}_{gv} = \langle\mu_{gv}\rangle_{gv} = g_*\langle\mu_v\rangle_v = g_*\bar{\mu}_v, \forall g \in G,$$

We say that a frame $\mu_{[\cdot]}$ is (weakly) equivariant if it is (weakly) equivariant at all $v \in V$. See Figure 2 for a visual.

Remark 3.2. If a frame is equivariant, then it is also weakly equivariant. Moreover, the two definitions are equivalent at points v with trivial stabilizer.

Though our definitions allow for general probability measures, we will only be interested in measures with finite support. Thus, the main difference from Puny et al. (2021) is that we allow group elements to have non-uniform weights.

The goal of using general weighted frames is to provide a more efficient BEC operator than the full Reynolds operator. To quantify this, we define the *cardinality* of the frame μ as

$$\sup_{v\in V} |\{g \in G : \ \mu_v(g) > 0\}|.$$

Note that the cardinality is the worst-case number of evaluations of f required to evaluate $\mathcal{I}_{\text{weighted}}$ or $\mathcal{E}_{\text{weighted}}$.

Our primary motivation in defining *weakly* equivariant weighted frames is computational. Indeed, note that for every $s \in G_v$, equivariance of a frame implies that $\mu_v = \mu_{sv} = s_*\mu$. This implies that the support of μ_v must be at least the size of the stabilizer of v, which can be problematic when the stabilizer is large. Weak equivariance bypasses this issue, drawing on the intuition that if $s \in G_v$, then $v = sv$ implies $f(sv) = f(v)$. Therefore, in terms of *invariant* averaging, equivalence up to G_v should not affect the projection operator. We now make this precise.

Both equivariant and weakly equivariant weighted frames can be used to define invariant projection operators. Namely,

$$\mathcal{I}_{\text{weighted}}[f](v) := \int_G f(g^{-1}v)d\mu_v(g).$$

$\mathcal{E}_{\text{weighted}}$ is the same, but with integrand $gf(g^{-1}v)$.

Proposition 3.3. (Invariant frame averaging) Let $\mu_{[\cdot]}$ be a weighted frame. If $\mu_{[\cdot]}$ is *weakly equivariant*, then $\mathcal{I}_{\text{weighted}}$ is a bounded, invariant projection operator.

To preserve continuity, we will need a natural notion of continuity of $\mu_{[\cdot]}$. We discuss this in the next section.

Remark 3.4 (Equivariant frame averaging)**.** If $\mu_{[\cdot]}$ is *equivariant*, then $\mathcal{E}_{\text{weighted}}$ is a bounded projection operator. However, if $\mu_{[\cdot]}$ is only *weakly* equivariant, then $\mathcal{E}_{\text{weighted}}$ may not be an equivariant projection operator; we discuss possible solutions to this in Section 6.

3.1. Continuous weighted frames

In this subsection, we define a natural notion of a *continuous weakly equivariant weighted frame*. We will show that with this notion, the operator $\mathcal{I}_{\text{weighted}}$ preserves continuity.

A natural definition of continuity would be to require that whenever $v_n \to v$, $\mu_{v_n} \to \mu_v$ in the weak topology on probability measures on G (i.e. for any continuous function $f : G \to \mathbb{R}$, $\int f d\mu_{v_n} \to \int f d\mu_v$). However, we allow the following even weaker notion of continuity, which takes into account the ambiguity at points with non-trivial stabilizers.

Definition 3.5 (Continuous weighted frames)**.** Let μ be a weakly equivariant weighted frame. We say that μ is continuous at the point v if for every sequence v_k converging to v, we have $\langle \mu_{v_k} \rangle_v \to \langle \mu_v \rangle_v = \bar{\mu}_v$ in the weak topology. μ is continuous if it is continuous at every $v \in V$.

We will use the term **robust frames** to refer to continuous, weakly equivariant, weighted frames.

Proposition 3.6. *$\mathcal{I}_{\text{weighted}}[f]$ is a BEC operator iff μ is a robust frame.*

4. Weighted frames for permutations

In Corollary 2.9, we showed that the only (unweighted) frame which preserves continuity with respect to the action of the permutation group S_n on $\mathbb{R}^{d\times n}_{\text{distinct}}$ is the full Reynolds frame. In 4.1, we will show there is a *weighted frame* which preserves continuity in this case, with cardinality of only $n(d-1)$. In 4.2, we will discuss the harder task of continuity preservation for the action of S_n on all of $\mathbb{R}^{d\times n}$.

4.1. Robust frames for permutations on $\mathbb{R}^{d\times n}_{\text{distinct}}$

The frames we construct are based on one-dimensional sorting. For $X \in \mathbb{R}^{d\times n}_{\text{distinct}}$ and $a \in \mathbb{R}^d$, we say that X is a-separated if there exists a unique permutation τ such that

$$a^T x_{\tau(1)} < a^T x_{\tau(2)} < \ldots < a^T x_{\tau(n)}. \tag{2}$$

Our goal is to find a finite, relatively small number of vectors $a_1, \ldots, a_m$, such that every $X \in \mathbb{R}^{d\times n}_{\text{distinct}}$ is a_i separated for at least one i. When this occurs, we call $a_1, \ldots, a_m$ a *globally separated collection*. The next theorem shows that this is possible if and only if $m \geq n(d-1)$.

Theorem 4.1. *Let $n, d > 1$ be natural numbers. Then Lebesgue almost every $a_1, \ldots, a_{n(d-1)} \in \mathbb{R}^d$ form a globally separated collection. Conversely, every globally separated collection must contain at least $n(d-1)$ vectors.*

A similar result, with slightly higher cardinality, was obtained in (Ye et al., 2024), where this result is used to represent functions equivariant to the anti-symmetric group, for quantum chemistry simulations.

We now explain how we can construct a robust frame $\mu^{\text{separated}}$ on $\mathbb{R}^{d\times n}_{\text{distinct}}$ using a globally separated collection $a_1, \ldots, a_m$ from Theorem 4.1. Our frame will be of the form $\mu_X^{\text{separated}} = \sum_{i=1}^m w_i(X)\delta_{g_i^{-1}(X)}$, where $g_i(X)$ is a permutation τ satisfying (2). Note that if X is not separated in the direction a_i, then τ is not uniquely defined. In this case we choose τ arbitrarily, which is not a problem because we will define the weights $w_i(X)$ to be zero in this case:

$$\tilde{w}_i(X) = \min_{s\neq t} |a_i^T(x_s - x_t)|, \; w_i(X) = \frac{\tilde{w}_i(X)}{\sum_{j=1}^m \tilde{w}_j(X)}$$

Note that the division by $\sum_{j=1}^m \tilde{w}_j(X)$ is well-defined because $a_1, \ldots, a_m$ is a globally separated collection.

Lemma 4.2. *$\mu^{\text{separated}}$ is a robust frame for $\mathbb{R}^{d\times n}_{\text{distinct}}$.*

4.2. Robust frames for all of $\mathbb{R}^{d\times n}$

The frame $\mu^{\text{separated}}$ is robust on $\mathbb{R}^{d\times n}_{\text{distinct}}$, but cannot be extended to a robust frame on all of $\mathbb{R}^{d\times n}$. We now provide a robust frame $\mu^{\mathrm{S_n}}$ for all of $(\mathbb{R}^{d\times n}, S_n)$ which averages over permutations obtained from all possible directions, rather than considering a fixed collection of directions. For a given X, $\mu_X^{\mathrm{S_n}}$ is defined by assigning to each $g^{-1} \in S_n$ the weight

$$w_{g^{-1}}(X) = \mathbb{P}_{a\sim S^{d-1}}\left[g = \text{argsort}(a^T X)\right].$$

Here the probability $\mathbb{P}$ is over directions a distributed uniformly on S^{d-1}, and $\text{argsort}(w)$ is the unique permutation $g \in S_n$ which sorts w while preserving the ordering of equal entries.

We then have that μ^{S_n} is a robust frame of moderate size.

Proposition 4.3. *$\mu^{\mathrm{S_n}}$ is a robust frame for S_n acting on $\mathbb{R}^{d\times n}$, with cardinality bounded by $2 \cdot \sum_{k=0}^{d-1} \binom{\frac{n^2-n-2}{2}}{k}$.*

In most applications, $d << n$ (e.g. $d = 3$), in which case the bound above is $O(n^{2(d-1)})$. This is significantly worse than $\mu^{\text{separated}}$, which had cardinality of $n(d-1)$, but also a significant improvement over the $(n!)$-sized Reynolds operator. The frame μ^{S_n} may be too large to compute exactly, but can be implemented in an augmentation-like style by randomly drawing $a \in S^{d-1}$ to sort along.

Our final result for S_n acting on $\mathbb{R}^{d\times n}$ is a lower bound (Proposition C.1) on the cardinality of any robust frame. When $d \ll n$, the lower bound is $\sim (d-1)(n/2)$.

5. Weighted frames for rotations

We now show how to define robust frames for the action of $SO(d)$ on $\mathbb{R}^{d\times n}$ with cardinality $n(n-1)\cdots(n-d+2)$. We note that our $d=3$ construction is very similar to the "smooth frames" introduced in Pozdnyakov & Ceriotti (2023); in a sense, our contribution is to generalize their frames to all dimensions, and to formally define and prove their robustness. Using essentially the same idea, we also construct robust frames for the action of $O(d)$ on $\mathbb{R}^{d\times n}$ with cardinality $2n(n-1)\cdots(n-d+2)$ in Appendix D.

5.1. The case of $d=2$

We use the identification $\mathbb{R}^{2\times n}\equiv\mathbb{C}^n$ and work in complex notation for simplicity. We denote a vector in $\mathbb{C}^n$ as $Z=(z_1,\ldots,z_n)$, and define a weighted frame of the form

$$\mu_Z^{SO(2)}=\sum_{i=1}^n w_i(Z)\delta_{g_i(Z)}.$$

The group elements $g_i(Z)=z_i/\|z_i\|\in S^1$ are defined to be the phase of the i-th entry, so that multiplying by g_i^{-1} will rotate Z so that z_i is real and positive. This is not well-defined when $z_i=0$, in which case we somewhat arbitrarily set $g_i(Z)=1$.

The weight functions are defined as follows. We fix $\eta\in(0,1)$ and a continuous function ϕ_η which is zero on $(-\infty,\eta)$, one on $[1,\infty)$, and satisfies $0\le\phi_\eta(t)\le 1$ elsewhere. For $Z\neq 0$, we set $\tilde{w}_i(Z)=\phi_\eta\left(\frac{\|z_i\|}{\max_j\|z_j\|}\right)$. We then define $w_i(Z)=\frac{1}{n}$ for $Z=0_n$, and $\frac{\tilde{w}_i(Z)}{\sum_j\tilde{w}_j(Z)}$ otherwise.

The functions w_i are S^1 invariant, non-negative, and sum to one everywhere. They are continuous on $\mathbb{C}^n\setminus\{0_n\}$, and they ensure that at a point Z with some zero and some non-zero coordinates, only the non-zero coordinates will be "active". The frame $\mu^{SO(2)}$ does have a singularity at 0_n, where all coordinates are zero. However, this singularity is "harmless" because the stabilizer is the whole group S_1.

Proposition 5.1. *$\mu^{SO(2)}$ is a robust frame.*

Proof idea. We explain why $\mu^{SO(2)}$ is weakly equivariant, leaving the full proof for the appendix. At points $Z\neq 0_n$ this follows from the invariance of the weight functions w_i and the equivariance of $g_i(Z)$ at all points, except for points with $Z_i=0, X\neq 0_n$ for which $w_i(Z)=0$. For $Z=0_n$, weak equivariance follows from the fact that G_{0_n} is all of S^1. Thus, for any distribution μ on S^1, the average measure $\langle\mu\rangle_{0_n}$ is the same: the Haar measure on S^1. □

As a corollary of our general results on robust frames, we deduce that projecting a dense set of continuous functions Q using the invariant operator $\mathcal{I}_{\text{weighted}}$ induced from $\mu^{SO(2)}$, will give a dense set of continuous invariant functions $\mathcal{I}_{\text{weighted}}(Q)$. Explicitly, for a given $q\in Q$, the function $\mathcal{I}_{\text{weighted}}[q]$ will be of the form

$$\mathcal{I}_{\text{weighted}}[q](Z)=\sum_{i=1}^n w_i(Z)q\left(\frac{\bar{z}_i}{\|z_i\|}\cdot Z\right)$$

5.2. The case of $d\geq 3$

Generalizing the cases of $d=2$ and $d=3$, which essentially appear in (Pozdnyakov & Ceriotti, 2023), to higher dimensions requires a much more involved construction and proof, which is given in detail in Appendix D. The basic idea is to associate a weight to every sequence of r columns of $X\in\mathbb{R}^{d\times n}$, where $r=\min(d-1,\text{rank}(X))$. For each such sequence of columns, a rotation can be obtained by mapping the column vectors into a standard position, determined by a Gram-Schmidt orthogonalization with the given ordering. By carefully choosing the weights to vanish whenever this rotation is not uniquely defined (up to an element of G_X), we can ensure that the resulting frame is robust.

6. From Invariant to Equivariant Projections

We saw that in the invariant setting, robust frames induced a BEC projection operator $\mathcal{I}_{\text{weighted}}$. In the equivariant setting, the situation is more complex. First, note that weighted frames $\mu_{[\cdot]}$ which are *fully* equivariant do induce well-defined equivariant projection operators $\mathcal{E}_{\text{weighted}}$. However, this comes at a computational cost for inputs v where $|G_v|$ is large. When considering robust (and therefore only *weakly* equivariant) frames, the natural "equivariant" operator $\mathcal{E}_{\text{weighted}}$ may not produce equivariant functions (Example 6.2). However, we can remedy this by requiring that the backbone architecture parametrize only *stable functions*, which remain equivariant under $\mathcal{E}_{\text{weighted}}$. In Appendix E, we also define *stable frames* as an alternative approach, which can be applied to an arbitrary backbone architecture. In Appendix D and Appendix E we show how both of these ideas can be implemented efficiently to achieve *equivariant*, continuous universal models for $(\mathbb{R}^{d\times n},SO(d))$ for $d=2,3$. Below we will discuss only stable functions for $d=2$.

We call a function $f:V\to W$ *stable* if $G_v\subseteq G_{f(v)}$ $\forall v\in V$. Note that any equivariant function $f:V\to W$ is stable, since for every $s\in G_v$ we have $f(v)=f(sv)=sf(v)$.

Proposition 6.1. *Let $\mu_{[\cdot]}$ be a robust frame. The restriction of $\mathcal{E}_{\text{weighted}}$ to* stable *input functions is a continuity-preserving, bounded, equivariant projection.*

Example 6.2 ($SO(2)$ equivariance)**.** Let us return to the robust frame $\mu^{SO(2)}$ from Subsection 5.1 for the action of $S^1\cong SO(2)$ on $\mathbb{C}^n\cong\mathbb{R}^{2\times n}$. Note that $G_Z=\{e\}\forall Z\in\mathbb{C}^n$ except $Z=0_n$ (whose stabilizer is all of S^1). Thus $f:\mathbb{C}^n\to\mathbb{C}^n$ is stable if and only if $f(0_n)=0_n$.

We first note that applying $\mathcal{E}_{\text{weighted}}$ induced from $\mu^{SO(2)}$ to a function which is not stable gives a function $\mathcal{E}_{\text{weighted}}[f]$ which is also not stable, and therefore not equivariant:

$$\mathcal{E}_{\text{weighted}}[f](0) = \frac{1}{n}\sum_{j=1}^{n} f(0) = f(0) \neq 0.$$

By Proposition 6.1, this problem can be avoided if we apply $\mathcal{E}_{\text{weighted}}$ only to stable functions, which here means functions q satisfying $q(0_n) = 0_n$. This condition is easily enforced: if Q is a dense space of continuous functions $q : \mathbb{C}^n \to \mathbb{C}^n$, then we obtain a dense set of stable functions via $\hat{q}(Z) = q(Z) - q(0_n)$. Applying $\mathcal{E}_{\text{weighted}}$ to these $\hat{q}$ functions yields a continuous, universal, equivariant model.

7. Experiments

In this section, we provide experimental evidence showing the advantage of preserving both continuity and invariance using robust frames. We consider the action of the permutation group S_n on two dimensional point clouds and leave the investigation of other group actions to future work. In Appendix F.1, we also experimentally verify the presence of discontinuities in a trained canonicalization pipeline for point clouds from the equiadapt library (Kaba et al., 2022; Mondal et al., 2023).

7.1. Comparison of S_n-frames

In this experiment[2], we tested permutation invariant frames on the following classification toy problem. Starting from the MNIST dataset, we processed the image of each digit into a two-dimensional point cloud containing 100 points, ordered randomly. We then trained a standard multi-layer perceptron (MLP) to classify the corresponding digit from this collection of point clouds, with invariance enforced in one of five ways: no invariance, invariance using a discontinuous canonicalization (sorting along the x-axis), invariance using each of the two robust frames introduced in Section 4, and invariance using the Reynolds operator (i.e. averaging over the entire group). Due to their size, each of the weighted frames (including the Reynolds operator) was implemented using empirical averaging, with one randomly drawn sample in each train step and 1, 5, 10, or 25 samples for inference (i.e. testing). All models were trained for 60 epochs using SGD with momentum 0.9 and a step size of 0.01, dropping to 0.001 after 30 epochs. The network was an MLP with 3 hidden layers of sizes 150, 100, and 50, with an input size of 200 and output of size 10. The results of this experiment are shown in Table 2.

Invariance Method	Test Accuracy (%)
No Invariance	25.5
Discontinuous Canonicalization	85.6
Robust Frames (Sec. 4.1)	75.5 / 85.6 / 87.1 / 88.4
Robust Frames (Sec. 4.2)	74.2 / 85.9 / 87.6 / 88.7
Reynolds Operator	21.0 / 22.4 / 22.6 / 22.6

Table 2. Comparison between permutation canonicalization and various frames. The right hand column shows 1/5/10/25 samples drawn during testing for the weighted frames.

From these results, we draw the following conclusions. First, no invariance and the (sampled) Reynolds operator do not work well, since the permutation group is so large that both of these are essentially not enforcing any permutation invariance and the dataset is far too small to enable learning without the permutation symmetry enforced. Second, a discontinuous canonicalization performs much better than the prior two methods without canonicalization (since permutation invariance is now enforced), but lack of continuity still hurts the test accuracy relative to the robust frames, which enforce both continuity and invariance. We also see that enforcing continuity on the entire input space $\mathbb{R}^{d\times n}$ performs slightly better than only enforcing continuity at $\mathbb{R}^{d\times n}_{\text{distinct}}$. However, when empirically implementing robust frames, we do need to average quite a few samples from the frame during inference to obtain a good result.

[2]Code for reproducing this experiment can be found at https://github.com/jwsiegel2510/Sn-invariant-weighted-frames

8. Conclusion and open questions

In this work, we illuminated a critical problem with group canonicalization: it can destroy the continuity of the function being canonicalized. Moreover, even frames may have this problem if they aren't sufficiently large. As a solution, we introduced *robust* frames, which are not only weighted but also continuity-preserving. Robust frames also deal intelligently with self-symmetric inputs, a facet that has not to our knowledge been previously analyzed. Finally, we construct several examples of robust frames. As frames (whether learned or deterministic) become more prevalent in the world of equivariant learning, we hope that our results will provide a guiding light for practitioners.

Our work leaves open a few questions, such as stronger lower bounds on the cardinality of robust $SO(d)$ frames, and whether a continuous canonicalization exists for unordered point clouds. More broadly, one may ask under what conditions (and frame sizes) one can expect stronger notions of smoothness, such as bounded Lipschitz constants. In the equivariant case, it also remains to develop stable frames and/or functions for a wider variety of groups. Finally, Kim et al. (2023) and Mondal et al. (2023) *probabilistically* sample from their weighted frames; to analyze this setting, one might imagine concentration bounds replacing cardinality as the relevant measure of a frame's efficiency.

Acknowledgements ND is supported by Israel Science Foundation grant no. 272/23. HL is supported by the Fannie and John Hertz Foundation and the NSF Graduate Fellowship under Grant No. 1745302. JWS is supported by the National Science Foundation (DMS-2424305 and CCF-2205004) as well as the Office of Naval Research (MURI ONR grant N00014-20-1-2787).

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. We do not foresee any direct negative societal impact resulting from this work

References

Amir, T., Gortler, S., Avni, I., Ravina, R., and Dym, N. Neural injective functions for multisets, measures and graphs via a finite witness theorem. In Oh, A., Neumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S. (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 42516–42551. Curran Associates, Inc., 2023.

Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J. P., Kornbluth, M., Molinari, N., Smidt, T. E., and Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1), 2022. doi: 10.1038/s41467-022-29939-5. URL https://par.nsf.gov/biblio/10381731.

Bietti, A., Venturi, L., and Bruna, J. On the sample complexity of learning under geometric stability. In Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=vlf0zTKa5Lh.

Borsuk, K. Drei sätze über die n-dimensionale euklidische sphäre. *Fundamenta Mathematicae*, 20(1):177–190, 1933.

Cahill, J., Iverson, J. W., Mixon, D. G., and Packer, D. Group-invariant max filtering. *Foundations of Computational Mathematics*, pp. 1–38, 2024.

Corso, G., Stark, H., Jing, B., Barzilay, R., and Jaakkola, T. Diffdock: Diffusion steps, twists, and turns for molecular docking. In *ICLR*, 2023.

Du, W., Zhang, H., Du, Y., Meng, Q., Chen, W., Zheng, N., Shao, B., and Liu, T.-Y. SE(3) equivariant graph neural networks with complete local frames. In *ICML*, 2022.

Du, W., Du, Y., Wang, L., Feng, D., Wang, G., Ji, S., Gomes, C., and Ma, Z. A new perspective on building efficient and expressive 3d equivariant graph neural networks. In *NeurIPS*, 2023.

Duval, A., Schmidt, V., Garcia, A. H., Miret, S., Malliaros, F. D., Bengio, Y., and Rolnick, D. Faenet: Frame averaging equivariant gnn for materials modeling. In *ICML*, 2023.

Dym, N. and Gortler, S. J. Low-dimensional invariant embeddings for universal geometric learning. *Foundations of Computational Mathematics*, pp. 1–41, 2024.

Dym, N. and Maron, H. On the universality of rotation equivariant point cloud networks. In *International Conference on Learning Representations*, 2020.

Elesedy, B. Provably strict generalisation benefit for invariance in kernel methods. *Advances in Neural Information Processing Systems*, 34:17273–17283, 2021.

Frasca, F., Bevilacqua, B., Bronstein, M., and Maron, H. Understanding and extending subgraph gnns by rethinking their symmetries. In *NeurIPS*, 2022.

Hatcher, A. *Algebraic topology*. Cambridge University Press, Cambridge, 2002. ISBN 0-521-79160-X; 0-521-79540-0.

Hopf, H. Über die abbildungen der dreidimensionalen sphäre auf die kugelfläche. *Mathematische Annalen*, 104 (1):637–665, 1931.

Hordan, S., Amir, T., and Dym, N. Weisfeiler leman for euclidean equivariant machine learning. *arXiv e-prints*, pp. arXiv–2402, 2024.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

Kaba, S.-O., Mondal, A., Zhang, Y., Bengio, Y., and Ravanbakhsh, S. Equivariance with learned canonicalization functions. In *Symmetry and Geometry in Neural Representations Workshop*, 2022.

Kim, J., Nguyen, T. D., Suleymanzade, A., An, H., and Hong, S. Learning probabilistic symmetrization for architecture agnostic equivariance. In *NeurIPS*, 2023.

Kurlin, V. Polynomial-time algorithms for continuous metrics on atomic clouds of unordered points. *Match: Communications in Mathematical and in Computer Chemistry*, 2023.

Liao, Y.-L., Wood, B., Das, A., and Smidt, T. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. 2023. URL https://arxiv.org/abs/2306.12059.

Luo, S., Li, J., Guan, J., Su, Y., Cheng, C., Peng, J., and Ma, J. Equivariant point cloud analysis via learning orientations for message passing. In *CVPR*, 2022.

Mei, S., Misiakiewicz, T., and Montanari, A. Learning with invariances in random features and kernel models. In *Conference on Learning Theory*, pp. 3351–3418. PMLR, 2021.

Mondal, A. K., Panigrahi, S. S., Kaba, S., Rajeswar, S., and Ravanbakhsh, S. Equivariant adaptation of large pretrained models. In *NeurIPS*, 2023.

Passaro, S. and Zitnick, C. L. Reducing SO(3) convolutions to SO(2) for efficient equivariant GNNs. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J. (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 27420–27438. PMLR, 23–29 Jul 2023.

Petrache, M. and Trivedi, S. Approximation-generalization trade-offs under (approximate) group equivariance. *Advances in Neural Information Processing Systems*, 36, 2024.

Pozdnyakov, S. and Ceriotti, M. Smooth, exact rotational symmetrization for deep learning on point clouds. 2023.

Puny, O., Atzmon, M., Ben-Hamu, H., Smith, E. J., Misra, I., Grover, A., and Lipman, Y. Frame averaging for invariant and equivariant network design. *CoRR*, abs/2110.03336, 2021.

Satorras, V. G., Hoogeboom, E., and Welling, M. E(n) equivariant graph neural networks. In *Proceedings of the 38th International Conference on Machine Learning*. PMLR, 2021.

Urysohn, P. Über die mächtigkeit der zusammenhängenden mengen. *Mathematische Annalen*, 94(1):262–295, 1925.

Villar, S., Hogg, D. W., Storey-Fisher, K., Yao, W., and Blum-Smith, B. Scalars are universal: Gauge-equivariant machine learning, structured like classical physics. *CoRR*, abs/2106.06610, 2021. URL https://arxiv.org/abs/2106.06610.

Widdowson, D. and Kurlin, V. 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*, pp. 1275–1284, 2023.

Yarotsky, D. Universal approximations of invariant maps by neural networks. *Constructive Approximation*, 55(1): 407–474, 2022.

Ye, H., Li, R., Gu, Y., Lu, Y., He, D., and Wang, L. $\widetilde{O}(n^2)$ representation of general continuous anti-symmetric function, 2024.

Zhang, Y., Hare, J. S., and Prügel-Bennett, A. Learning representations of sets through optimized permutations. In *7th International Conference on Learning Representations, ICLR*, 2019.

Zhang, Y., Hare, J. S., and Prügel-Bennett, A. Fspool: Learning set representations with featurewise sort pooling. In *8th International Conference on Learning Representations, ICLR*, 2020.

.

Appendix structure

The structure of the appendix is as follows. Appendix A details some additional mathematical assumptions and background.

We then discuss results which were not fully stated in the paper: Appendix B proves the existence of continuous canonicalizations for $SO(d)$ when $n < d$, and for $O(d)$ when $n \leq d$. Appendix C proves a lower bound on the cardinality of robust frames for the module $(\mathbb{R}^{d\times n}, S_n)$ with $d > 1$. Appendix D constructs robust frames for the actions of $SO(d)$ and $O(d)$, and Appendix E discusses stable frames.

The last and largest appendix G contains the proofs of all claims in the paper, listed in chronological order.

A. Additional Background

As stated in the main text, throughout the paper we considered compact groups G acting linearly and continuously on (typically) finite dimensional real vector spaces V and W, or else on subsets of V and W closed under the action of G. These are common assumptions, essentially the same as the G-modules used in the definition of a module in (Yarotsky, 2022). In this appendix we lay out in more detail what a module (V, G) means:

1. We assume V is a finite dimensional real Hilbert space, i.e. a finite dimensional vector space endowed with a positive definite inner product.
2. G is a compact group. That is, G is a group endowed with a topology under which G is a compact Hausdorff space, and moreover the multiplication and inverse operations are continuous.
3. G acts on V and for every fixed g, the map $g : V \to V$ is a linear transformation, and the map $(g, v) \mapsto gv$ is continuous. Note that this is equivalent to a continuous group homomorphism $G \to GL(V)$, where $GL(V)$ denotes the general linear group of V, i.e. the group of all invertible linear transformations of V.

Definition A.1. If G acts on a set V and $V' \subseteq V$, then we say that V' is a G**-stable set**, or that V' is closed under the action of G, if

$$gV' := \{gv,\ v \in V'\} \subset V'.$$

Note that a group action on a set V induces a well-defined group action any G-stable subset $V' \subset V$. In our discussion in the paper, in the (V, G) pairs we discuss, we allow V to either be the whole vector space or a G stable subset.

B. Continuous orthogonal canonicalizations

In this section, we show that Proposition 2.7 is sharp, i.e. that there exists a continuous canonicalization for $SO(d)$ when $n < d$ and for $O(d)$ when $n \leq d$. The key to this is the following construction.

Proposition B.1. *The module* $(\mathbb{R}^{d\times d}, O(d))$ *has a continuous canonicalization.*

Proof. We first recall the classical fact that the positive semi-definite Gram matrix $X^T X$ is a complete invariant for the action of $O(d)$ on $X \in \mathbb{R}^{d\times d}$. We define a continuous canonicalization $y : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$ by

$$y_X = (X^T X)^{1/2}. \tag{3}$$

Here the square root is the standard square root of a positive semi-definite matrix, i.e. $A^{1/2}$ is the unique positive semi-definite matrix B such that $B^2 = A$. It is clear that this is a canonicalization by construction, since y_X is symmetric so that $(y_X)^T y_X = (y_X)^2 = X^T X$.

The continuity follows immediately from the continuity of the matrix square root on the set of positive semi-definite matrices. This fact is elementary, and can be proven for instance using the Taylor series expansion

$$M^{1/2} = \sum_{n=0}^{\infty} (-1)^{n-1} \frac{(2n)!}{4^n (n!)^2 (2n-1)} (M - I_d)^n. \tag{4}$$

Using Sterling's formula, we easily see that the coefficients satisfy

$$\frac{(2n)!}{4^n (n!)^2 (2n-1)} = O\left(\frac{1}{n^{3/2}}\right),$$

which implies that the series (4) converges absolutely whenever $\|M - I_d\| \leq 1$ (here $\|\cdot\|$ denotes the operator norm so that $\|X^k\| \leq \|X\|^k$). This implies that the matrix square root is a continuous function for all positive semi-definite matrices M such that $0 \preceq M \preceq 2I_d$, since for such matrices we clearly have $\|M - I_d\| \leq 1$. Finally, the homogeneity of the matrix square root extends this continuity to all positive semi-definite matrices. □

By appending zero columns to X, this immediately implies that $O(d)$ has a continuous canonicalization whenever $n \leq d$. We can also use it in a straightfoward manner to obtain a continuous canonicalization for $SO(d)$ acting on $\mathbb{R}^{d\times n}$ when $n < d$.

Corollary B.2. *The module* $(\mathbb{R}^{d\times(d-1)}, SO(d))$ *has a continuous canonicalization.*

Proof. This follows since the orbits of $\mathbb{R}^{d\times(d-1)}$ under the action of $SO(d)$ and $O(d)$ are the same. Indeed, if X, Y are in the same orbit under the action of $SO(d)$ they are clearly in the same orbit under the action of $O(d)$ since $SO(d) \subset O(d)$. On the other hand, suppose that X, Y are in the same orbit under the action of $O(d)$, i.e. that there exists a $U \in O(d)$ such that $Y = UX$. Let R_Y be a reflection which leaves the space spanned by Y invariant (this exists since Y consists of $d-1$ vectors). Then $Y = R_Y UX$ and either $U \in SO(d)$ or $R_Y U \in SO(d)$. Thus X, Y are in the same orbit under $SO(d)$.

Since the orbits of $\mathbb{R}^{d\times(d-1)}$ under the action of $SO(d)$ and $O(d)$ are the same, the canonicalization for $O(d)$ acting on $\mathbb{R}^{d\times(d-1)}$ gives a canonicalization for $SO(d)$ as well. □

C. Lower bound on permutation robust frames

In this appendix, we state and prove the precise lower bound on the cardinality of a robust frame for $(\mathbb{R}^{d\times n}, S_n)$ with $d > 1$ mentioned in Section 4.

Proposition C.1. *Any robust frame for the module $(\mathbb{R}^{d\times n}, S_n)$ with $d > 1$ has cardinality at least*

$$k(d,n) := (d-1)\lfloor n/2\rfloor + 1 - \sum_{i=1}^{\lfloor n/2\rfloor}(d-1-2^i)_+.$$

Proof. Suppose without loss of generality that n is even. If not, we restrict to the set where the first vector is 0.

Suppose that we are given a robust frame $\mu : \mathbb{R}^{d\times n} \to \Pi(S_n)$, where $\Pi(S_n)$ denotes the space of probability measures on the symmetric group S_n. For an element $X \in \mathbb{R}^{d\times n}$, we let $\tilde{\mu}_X$ denote the pushforward of μ_X under the coset map $G \to G/G_X$. Thus $\tilde{\mu}_X$ is a measure on the set of cosets G/G_X. We will show that there exists a point $X \in \mathbb{R}^{d\times n}$ such that

$$|\mathrm{supp}(\tilde{\mu}_X)| \geq \frac{(d-1)n}{2} + 1 - \sum_{i=1}^{n/2}(d-1-2^i)_+.$$

Let $x_1, ..., x_{n/2} \in \mathbb{R}^d$ be distinct and consider the point

$$X_0 = (x_1, x_1, x_2, x_2, \cdots, x_{n/2}, x_{n/2}) \in \mathbb{R}^{dn}. \tag{5}$$

We will inductively construct a sequence of points $X_1, ..., X_{n/2}$ of the form

$$X_i = (x_1', x_1^*, \cdots x_i', x_i^*, x_{i+1}, x_{i+1}, \cdots, x_{n/2}, x_{n/2}), \tag{6}$$

where $x_i' \neq x_i^*$ are close to x_i and such that $|\mathrm{supp}(\tilde{\mu}_{X_i})| \geq |\mathrm{supp}(\tilde{\mu}_{X_{i-1}})| + \min\{|\mathrm{supp}(\tilde{\mu}_{X_{i-1}})|, d-1\}$. Since $|\mathrm{supp}(\tilde{\mu}_{X_0})| \geq 1$, we get

$$|\mathrm{supp}(\tilde{\mu}_{X_{n/2}})| \geq \frac{(d-1)n}{2} + 1 - \sum_{i=1}^{n/2}(d-1-2^i)_+,$$

as desired.

Suppose that the point X_i can been constructed and let $m = \min\{|\mathrm{supp}(\tilde{\mu}_{X_i})|, d-1\}$. Consider points of the form

$$X(v_\epsilon) = (v_1', v_1^*, \cdots v_i', v_i^*, v_{i+1} + v_\epsilon, v_{i+1} - v_\epsilon, \cdots, v_{n/2}, v_{n/2}) \tag{7}$$

for a vector $v_\epsilon \in \mathbb{R}^d$. By definition, the continuity of the frame implies that

$$\lim_{v_\epsilon \to 0}\langle \mu_{X(v_\epsilon)}\rangle_{X_i} = \bar{\mu}_{X_i} \tag{8}$$

in the weak topology, which coincides with pointwise convergence of the probabilities since S_n is a finite group.

Let $C_1, ..., C_m \in S_n/G_{X_i}$ be distinct cosets in the support of $\tilde{\mu}_{X_i}$, i.e. we have $\mu_{X_i}(C_j) > 0$. Equation (8) implies that for sufficiently small v_ϵ, we will have $\mu_{X(v_\epsilon)}(C_j) > 0$ for all $j = 1, ..., m$.

Observe that for any $v_\epsilon \neq 0$, the stabilizer $H := G_{X(v_\epsilon)}$ is independent of v_ϵ and has index 2 in G_{X_i}. This means that each coset C_j splits into two cosets of the smaller subgroup H, which we denote by $C_j^+, C_j^- \in G/H$. Define the following function $f : \mathbb{R}^d \to \mathbb{R}^m$

$$f(v_\epsilon)_j = \frac{\mu_{X(v_\epsilon)}(C_j^+)}{\mu_{X(v_\epsilon)}(C_j)}. \tag{9}$$

The function f is well-defined for sufficiently small inputs v_ϵ, since by the previous remark the denomintor is then > 0. Moreover, the continuity of the frame μ implies that f is a continuous function of $v_\epsilon \neq 0$ in such a sufficiently small neighborhood (since the stabilizer H of $X(v_\epsilon)$ is constant in this neighborhood and so the continuity condition becomes continuity of the averaged frame $\bar{\mu}_{X(v_\epsilon)}$). In addition, the invariance of the frame means that

$$f(-v_\epsilon)_j = 1 - f(v_\epsilon)_j \tag{10}$$

since transposing the $2i + 1$ and $2i + 2$ elements of X maps $X(v_\epsilon)$ to $X(-v_\epsilon)$ and also swaps the cosets C_j^+ and C_j^-.

We complete the proof by noting that since $m \leq d - 1$, the Borsuk-Ulam Theorem (Borsuk, 1933) implies that there must be a point v_ϵ such that $f(v_\epsilon) = f(-v_\epsilon)$. Combined with (10) this means that $f(v_\epsilon)_j = 1/2$ for all j, and so

$$\mu_{X(v_\epsilon)}(C_j^+) = \mu_{X(v_\epsilon)}(C_j^-) = 1/2 > 0.$$

Choosing $X_{i+1} = X(v_\epsilon)$ then completes the inductive step. □

D. Frames for $SO(d)$ and $O(d)$ acting on $\mathbb{R}^{d\times n}$

In this appendix, we construct robust frames for $(\mathbb{R}^{d\times n}, SO(d))$ and for $(\mathbb{R}^{d\times n}, O(d))$ using similar ideas to those laid out in (Pozdnyakov & Ceriotti, 2023). Our contribution is to generalize this construction to all dimensions and to rigorously prove that it preserves continuity.

We begin first with the action of $SO(d)$. For a point cloud $0 \neq X \in \mathbb{R}^{d\times n}$, the frame $\mu := \mu^{SO(d)}$ will be of the form

$$\mu_X = \sum_{i_1=1}^{n} w_{i_1}(X) \sum_{i_2=1}^{n} w_{i_1 i_2}(X) \cdots \sum_{i_r=1}^{n} w_{i_1 i_2 \cdots i_r}(X) \delta_{g_{i_1 i_2 \cdots i_r}(X)}, \tag{11}$$

where $r = \min(\text{rank}(X), d-1)$, $w_{i_1 i_2 \cdots i_t}(X)$ for t is a weight associated to the sequence of columns $i_1, i_2, \cdots i_t$ for $t \leq r$, and $g_{i_1 i_2 \cdots i_r}(X) \in SO(d)$ is a rotation associated to the sequence of columns $i_1, i_2, \cdots i_r$. When $X = 0$, the frame μ_X can be chosen arbitrarily.

We proceed to describe the weight functions $w_{i_1 \cdots i_t}(X)$ for $t = 1, ..., r$ and the rotations $g_{i_1 \cdots i_r}(X)$.

The rotations $g_{i_1 \cdots i_r}(X)$ are defined by

$$g_{i_1 \cdots i_r}(X)^{-1} \left(x_{i_1}, x_{i_2}, ..., x_{i_r}\right) = A, \tag{12}$$

where A is an upper triangular $d \times r$ matrix with non-negative diagonal entries which satisfies

$$A^T A = \begin{pmatrix} \langle x_{i_1}, x_{i_1}\rangle & \cdots & \langle x_{i_1}, x_{i_r}\rangle \\ \vdots & \ddots & \vdots \\ \langle x_{i_r}, x_{i_1}\rangle & \cdots & \langle x_{i_r}, x_{i_r}\rangle \end{pmatrix}. \tag{13}$$

The matrix A is uniquely determined if the columns $x_{i_1}, ..., x_{i_r}$ are linearly independent. If not, we simply choose one such A, and remark that in this case the choice will not matter because the corresponding weights will be equal to 0.

Another way of thinking about the matrix A is that it is determined by performing Gram-Schmidt orthogonalization on the column vectors $x_{i_1}, x_{i_2}, ..., x_{i_r}$ to obtain an orthonormal set $\hat{x}_{i_1}, ..., \hat{x}_{i_r}$. If the vectors $x_{i_1}, x_{i_2}, ..., x_{i_r}$ are linearly dependent, we must modify Gram-Schmidt as follows. If $x_{i_t} \in \text{span}(x_{i_1}, ..., x_{i_{t-1}})$, then in the t-th step of Gram Schmidt we simply choose $\hat{x}_{i_t}$ to be an arbitrary unit vector orthogonal to $\hat{x}_{i_1}, ..., \hat{x}_{i_{t-1}}$ (this is where the non-uniqueness comes in).

If we perform this modified Gram-Schmidt orthogonalization procedure on the columns $x_{i_1}, ..., x_{i_r}$ to obtain an orthonormal set $\hat{x}_{i_1}, ..., \hat{x}_{i_r}$, which we then complete to an orthonormal basis, then the columns of A correspond to the representation of $x_{i_1}, ..., x_{i_r}$ with respect to this basis. From this it becomes clear that the first t columns of A only depend upon $x_{i_1}, ..., x_{i_t}$, and that the first t columns of A are a continuous function of $x_{i_1}, ..., x_{i_t}$ on the set where these vectors are linearly independent. These facts will become important later in the proof of continuity.

Since the vectors $x_{i_1}, ..., x_{i_r}$ have the same inner products as the first r columns in A, there exists an orthogonal transformation satisfying (12). Moreover, because $r \leq d-1$ this orthogonal transformation can be chosen to lie in $SO(d)$ (by reflecting across the plane spanned by the columns of A if necessary). Thus a rotation $g_{i_1 \cdots i_r}(X)$ satisfying (12) always exists, although it is only uniquely defined up to left multiplication by the stabilizer of the columns $x_{i_1}, x_{i_2}, ..., x_{i_r}$. In defining $g_{i_1 \cdots i_r}(X)$ we simply choose any rotation satisfying (12). As we will see, the weights $w_{i_1, ..., i_t}$ will be chosen so that if $x_{i_1}, x_{i_2}, ..., x_{i_r}$ have a larger stabilizer than the whole point cloud X, then the total weight corresponding to $g_{i_1 \cdots i_r}(X)$ will be 0.

Next, we describe the weight functions. If $x_{i_1}, ..., x_{i_{t-1}}$ are linearly independent, the weights $w_{i_1 \cdots i_t}(X)$ are defined by

$$w_{i_1 \cdots i_t}(X) = \frac{\Delta(x_{i_1}, ..., x_{i_t})}{\sum_{j=1}^{n} \Delta(x_{i_1}, ..., x_{i_{t-1}}, x_j)}, \tag{14}$$

where

$$\Delta(v_1, ..., v_t) = \sqrt{\left| \det \begin{pmatrix} \langle v_1, v_1\rangle & \cdots & \langle v_1, v_t\rangle \\ \vdots & \ddots & \vdots \\ \langle v_t, v_1\rangle & \cdots & \langle v_t, v_t\rangle \end{pmatrix} \right|}$$

is the area of the parallelopiped spanned by $v_1, ..., v_t$. We remark that if $x_{i_1}, ..., x_{i_{t-1}}$ are linearly independent, then since X has rank at least t there is some j such that $\Delta(x_{i_1}, ..., x_{i_{t-1}}, x_j) > 0$ so that the weight in (14) is well-defined. We remark that we could have used a cutoff function ϕ_η as in Section 5.1 and (Pozdnyakov & Ceriotti, 2023) in the definition of the weights (14). However, for simplicity of presentation we stick with the raw areas, although a cutoff function may be desirable in a practical implementation. The following construction and proof carry over easily to the case where a cutoff function is used with minor modifications.

If $x_{i_1}, ..., x_{i_{t-1}}$ are linearly dependent, then we simply set $w_{i_1\cdots i_{t-1}j}(X) = 1/n$ for $j = 1, ..., n$. From these definitions, it is clear that all weights are non-negative, and for any indices $i_1, ..., i_{t-1}$ we have

$$\sum_{i_t=1}^{n} w_{i_1 i_2 \cdots i_t}(X) = 1. \tag{15}$$

From this it follows that

$$\sum_{i_1=1}^{n} w_{i_1}(X) \sum_{i_2=1}^{n} w_{i_1 i_2}(X) \cdots \sum_{i_r=1}^{n} w_{i_1 i_2 \cdots i_r}(X) = 1, \tag{16}$$

so that μ is a well-defined frame. Moreover, the cardinality of μ is equal to the maximum number of sequences of indices $i_1, ..., i_r$ for $r \leq d-1$ such that $x_{i_1}, ..., x_{i_r}$ are linearly independent. Clearly in this case we can have no repeated indices so that the cardinality of μ is $n(n-1)\cdots(n-d+2)$. The main result of this Section is that the frame μ defined in this way is a robust frame.

Proposition D.1. *The frame $\mu^{SO(d)}$ is a robust frame.*

Proof. We first verify that the frame $\mu := \mu^{SO(d)}$ is weakly equivariant. Let $g \in SO(d)$ and $X \in \mathbb{R}^{d\times n}$. Since the action of $SO(d)$ preserves both inner products and the rank of X, it follows that all of the weights in (11) are invariant under the action of g. Also, since inner products are preserved it follows from (12) and (13) that

$$g_{i_1\cdots i_r}(X)^{-1}\left(x_{i_1}, x_{i_2}, ..., x_{i_r}\right) = g_{i_1\cdots i_r}(gX)^{-1} g \left(x_{i_1}, x_{i_2}, ..., x_{i_r}\right). \tag{17}$$

Observe also that if we define total weights via

$$W_{i_1,...,i_r}(X) = w_{i_1}(X) w_{i_1 i_2}(X) \cdots w_{i_1\cdots i_r}(X), \tag{18}$$

then (14) implies that $W_{i_1,...,i_r}(X) > 0$ iff $G_X = G_{(x_1,...,x_r)}$, i.e. if the whole point cloud X and the columns $(x_1, ..., x_r)$ have the same stabilizer. Indeed, $W_{i_1,...,i_r}(X) > 0$ iff $x_{i_1}, ..., x_{i_r}$ span a subspace of dimension r, which must coincide with the span of X if X has rank $< d$. In this case, the stabilizer of X and $(x_{i_1}, ..., x_{i_r})$ consist of all rotations that fix this subspace. If X has rank d, then its stabilizer is trivial, and $x_{i_1}, ..., x_{i_r}$ span a space of dimension $d-1$ so that the stabilizer consists of all rotations which fix this $(d-1)$-dimensional subspace. But any rotation fixing a $(d-1)$-dimensional subspace must be trivial so that the stabilizer of $(x_{i_1}, ..., x_{i_r})$ is also trivial in this case.

Utilizing this, we see that (17) implies that

$$g_{i_1\cdots i_r}(X)^{-1} G_X = g_{i_1\cdots i_r}(gX)^{-1} g G_X \implies G_X g_{i_1\cdots i_r}(X) = G_X g^{-1} g_{i_1\cdots i_r}(gX) \tag{19}$$

for every sequence of indices $(i_1, ..., i_r)$ for which $W_{i_1,...,i_r}(X) > 0$.

Plugging the invariance of the weights into (11), we see upon averaging over the stabilizer that

$$\bar{\mu}(gX) = \int_{G_{gX}} \sum_{i_1,...,i_r=1}^{n} W_{i_1,...,i_r}(X) \delta_{s g_{i_1\cdots i_r}(gX)} ds. \tag{20}$$

Now, we use that the stabilizers satisfy $G_{gX} = g G_X g^{-1}$ and the relation (19) (which holds whenever $W_{i_1,...,i_r}(X) > 0$) to rewrite this as

$$\bar{\mu}(gX) = \int_{G_X} \sum_{i_1,...,i_r=1}^{n} W_{i_1,...,i_r}(X) \delta_{gsg^{-1} g_{i_1\cdots i_r}(gX)} ds = \int_{G_X} \sum_{i_1,...,i_r=1}^{n} W_{i_1,...,i_r}(X) \delta_{gs g_{i_1\cdots i_r}(X)} ds = g^* \bar{\mu}(X), \tag{21}$$

as desired.

Next, we prove that the frame μ is continuous. To do this, fix $X \in \mathbb{R}^{d\times n}$ and suppose that X has rank r. If $r = 0$, i.e. if $X = 0$, then the stabilizer of X is all of $SO(d)$ and so continuity at X follows trivially from Definition 3.5 since we are averaging over the whole group.

So suppose that X has rank $r > 0$. We first observe that for sufficiently small $\epsilon > 0$ (depending upon X), $|Y - X| < \epsilon$ implies that if $x_{i_1}, ..., x_{i_r}$ are linearly independent, then $y_{i_1}, ..., y_{i_r}$ are also linearly independent. In particular, $|Y - X| < \epsilon$ implies that rank$(Y) \geq r$. For such a Y, the 'marginal' weights

$$W_{i_1,...,i_r}(Y) = w_{i_1}(Y)w_{i_1 i_2}(Y)\cdots w_{i_1\cdots i_r}(Y) = \sum_{i_{r+1}=1}^{n}\sum_{i_R=1}^{n} W_{i_1,...,i_R}(Y), \tag{22}$$

where $R = \text{rank}(Y)$ are well-defined. We first claim that

$$\lim_{Y\to X} W_{i_1,...,i_r}(Y) = W_{i_1,...,i_r}(X) \tag{23}$$

for any sequence of indices $i_1, ..., i_r$. To prove this, suppose first that $W_{i_1,...,i_r}(X) > 0$. In this case, $x_{i_1}, x_{i_2}, ..., x_{i_r}$ are linearly independent, so that by (14) each of the weight functions $w_{i_i,...,i_t}(X)$ for $1 \leq t \leq r$ is continuous in a neighborhood of X. This implies (23). If on the other hand $W_{i_1,...,i_r}(X) = 0$, then let t be the first index such that $w_{i_i,...,i_t}(X) = 0$. Then $x_{i_1}, ..., x_{i_{t-1}}$ are linearly independent and (14) shows that $w_{i_i,...,i_t}(X)$ is continuous in a neighborhood of X. This means that

$$\lim_{Y\to X} w_{i_i,...,i_t}(Y) = 0.$$

Since the weight functions are all bounded, we get

$$\lim_{Y\to X} W_{i_1,...,i_r}(Y) = 0 = W_{i_1,...,i_r}(X)$$

as desired.

The final ingredient we need to prove continuity is to observe that if $R = \text{rank}(Y) \geq r$, then for any indices $i_1, ..., i_R$ the product

$$g_{i_1,...,i_R}(Y)^{-1}\left(y_{i_1}, y_{i_2}, ..., y_{i_r}\right) \tag{24}$$

is independent of final indices $i_{r+1}, ..., i_R$. This is due to the fact (mentioned earlier) that the first r columns of the matrix A in (13) only depend upon the first r vectors $y_{i_1}, y_{i_2}, ..., y_{i_r}$. Moreover, as mentioned earlier we also have that the product in (24) (i.e. the first r columns of the matrix A in (13)) is a continuous function of $y_{i_1}, ..., y_{i_r}$ on the set where $y_{i_1}, ..., y_{i_r}$ are linearly independent.

We can now complete the proof of continuity. Let $i_1, ..., i_r$ be a sequence of indices such that $W_{i_1,...,i_r}(X) > 0$. This means that $x_{i_1}, ..., x_{i_r}$ are linearly independent and if $|Y - X| < \epsilon$, then $y_{i_1}, ..., y_{i_r}$ are linearly independent as well. The continuity of the Gram-Schmidt procedure (assuming linear independence) implies that for any set of indices $i_{r+1}, ..., i_R$ where R is the rank of Y, we have

$$\lim_{Y\to X} g_{i_1,...,i_R}(Y)^{-1}\left(y_{i_1}, y_{i_2}, ..., y_{i_r}\right) = g_{i_1,...,i_r}(X)^{-1}\left(x_{i_1}, x_{i_2}, ..., x_{i_r}\right). \tag{25}$$

Note that here the rank R may depend upon Y in the above limit. Since $y_i \to x_i$ and every $g \in SO(d)$ is an isometry, this implies that

$$\lim_{Y\to X}[g_{i_1,...,i_R}(Y)^{-1} - g_{i_1,...,i_r}(X)^{-1}]\left(x_{i_1}, x_{i_2}, ..., x_{i_r}\right) = 0. \tag{26}$$

Here the left term is viewed as a matrix in $\mathbb{R}^{d\times d}$. Since $x_{i_1}, ..., x_{i_r}$ is a basis for the range of X, we get that

$$\lim_{Y\to X}[g_{i_1,...,i_R}(Y)^{-1} - g_{i_1,...,i_r}(X)^{-1}]v = 0 \tag{27}$$

uniformly for v in any compact subset of range(X). If $r = d - 1$, then this actually holds with range(X) replaced by $\mathbb{R}^d$ since both matrices on the left hand side above are in $SO(d)$. In either case, by averaging over G_X we get that

$$\lim_{Y\to X}\int_{h\in G_X}[g_{i_1,...,i_R}(Y)^{-1} - g_{i_1,...,i_r}(X)^{-1}]hv = 0 \tag{28}$$

uniformly on compact subsets of the whole space $\mathbb{R}^d$, since either $\int_{h\in G_X} hv \in \text{range}(X)$ for all $v \in \mathbb{R}^d$, or $r = d-1$ and (27) already holds for all $v \in \mathbb{R}^d$. Putting this together, we get that

$$\lim_{Y\to X} \int_{h\in G_X} \delta_{h^{-1}g_{i_1,\ldots,i_R}(Y)} dh = \int_{h\in G_X} \delta_{h^{-1}g_{i_1,\ldots,i_r}(X)} dh \tag{29}$$

in the weak topology on $SO(d)$. Since this holds for all sets of indices $i_1, ..., i_R$ for which $W_{i_1,...,i_r}(X) > 0$, and we have $\lim_{Y\to X} W_{i_1,...,i_r}(Y) = W_{i_1,...,i_r}(X)$ for all indices $i_1, .., i_r$, we finally get using the definition (11) that

$$\langle \mu_Y \rangle_X = \int_{h\in G_X} h^*\mu_Y \to \int_{h\in G_X} h^*\mu_X = \langle \mu_X \rangle_X \tag{30}$$

as $Y \to X$, in the weak topology on $SO(d)$, which proves the continuity of the frame μ. □

We remark that essentially the same construction can be used to obtain a frame for the action of $O(d)$ on $\mathbb{R}^{d\times n}$ which has cardinality $2n(n-1)\cdots(n-d+2)$. Specifically, setting $r = \text{rank}(X)$ in (11), and repeating the exact same argument we get a robust frame for $O(d)$. This frame has cardinality $2n(n-1)\cdots(n-d+2)$, since the orthogonal transformation $g_{i_1\cdots i_d}(X)$ is determined up to a reflection through the plane spanned by $x_{i_1}, ..., x_{i_{d-1}}$ when the columns $x_{i_1}, ..., x_{i_d}$ are linearly independent.

Finally, we remark that we do not know whether these constructions are optimal. In specific instances, such as when $n = d-1$ for $SO(d)$, we know they are sub-optimal since in this case a canonicalization exists (see Corollary B.2. Determining the smallest possible weighted frames in these cases is an interesting further research direction.

D.1. $SO(3)$ and $O(3)$ equivariance via stable functions

In this section, we specialize to the case of $\mathbb{R}^3$ and discuss how to endorce equivariance using the previously constructed frames. To achieve continuous, equivariant, universal models for functions $f : \mathbb{R}^{3\times n} \to \mathbb{R}^{3\times n}$ with respect to the action of $O(3)$ or $SO(3)$, we need to characterize the space of stable functions. This can be done using results from Villar et al. (2021). Namely, in the $O(3)$ case we will have that $G_{f(X)} \subseteq G_X$ if and only if all columns of $f(X)$ are in the linear space spanned by the points of X. Thus, the stable functions in the $O(3)$ case can be parameterized as

$$f_k(X) = \sum_j f_j^k(X) x_j. \tag{31}$$

Moreover, the space of functions of this form with *continuous* coefficients f_j^k is dense (over compact sets) in the space of continuous equivariant functions (follows from Proposition 4 in (Villar et al., 2021))

In the $SO(3)$ case, stable functions have a slightly more complex parameterization, in which each column of $f(X)$ is of the form

$$f(X)_k = \sum_j f_j^k(X) x_j + \sum_{i,j} f_{i,j}^k(X)(x_i \times x_j), \tag{32}$$

where $f_j^k(X)$ and $f_{i,j}^k(X)$ are arbitrary, scalar-valued functions. Moreover the space of functions of this form with *continuous* coefficients f_j^k and $f_{i,j}^k(X)$ are dense (over compact sets) in the space of continuous equivariant functions (follows from Proposition 5 in (Villar et al., 2021))

Continuous equivariant universal models for $O(3)$ can thus be obtained by applying $\mathcal{E}_{\text{weighted}}$ to functions of the form (31), where the f_j are taken from some dense function space Q. Continuous equivariant universal models for $SO(3)$ can be obtained analogously via (32).

E. Equivariant Projections via Stable Frames

To obtain BEC projection operators from robust frames in the equivariant setting, we add the requirement that, replacing μ_v with $\bar{\mu}_v$ in the definition of $\mathcal{E}_{\text{weighted}}$, will not affect the operator. This requirement is satisfied automatically in the invariant case, but in the equivariant case it is an additional condition we must enforce.

Definition E.1 (Stable robust frame)**.** Let (V, G) and (W, G) be modules. We say that a robust frame $\mu_{[\cdot]}$ is *stable* at a point v, if for every sequence $v_n \to v$ and for every $f : V \to W$

$$\int gf(g^{-1}v)d\mu_v(g) = \int gf(g^{-1}v)d\bar{\mu}_v(g)$$
$$\lim_{k\to\infty} \int gf(g^{-1}v)d\mu_{v_k}(g) = \int gf(g^{-1}v)d\mu_v(g)$$

We say that $\mu_{[\cdot]}$ is a stable robust frame if it is robust and stable at all points $v \in V$.

Remark E.2. Note that the second requirement above resembles the standard definition of weak convergence, but it requires only convergence of integrals $\int F(g)d\mu_{v_k} \to \int F(g)d\mu(g)$ for functions of the form $F(g) = gf(g^{-1}v)$. Not all functions on G are of this form. Note for example that every function of this form is G_v equivariant since for all $s \in G_v$ we have

$$F(sg) = sgf(g^{-1}s^{-1}v) = sgf(g^{-1}v) = sF(g)$$

In particular, to check whether a robust frame is stable, it is sufficient to check whether for every v, and every G_v equivariant $F : G \to W$, we have

$$\int F(g)d\mu_v(g) = \int F(g)d\bar{\mu}_v(g)$$
$$\lim_{k\to\infty} \int F(g)d\mu_{v_k}(g) = \int F(g)d\mu_v(g)$$

for every sequence v_k converging to v.

Proposition E.3. *Let (V, G) and (W, G) be modules, and μ a stable robust frame. Then $\mathcal{E} : F(V, W) \to F(V, W)$ is a BEC operator.*

Proof. Let $f : V \to W$ be a function. We can now show that $\mathcal{E}_{\text{weighted}}[f]$ is an equivariant function because for every $h \in G$ and $v \in V$,

$$\begin{aligned}
\mathcal{E}_{\text{weighted}}[f](hv) &= \int_G gf(g^{-1}hv)d\mu_{hv}(g) \\
&= \int_G gf(g^{-1}hv)d\bar{\mu}_{hv}(g) \\
&= \int_G gf(g^{-1}hv)dh_*\bar{\mu}_v(g) \\
&= \int_G hgf((hg)^{-1}hv)d\bar{\mu}_v(g) \\
&= h\left[\int_G gf(g^{-1}v)d\bar{\mu}_v(g)\right] \\
&= h\left[\int_G gf(g^{-1}v)d\mu_v(g)\right] \\
&= h\mathcal{E}_{\text{weighted}}[f](v).
\end{aligned}$$

If f is equivariant then $\mathcal{E}_{\text{weighted}}[f] = f$ since $gf(g^{-1}v) = f(v)$ from equivariance.

Boundedness also follows easily since $d\mu_v$ is a probability measure, so that on every compact $K \subseteq V$ which is also closed under the action of G, we have for every $v \in K$ that

$$|\mathcal{E}_{\text{weighted}}[f](v)| \le \left|\int_G gf(g^{-1}v)d\mu_v(g)\right| \le \max_{g\in G} \|g\| \max_{w\in K} |f(w)|, \tag{33}$$

where $\|g\|$ denotes the operator norm of the linear operator g (which is bounded, since G is compact and acting linearly and continuously).

Continuity Let $f : V \to \mathbb{R}$ be a continuous functions. Let v_k be a sequence converging to $v \in V$. We need to show that $\mathcal{E}_{\text{weighted}}[f](v_k)$ converges to $\mathcal{E}_{\text{weighted}}[f](v)$. We observe that

$$\begin{aligned}
&|\mathcal{E}_{\text{weighted}}[f](v_k) - \mathcal{E}_{\text{weighted}}[f](v)| \\
&= \left| \int gf(g^{-1}v_k) d\mu_{v_k}(g) - \int gf(g^{-1}v) d\mu_v(g) \right| \\
&\leq \left| \int gf(g^{-1}v_k) d\mu_{v_k}(g) - \int gf(g^{-1}v) d\mu_{v_k}(g) \right| + \left| \int gf(g^{-1}v) d\mu_{v_k}(g) - \int gf(g^{-1}v) d\bar{\mu}_v(g) \right|.
\end{aligned}$$

The second term tends to 0 from the definition of a stable frame. The first term tends to zero because $gf(g^{-1}v_k)$ converges to $gf(g^{-1}v)$ uniformly in g as k tends to infinity. □

We conclude this appendix with an explanation of how stable robust frames can be constructed for our $SO(2)$ and $SO(3)$ examples.

Example E.4. For the action of S^1 on $\mathbb{C}^n$ we double the size of the weighted frame $\mu^{SO(2)}$ and define

$$g_{i,+}(Z) = g_i(Z) \text{ and } g_{i,-}(Z) = -g_i(Z)$$

$$w_{i,+}(Z) = w_{i,-}(Z) = \frac{1}{2} w_i(Z).$$

We then obtain the projection operator

$$Q_2(f)(Z) = \sum_{i=1}^{n} \sum_{s \in \{-1,1\}} w_{i,s}(Z) g_{i,s}(Z) \cdot f(g_{i,s}^{-1}(Z) \cdot Z)$$

We need to check that this frame is a stable robust frame. At any non-zero point, since the stabilizer is trivial, the robustness of the original frame implies stable robustness since the notions are equivalent.

The main interesting point is at zero, where we have a non-trivial stabilizer. To show that the frame $\mu_{[\cdot]}$ we defined is stable at 0_n, it is sufficient to show, using Remark E.2, that for every $G_{0_n} = S^1$ equivariant function $F : S^1 \to \mathbb{C}^n$, we will have that

$$\int F(g) d\mu_0(g) = \int F(g) d\langle \mu_0 \rangle_0(g) = 0$$

and that for every $Z_k \to 0$ we will have that

$$\lim_{k \to \infty} \int F(g) d\mu_{Z_k}(g) = \int F(g) d\mu_0(g) = 0.$$

Indeed, using the equivariance of F we have

$$\int F(g) d\mu_0(g) = \frac{1}{2}(F(1) + F(-1)) = \frac{1}{2}(F(1) - F(1)) = 0,$$

and

$$\int F(g) d\mu_{Z_k}(g) = \frac{1}{2} \sum_{i=1}^{n} w_i(Z_k)(F(g_i(Z_k)) + F(-g_i(Z_k))) = \frac{1}{2} \sum_{i=1}^{n} w_i(Z_k)(F(g_i(Z_k)) - F(g_i(Z_k))) = 0.$$

Example E.5. We now define a stable weakly equivariant weighted frame for the module $(\mathbb{R}^{3\times n}, SO(3))$. This cardinality of the stable version $\mu^{SO(3)\text{stable}}$ of $\mu^{SO(3)}$ will be four times larger.

$$\mu^{SO(3)\text{stable}}(X) = \sum_{j \neq i} \sum_{k=1}^{4} \frac{1}{4} w_i(X) w_{ij}(X) \delta_{g_{ij} d^{(k)}(X)}$$

where $d^{(k)}, k = 1, 2, 3, 4$ are the four diagonal matrices in $SO(3)$, that is

$$d^{(1)} = I_3, d^{(2)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$
$$d^{(3)} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} d^{(4)} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

Let us explain why this frame is stable at points with non-trivial stabilizer. At $X = 0$ the reasoning is the same as in the $SO(2)$ example. If X has rank 1, then all points x_i with $x_i = 0$ have weight 0, and at all other points $x_i\ g_{ij}^{-1}$ is a rotation taking x_i to $\|x_i\|e_1$, which means that $g_{ij}d^{(k)}g_{ij}^{-1} \in G_X$ for $k = 1, 2$. Now, for any G_X equivariant function $F : SO(3) \to \mathbb{R}^{3\times n}$ and X with rank one, we obtain (denoting $g_{ij} = g_{ij}(X)$ and $\hat{w}_{ij} = w_i(X)w_{ij}(X)$ for simplicity)

$$\begin{aligned}
\int_G F(g)d\mu_X(g) &= \frac{1}{4}\sum_{i:x_i\neq 0}\sum_j \hat{w}_{ij}\sum_{k=1}^{4} F(g_{ij}d^{(k)}) \\
&= \frac{1}{4}\sum_{i:x_i\neq 0}\sum_j \hat{w}_{ij}(F(g_{ij}) + F(g_{ij}d^{(2)}g_{ij}^{-1}g_{ij}) \\
&+ F(g_{ij}d^{(2)}g_{ij}^{-1}g_{ij}d^{(4)}) + F(g_{ij}d^{(4)})) \\
&= \frac{1}{4}\sum_{i:x_i\neq 0}\sum_j \hat{w}_{ij}((d^{(1)} + g_{ij}d^{(2)}g_{ij}^{-1})F(g_{ij}) \\
&+ (d^{(1)} + g_{ij}d^{(2)}g_{ij}^{-1})F(g_{ij}d^{(4)}))
\end{aligned}$$

The stability then follows from the fact that for all $w \in W$,

$$\int_{G_X} swds = \frac{1}{2}g_{ij}(d^{(1)} + d^{(2)})g_{ij}^{-1}w,$$

so that the expression above is given by

$$\int_G F(g)d\mu_X(g) = \frac{1}{2}\int_{G_X} s\left(\sum_{i:x_i\neq 0}\sum_j \hat{w}_{ij}(F(g_{ij}) + F(g_{ij}d^{(4)}))\right)$$

and therefore is equal to

$$\int_G F(g)d\bar{\mu}_X(g) = \int_{G_X}\int_G F(sg)d\mu_X(g) = \int_{G_x} s\left(\int_G F(g)d\mu_X(g)\right).$$

F. Experiments

F.1. Empirically verifying discontinuities

We also include a practical demonstration of a discontinuity in a canonicalization code library, equiadapt (associated with Mondal et al. (2023) and Kaba et al. (2022)). We consider their point cloud implementation, which is $O(3)$-equivariant. They use a vector neuron architecture composed with the Gram-Schmidt algorithm (which we together call C) to learn an orthogonal matrix, and compose it with a downstream non-equivariant network (which we call f). We train their network on the ModelNet40 dataset, and demonstrate that the resultant pipeline is discontinuous at certain points. (Our theoretical results show that, for the groups we consider, no such canonicalization method can always preserve continuity; this demonstration shows that they do not preserve continuity in practice, for the particular f that is learned.) More specifically, we consider an input point cloud x, and a "bad" point cloud b generated such that every point points in the same direction (i.e. the matrix of points is rank 1, so this point cloud has infinite stabilizer in $O(3)$). We predict that the pipeline $f(C(\cdot))$ is discontinuous at the input point cloud b, and verify this experimentally as follows. First, note that $C(b)$ itself yields a "NaN" output.

To further verify the discontinuity, we compute the average pairwise normalized L_2 distance between points close to b, which are generated as convex combinations $\epsilon p + (1-\epsilon) * b$ for many randomly chosen point clouds p (generated such that they have trivial stabilizer with probability 1). For the sake of comparison, we also repeat this exact process, with a random asymmetric point cloud g replacing b. If $f \circ C$ is discontinuous at b, we expect that there is no valid limiting value at b, and therefore the average pairwise distance nearby b should be much larger than it is for g. As shown in Table 3, this is indeed what we observe. We find that both the canonicalization C and the composition $f \circ C$ show strong evidence of discontinuity.

Pairwise Error Metric	x_1, x_2 near a singularity b	x_1, x_2 near a generic point g
$\frac{\|\|C(x_1)-C(x_2)\|\|}{\|\|C(x_1)\|\|}$	1.1088	1.7035e-5
$\frac{\|\|f(C(x_1))-f(C(x_2))\|\|}{\|\|f(C(x_1))\|\|}$	0.0406	0.0009

Table 3. Average distance between pairs of points, near a singular point cloud and near a random point cloud.

It is also straightforward to understand the source of this behavior for the particular architecture used in equiadapt. In particular, the Gram-Schmidt procedure is very unstable when the vectors v_1 and v_2 it is provided are close to linearly dependent. And, since v_1 and v_2 are coming from an equivariant network applied to an input that is a perturbation away from having stabilizer isomorphic to $SO(2)$, they are nearly linearly dependent. This simple experiment verifies that discontinuity arising from canonicalization is not just a hypothetical, but a real problem for practically-used architectures after training.

G. Proofs

Proposition 2.2. *Let $\mathcal{E} : F(V,W) \to F(V,W)$ be a BEC operator, and $Q \subseteq C(V,W)$ a dense subset . Then $\mathcal{E}(Q) = \{\mathcal{E}(q) |\ q \in Q\}$ contains only continuous equivariant functions, and is dense in $C_{\text{equi}}(V,W)$.*

Proof. It is clear directly from the definition that $\mathcal{E}(Q)$ contains only continuous functions, so we only need to prove the density. Let $K \subset V$ be a compact subset and $\epsilon > 0$ be given. We need to show that for every $f \in C_{\text{equi}}(V,W)$, there exists a $q \in Q$ such that

$$\sup_{x\in K} |f(x) - \mathcal{E}(q)(x)| < \epsilon.$$

Since Q is dense in $C(V,W)$, there exists a $q \in Q$ (not necessarily equivariant), such that $\sup_{x\in K} |f(x) - q(x)| < \delta$ for any $\delta > 0$. Since $\mathcal{E}$ is bounded and $\mathcal{E}(f) = f$ (since f is equivariant), we get the following bound

$$\sup_{x\in K} |f(x) - \mathcal{E}(q)(x)| = \sup_{x\in K} |\mathcal{E}(f)(x) - \mathcal{E}(q)(x)| \leq \|\mathcal{E}\| \sup_{x\in K} |f(x) - q(x)| < \|\mathcal{E}\|\delta. \tag{34}$$

Choosing δ sufficiently small completes the proof. □

Proposition 2.5. *Let $y : V \to V$ be a canonicalization. Then $\mathcal{I}_{\text{can}} : F(V,\mathbb{R}) \to F(V,\mathbb{R})$ preserves continuity if and only if y is continuous.*

Proof. Let v_n be an arbitrary convergent sequence in V, i.e. $\lim_{n\to\infty} v_n = v$. If y is continuous, then $y_{v_n} \to y_v$, so that if f is continuous, we have

$$\lim_{n\to\infty} \mathcal{I}_{\text{can}}[f](v_n) = \lim_{n\to\infty} f(y_{v_n}) = f(y_v) = \mathcal{I}_{\text{can}}[f](v). \tag{35}$$

The proves that $\mathcal{I}_{\text{can}}$ preserves continuity if the canonicalization y is continuous.

For the reverse direction, suppose that y is not continuous. This means that there exists a convergent sequence v_n with $\lim_{n\to\infty} v_n = v$, but such that y_{v_n} does not converge to y_v. By choosing an appropriate subsequence, we may assume that there exists an $\epsilon > 0$ such that $|y_{v_n} - y_v| > \epsilon$ for all n. Let f be a continuous function such that $f(y_v) = 1$ and $f(w) = 0$ for all $w \in V$ satisfying $|w - y_v| > \epsilon$ (such a function exists by well-known considerations, for example the Urysohn Lemma). Then we will have

$$\lim_{n\to\infty} \mathcal{I}_{\text{can}}[f](v_n) = \lim_{n\to\infty} f(y_{v_n}) = 0 \neq 1 = f(y_v) = \mathcal{I}_{\text{can}}[f](v). \tag{36}$$

Hence the canonicalization of $\mathcal{I}_{\text{can}}[f]$ is not continuous and so $\mathcal{I}_{\text{can}}$ does not preserve continuity. □

We stated the following proposition in the main text:

Proposition 2.7. *Consider $O(d)$ and $SO(d)$ acting on $\mathbb{R}^{d\times n}$. If $n > d \geq 1$ for $O(d)$, or $n \geq d \geq 2$ for $SO(d)$, then there is no continuous canonicalization.*

We in fact prove the following two slightly more precise, separate results for $SO(d)$ and $O(d)$, which subsume the previous result. In the proofs, we will also abuse notation slightly and simply write y for both the canonicalization $V \to V$ and the induced map on the quotient space that we previously called $\tilde{y} : V/G \to V$.

Proposition G.1. *If $n \geq d \geq 2$, $SO(d)$ acting on $\mathbb{R}^{d\times n}$ does not have a continuous canonicalization.*

Proof. Observe first that by considering the subspace

$$Y = \{(x_1, ..., x_d, 0, ..., 0),\ x_i \in \mathbb{R}^d\} \subset \mathbb{R}^{d\times n} \tag{37}$$

a continuous canonicalization for $\mathbb{R}^{d\times n}$ gives a continuous canonicalization for $\mathbb{R}^{d\times d}$, so it suffices to consider the case $n = d$.

Consider first the case $d = 2$. Denote $G = SO(2)$. Consider the subspace $Y \subset \mathbb{R}^{2\times 2}$ defined by

$$Y = \{(x_1, x_2),\ |x_1|^2 + |x_2|^2 = 1\}. \tag{38}$$

It is clear that Y is G-invariant and that $Y \approx S^3$. Moreover, we claim that $Y/G \approx S^2$. Indeed, consider the G-equivariant map $Y \to S^2$ given by

$$H_f : (x_1, x_2) \to \begin{pmatrix} \sqrt{1-z^2}\cos(\theta) \\ \sqrt{1-z^2}\sin(\theta) \\ z \end{pmatrix} \in S^2, \tag{39}$$

where $\theta := \theta(x_1, x_2) \in [0, 2\pi)$ is the (counterclockwise) angle from x_1 to x_2 (when $x_1 = 0$ or $x_2 = 0$ this angle is not well-defined and we simply set $\theta(x_1, x_2) = 0$), and the height $z = z(x_1, x_2)$ is given by

$$z(x_1, x_2) = \begin{cases} \frac{2}{\pi}\arctan(\log(|x_2|) - \log(|x_1|)) & 0 < |x_1|, |x_2| < 1 \\ 1 & |x_1| = 0 \\ -1 & |x_2| = 0. \end{cases} \tag{40}$$

It is straightforward to verify that this map is both G-invariant and continuous. Indeed, both the angle θ and the lengths $|x_1|$ and $|x_2|$ are G-invariant. Since the map H_f only depends upon these functions of the input it is clearly G-invariant. Moreover, continuity is evident away from the points where $x_1 = 0$ or $x_2 = 0$, since in this regime the map H_f is a composition and product of continuous functions. Note that although θ is not a continuous function of $x_1, x_2 \neq 0$, both $\sin(\theta)$ and $\cos(\theta)$ are continuous. Indeed, if we view x_1 and x_2 as elements of the complex plane, then

$$\cos(\theta) + i\sin(\theta) = \frac{x_2|x_1|}{x_1|x_2|},$$

which verifies continuity away from the set where $x_1 = 0$ or $x_2 = 0$.

Next, we verify continuity when $x_1 = 0$ (the corresponding calculation when $x_2 = 0$ is completely analogous). We observe that for any unit vector x_2, we have

$$H_f(0, x_2) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Given any sequence $(x_1^n, x_2^n) \to (0, x_2)$ we note that since $x_1^n \to 0$, we have that $z(x_1^n, x_2^n) \to (2/\pi)\arctan(\infty) = 1$. Plugging this into (39), we get that

$$H_f(x_1^n, x_2^n) \to \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

verifying continuity at $(0, x_2)$. Thus, H_f does indeed define a continuous map $Y/G \to S^2$.

Finally, we must verify the surjectivity and injectivity of H_f. These follow since given any $z \in [-1, 1]$ we can solve for z in (40) uniquely for the two lengths $|x_1|$ and $|x_2|$ satisfying $|x_1|^2 + |x_2|^2 = 1$ (since arctan and log are both increasing functions). If $z \neq \pm 1$, these lengths will both be non-zero and the angle θ is well-defined and uniquely determined up to a rotation of both x_1 and x_2. At the poles where $z = \pm 1$ (and θ becomes irrelevant), one of the vectors $x_i = 0$ and the other is a unit vector, and all of these configurations are equivalent up to a rotation as well. In fact, the map H_f is easily seen to be the well-known Hopf fibration (Hopf, 1931).

We can now complete the proof that no continuous canonicalization can exist when $d = 2$. Indeed, if there were such a canonicalization y, then the induced map $\tilde{y}$ and the quotient map q would satisfy

$$S^2 \xrightarrow{\tilde{y}} S^3 \xrightarrow{q} S^2 \tag{41}$$

whose composition is the identity. Algebraic topology provides a variety of obstructions to such a scenario. Perhaps the simplest comes from the homology groups (see (Hatcher, 2002), Chapter 2). Indeed, the homology groups are $H_2(S^2) = \mathbb{Z}$ and $H_2(S^3) = \{0\}$. The induced maps on homology groups would have to satisfy

$$\mathbb{Z} \xrightarrow{\tilde{y}^*} \{0\} \xrightarrow{q^*} \mathbb{Z} \tag{42}$$

with composition equal to the identity which is clearly impossible.

Next, we consider the case where $d > 2$. We will prove by induction on d that no canonicalization exists for $(\mathbb{R}^{d\times d}, SO(d))$. The case $d = 2$ forms the base case.

Suppose that there exists a continuous canonicalization y for $(\mathbb{R}^{d\times d}, SO(d))$. Consider the element

$$X_0 = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{d\times d}$$

whose first $d-1$ columns are 0 and whose remaining column consists of the basis vector e_d. Note that we can compose y with any element of $SO(d)$ to obtain a new continuous canonicalization. Thus, we can assume without loss of generality that $y_{X_0} = X_0$ by composing with a rotation which moves the last (and only non-zero) column of y_{X_0} to the standard basis vector e_d.

We will use y to construct a continuous canonicalization $\hat{y}$ for $(\mathbb{R}^{(d-1)\times(d-1)}, SO(d-1))$, which cannot exist by the inductive hypothesis. Consider the norm on $\mathbb{R}^{d\times d}$ defined by (note that this is not the usual ℓ^∞ norm of a vector)

$$\|X\|_\infty := \max_i \|x_i\|_2, \tag{43}$$

i.e. the maximum length of the columns of X. Let $\epsilon > 0$ be chosen so that $\|y_X - y_{X_0}\|_\infty = \|y_X - X_0\|_\infty < 1/2$ whenever $\|X - X_0\|_\infty \leq \epsilon$ (this can always be done by the continuity of the canonicalization y). Let

$$B^{d-1}_\infty = \{X \in \mathbb{R}^{(d-1)\times(d-1)} : \|X\|_\infty = 1\} \tag{44}$$

denote the unit ball in $\mathbb{R}^{(d-1)\times(d-1)}$ with respect to the norm (43). We will first construct the canonicalization $\hat{y}$ on the set B^{d-1}_∞ and then extend it homogeneously to all of $\mathbb{R}^{(d-1)\times(d-1)}$.

We define a (continuous) map $i_0 : B^{d-1}_\infty \to \mathbb{R}^{d\times d}$ by

$$i_0(X) = \begin{pmatrix} \epsilon X & 0 \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{d\times d}. \tag{45}$$

In otherwords, i_0 simply puts ϵ times X into the upper $(d-1)\times(d-1)$ block of X_0. We clearly have

$$\|i_0(X) - X_0\|_\infty \leq \epsilon$$

for every $X \in B^{d-1}_\infty$.

Applying the canonicalization y to $i_0(X)$ gives a matrix

$$y_{i_0(X)} = U(X)\begin{pmatrix} \epsilon X & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} A & v \\ w^T & r \end{pmatrix}, \tag{46}$$

for a matrix $U(X) \in SO(d)$ (since y is a canonicalization and thus maps to the same orbit). Here the submatrices $A \in \mathbb{R}^{(d-1)\times(d-1)}$, $w \in \mathbb{R}^{d-1}$, $v \in \mathbb{R}^{d-1}$, and $r \in \mathbb{R}$ are all continuous functions of X since y and i_0 are continuous (we have suppressed this dependence for notational simplicity). Finally, we observe that since the last columns of $i_0(X)$ is e_d, which has norm 1, we have $|v|^2 + r^2 = 1$.

Since $\|i_0(X) - X_0\|_\infty \leq \epsilon$ we have that $\|y_{i_0(X)} - X_0\|_\infty < 1/2$, which means that $r \geq 1/2$. We define the matrix

$$V(X) := I_d + (r-1)\left[e_d e_d^T + \frac{\bar{v}\bar{v}^T}{1-r^2}\right] + \left[e_d\bar{v}^T - \bar{v}e_d^T\right], \tag{47}$$

where I_d is the $d\times d$ identity matrix and $\bar{v}$ is v augmented with a 0 in the d-th coordinate. This is clearly a continuous function of both $r > -1$ and v since $(r-1)/(1-r^2) = -1/(r+1)$, and thus it is also a continuous function of X. Moreover, we claim that since $r^2 + |v|^2 = 1$, the matrix $V(X) \in SO(d)$ and

$$V(X)\begin{pmatrix} v \\ r \end{pmatrix} = e_d. \tag{48}$$

Indeed, in the case $v = 0$, $r = 1$ we clearly have $V(X) = I_d$. If $v \neq 0$, then the space orthogonal to the vectors e_d and $\bar{v}$, which we denote by $W^{\perp}$, is clearly invariant under $V(X)$, while we calculate

$$V(X)e_d = e_d + (r-1)e_d - \bar{v} = re_d - \bar{v} \tag{49}$$

and

$$V(X)\bar{v} = \bar{v} + (r-1)\frac{|v^2|}{1-r^2}\bar{v} + |v|^2 e_d = r\bar{v} + |v|^2 e_d, \tag{50}$$

since $|v|^2 = 1 - r^2$. Thus, with respect to the orthonormal basis consisting of $e_d, \bar{v}/|v|$ and an orthonormal basis for $W^{\perp}$, the matrix $V(X)$ consists of a $(d-2)\times(d-2)$ identity matrix block (corresponding to the orthonormal basis of $W^{\perp}$), and a 2×2 block of the form

$$\begin{pmatrix} r & -|v| \\ |v| & r \end{pmatrix} \tag{51}$$

corresponding to e_d and $\bar{v}/|v|$. This implies that $V(X) \in SO(d)$. We also calculate

$$V(X)\begin{pmatrix} v \\ r \end{pmatrix} = rV(X)e_d + V(x)\bar{v} = r^2 e_d - r\bar{v} + r\bar{v} + |v|^2 e_d = (r^2 + |v|^2)e_d = e_d \tag{52}$$

as desired. Geometrically, $V(X)$ is the rotation through the plane spanned by e_d and $\bar{v}$ which maps the last column of $y_{i_0(X)}$ to e_d. Algebraically, this is given by the formula (47).

We now define the canonicalization $\hat{y} : B_{\infty}^{d-1} \to \mathbb{R}^{(d-1)\times(d-1)}$ via

$$V(X)y_{i_0(X)} = \begin{pmatrix} \epsilon\hat{y}_X & 0 \\ 0 & 1 \end{pmatrix}. \tag{53}$$

Since both $V(r, v)$ and $y_{i_0(X)}$ are continuous functions of X, it is clear that $\hat{y}$ is continuous.

We claim that $\hat{y}$ is a canonicalization for $(\mathbb{R}^{(d-1)\times(d-1)}, SO(d-1))$. Note that since y is $SO(d)$ invariant, we have that $\hat{y}$ is $SO(d-1)$ invariant (recall that $V(X)$ only depends upon $y_{i_0(X)}$ and is thus also invariant). Further, since

$$\begin{pmatrix} \epsilon\hat{y}_X & 0 \\ 0 & 1 \end{pmatrix} = V(X)U(X)\begin{pmatrix} \epsilon X & 0 \\ 0 & 1 \end{pmatrix}, \tag{54}$$

where by construction the last column of $V(X)U(X) \in SO(d)$ is e_d, we have that $\hat{y}_X = \hat{O}(X)X$ for $\hat{O}(X) \in SO(d-1)$ being the upper right $(d-1)\times(d-1)$-block of $V(X)U(X)$. This implies that $\hat{y}$ is a canonicalization as desired.

We finally extend $\hat{y}$ 1-homogeneously to all of $\mathbb{R}^{(d-1)\times(d-1)}$ via

$$\hat{y}_X = \begin{cases} \|X\|_{\infty}\tilde{y}_{X/\|X\|_{\infty}} & X \neq 0 \\ 0 & X = 0. \end{cases}$$

This gives a continuous canonicalization for $(\mathbb{R}^{(d-1)\times(d-1)}, SO(d-1))$, which is impossible by the inductive hypothesis. □

Proposition G.2. *If $n > d \geq 1$, then $O(d)$ acting on $\mathbb{R}^{d\times n}$ does not have a continuous canonicalization.*

Proof. Observe that be considering the subspace

$$Y = \{(x_1, ..., x_{d+1}, 0, ..., 0),\ x_i \in \mathbb{R}^d\} \subset \mathbb{R}^{d\times n} \tag{55}$$

a continuous canonicalization for $\mathbb{R}^{d\times n}$ gives a continuous canonicalization for $\mathbb{R}^{d\times(d+1)}$, so it suffices to consider the case $n = d + 1$.

We first address the case $d = 1$, in which case $O(d) = \{\pm 1\}$. Consider the subspace

$$Y = S^1 = \{(x_1, x_2) :\ |x_1|^2 + |x_2|^2 = 1\} \subset \mathbb{R}^{1\times 2}.$$

The quotient $Y/G \approx S^1$ (identifying antipodal points on a circle again gives a circle) and the quotient map $q : S^1 \to S^1$ is given by identifying antipodal points which corresponds to the map $t \to 2t$ (here we are viewing S^1 as $\mathbb{R}/\mathbb{Z}$). Suppose that y is a continuous canonicalization and $\tilde{y} : Y/G \to Y$ is the corresponding map on the quotient space. We consider the induced maps on the fundamental group $\pi_1(S^1) = \mathbb{Z}$ corresponding to the quotient map q and map $\tilde{y}$. These would satisfy $q_*(z) = 2z$ and $q^* \circ \tilde{y}^* = I$. This is clearly impossible since q_* is not surjective.

We now prove the result for general d by induction on d with $d = 1$ being the base case. In particular, a continuous canonicalization for $(\mathbb{R}^{d\times(d+1)}, O(d))$ can be used to obtain a continuous canonicalization for $(\mathbb{R}^{(d-1)\times d}, O(d-1))$ utilizing exactly the same construction as in the proof of Proposition G.1. ☐

Theorem 2.8. *Let G be a finite group acting continuously on a metric space V, let V_{free} denote the points in V with trivial stabilizer, and let $\mathcal{F}$ be a frame which preserves continuity on V_{free}. If V_{free} is connected, then $\mathcal{F}(v) = G$ for all v in the closure of V_{free}.*

Proof. We begin by showing that the identity element $e \in G$ is in $\mathcal{F}(v_0)$ for some $v_0 \in V_{free}$. Indeed, choose some $v \in V_{free}$, and choose some $g \in \mathcal{F}(v)$. Then we can set $v_0 = g^{-1}v$. This point is in V_{free} and the frame equivariance implies that

$$e = g^{-1}g \in g^{-1}\mathcal{F}(v) = \mathcal{F}(g^{-1}v)$$

Next, we will show that $e \in \mathcal{F}(v)$ *for all* $v \in V_{free}$. Since V_{free} is connected, it is sufficient to show that the sets

$$V_0 = \{v \in V_{free} | \quad e \in \mathcal{F}(v)\} \text{ and } V_1 = \{v \in V_{free} | \quad e \notin \mathcal{F}(v)\}$$

are both open in V_{free}. This implies that one of these sets must be empty and the other all of V_{free}. Since we say V_0 is not empty we will get that it is all of V_{free}.

We now show that V_0 and V_1 are open in V_{free}. Note that the continuity of the group action and the finiteness of G implies that V_{free} is open, so that we need to prove that V_0 and V_1 are open in the original topology of V.

Fix some $v \in V_{free}$. Since v has a trivial stabilizer we have that $gv \neq v$ for all distinct $g \in G$. By continuity of the action and the finiteness of the group there exists an open neighborhood U of v so that $g^{-1}U \cap U = \emptyset$ for all distinct $g \in G$. Thus, for any fixed $v \in V_{free}$ we can define a continuous function f_v which is identically 1 on U and zero on $g^{-1}U$ for all $g \neq e$.

It follows that for all $y \in U$ we have that $f_v(g^{-1}y)$ is one if $g = e$ and is zero otherwise. In particular

$$\mathcal{I}_{\text{frame}}[f_v](y) = \frac{1}{|\mathcal{F}(y)|} \sum_{g \in \mathcal{F}(y)} f_v(g^{-1}y) = \begin{cases} 1/|\mathcal{F}(y)| & \text{if } e \in \mathcal{F}(y) \\ 0 & \text{if } e \notin \mathcal{F}(y) \end{cases}, \quad \forall y \in U.$$

It follows that if $v \in V_0$, so that $e \in \mathcal{F}(v)$, then by continuity of $\mathcal{I}_{\text{frame}}[f_v]$ we will have that $e \in \mathcal{F}(y)$ (and also $|\mathcal{F}(y)| = |\mathcal{F}(v)|$) for all $y \in U$. This shows that V_0 is open. Similarly, if $v \in V_1$, so that $e \notin \mathcal{F}(v)$, then by continuity of $\mathcal{I}_{\text{frame}}[f_v]$ we will have that $e \notin \mathcal{F}(y)$ for all $y \in U$ so that V_1 is open. We have thus showed that $V_0 = V_{free}$.

We have shown that the identity element is in $\mathcal{F}(v)$ for all $v \in V_{free}$. The same is true for all group elements: let $g \in G$ and $v \in V_{free}$. We want to show that $g \in \mathcal{F}(v)$. Note that $g^{-1}v$ also is in V_{free}. Therefore $e \in \mathcal{F}(g^{-1}v)$, and by equivariance

$$g \in g\mathcal{F}(g^{-1}v) = \mathcal{F}(gg^{-1}v) = \mathcal{F}(v).$$

We now proved the claim for all points in V_{free} and we need to extend the claim to the closure of V_{free}.

Let v be a point in the closure of V_{free}, and let $g \in G$. We need to show again that $g \in \mathcal{F}(v)$. Note that $h^{-1}v = g^{-1}v$ if and only if $g = sh$ for some s in the stabilizer G_v. In this case we will say that g, h are G_v equivalent and denote $g \sim_v h$.

Similarly to earlier in the proof, we can find an open set $U \subseteq V$, and a continuous function $f_{v,g} : V \to \mathbb{R}$, such that for all $y \in U$ and $h \in G$, we have that $f_{v,g}(hy) = 1$ if $h \sim_v g$, and otherwise $f_{v,g}(hy) = 0$.

It follows that the continuous function $\mathcal{I}_{\text{frame}}[f_{v,g}](y)$ is equal to

$$\mathcal{I}_{\text{frame}}[f_{v,g}](y) = \frac{1}{|\mathcal{F}(y)|} \sum_{g \in \mathcal{F}(y)} f_v(g^{-1}y) = \frac{|G_v \cdot g \cap \mathcal{F}(y)|}{|\mathcal{F}(y)|}, \quad \forall y \in U.$$

Since v is in the closure of V_{free}, there is a sequence v_n of elements in $V_{free} \cap U$ converging to v, and since $\mathcal{F}(v_n) = G$ we obtain

$$\frac{|G_v|}{|G|} = \lim_{n\to\infty} \mathcal{I}_{\text{frame}}[f_{v,g}](v_n) = \mathcal{I}_{\text{frame}}[f_{v,g}](v) = \frac{|G_v \cdot g \cap \mathcal{F}(v)|}{|\mathcal{F}(v)|}$$

so that the intersection $G_v \cdot g \cap \mathcal{F}(v)$ is not empty: there exists some $s \in G_v$ such that $sg \in \mathcal{F}(v)$. By equivariance of the frame

$$g = s^{-1}sg \in s^{-1}\mathcal{F}(v) = \mathcal{F}(s^{-1}v) = \mathcal{F}(v).$$

This concludes the proof of the theorem. □

Corollary 2.9. *Let $d, n > 1$, and consider S_n acting on $\mathbb{R}^{d\times n}$. If $\mathcal{F}$ is a continuity-preserving frame, then $\mathcal{F}(X) = S_n$ for all $X \in \mathbb{R}^{d\times n}$.*

Proof. The matrices $X \in \mathbb{R}^{d\times n}$ which have trivial stabilizers are the set $\mathbb{R}^{d\times n}_{\text{distinct}}$ of matrices X whose columns are pairwise distinct. Clearly $\mathbb{R}^{d\times n}_{\text{distinct}}$ is dense in $\mathbb{R}^{d\times n}$, and therefore, due to Theorem 2.8, it is sufficient to show that $\mathbb{R}^{d\times n}_{\text{distinct}}$ is connected.

We will show that $\mathbb{R}^{d\times n}_{\text{distinct}}$ is path connected and hence connected. Let $X^{(0)}$ be any point with a trivial stabilizer. We will show that it can be connected by two straight lines to the point $X^{(2)}$ defined by $X^{(2)}_{ij} = j$, that is the matrix (clearly also with trivial stabilizer) whose rows are all identical and given by

$$\begin{pmatrix} 1 & 2 & \dots & n \end{pmatrix} \tag{56}$$

As a first step, we find a permutation τ which sorts the first row of $X^{(0)}$, so that

$$X^{(0)}_{1\tau(1)} \leq X^{(0)}_{1\tau(2)} \leq \ldots \leq X^{(0)}_{1\tau(n)}.$$

We then choose a point $X^{(1)}$ whose first row satisfies this same inequality strictly

$$X^{(1)}_{1\tau(1)} < X^{(1)}_{1\tau(2)} < \ldots < X^{(1)}_{1\tau(n)}.$$

and whose remaining rows are equal to the rows (56) of $X^{(2)}$. We note that all points in the straight line between $X^{(0)}$ and $X^{(1)}$ have a trivial stabilizer. For $X^{(0)}$ this is by assumption, and for all other points this is because the first row, sorted by τ, is strictly separated.

Next, we observe that all points in the straight line between $X^{(1)}$ and $X^{(2)}$ also have a trivial stabilizer. This is because all but the first row are equal to (56). Thus we have shown that $\mathbb{R}^{d\times n}_{\text{distinct}}$ is connected.

□

Theorem 2.10. *Consider $SO(2)$ acting on $\mathbb{R}^{2\times n}$ with $n \geq 2$. If $\mathcal{F}$ is a continuity-preserving frame, then*

$$\sup_{X\in\mathbb{R}^{d\times n}\setminus\mathbf{0}} |\mathcal{F}(X)| = \infty,$$

i.e. there does not exist a finite (unweighted) frame which preseves continuity.

Proof. By considering the subspace

$$Y = \{(x_1, x_2, 0, ..., 0) : \ x_1, x_2 \in \mathbb{R}^2\} \tag{57}$$

it suffices to consider the case $n = 2$.

We will in fact prove something a bit stronger. Let G be a group acting on a vector space V and define an unweighted frame of size N to be a map

$$\mathcal{F}_N : V \to G^N/S_N \tag{58}$$

from V to the set G^N/S_N of unordered N-tuples of elements of G (potentially with repetition). The corresponding invariant projection operator is given by

$$\mathcal{I}_{\text{can}}[f](v) := \frac{1}{N}\sum_{g\in\mathcal{F}_N(v)} f(g^{-1}v). \tag{59}$$

Given a frame $\mathcal{F}$ such that $|\mathcal{F}(v)| \leq M$ for all $v \in V$, we can construct a finite unweighted frame of size $M!$ by repeating each element of $\mathcal{F}(v)$ k_v times where $k_v = M!/|\mathcal{F}(v)|$. However, the notion of an unweighted frame of size N is more general since it allows different weights through repetition, although the weights must all be divisible by $1/N$.

We proceed to show that a continuity-preserving finite unweighted frame $\mathcal{F}_N$ cannot exist for $SO(2) = S^1$ acting on $\mathbb{R}^{2\times 2}$ for any finite N (here S^1 denotes the unit circle of rotations). Suppose to the contrary that $\mathcal{F}_N$ is such a frame. Consider the subspace Y defined in (38) in the proof of Theorem 2.7, and the subspace

$$Y^o := \{(x_1, x_2),\ |x_1|^2 + |x_2|^2 = 1,\ x_1 \neq 0,\ x_2 \neq 0\} \tag{60}$$

which is the same as Y but with the points where $x_1 = 0$ and $x_2 = 0$ removed. Since all points of Y have trivial stabilizer, the map $\mathcal{F}_N$ restricted to Y must be continuous and equivariant under the action of S^1.

Next, consider the space $S^1_{N,u} := (S^1)^N/S_N$ of unordered N-tuples of rotations. Observe that the group S^1 of rotations naturally acts componentwise on $S^1_{N,u}$ via the map

$$g \cdot (g_1, ..., g_N) = (gg_1, ..., gg_N). \tag{61}$$

In addition, since S^1 is abelian, there is a natural continuous map $S^1_{N,u} \to S^1$ given by multiplying all of the rotations together, i.e.

$$(g_1, ..., g_N) \to g_1 g_2 \cdots g_N \in S^1. \tag{62}$$

Consider the following map defined on Y^o

$$\arg_1(X) = \frac{x_1}{|x_1|} \in S^1, \tag{63}$$

which gives the angle of the first vector (well-defined and continuous since we have removed the points where $x_1 = 0$). We now define the continuous map $\tilde{\mathcal{F}}_N : Y^o \to S^1_{N,u}$ by

$$\tilde{\mathcal{F}}_N(X) = \arg_1(X)^{-1} \cdot \mathcal{F}_N(X) \in S^1_{N,u}, \tag{64}$$

which gives the unordered collection of angles that the first vector is rotated to under the frame $\mathcal{F}_N$ (here the multiplication in equation (64) is the action described in (61)). The map $\tilde{\mathcal{F}}_N$ is invariant under the action of S^1 and continuous on Y^o (but does not have a continuous extension to Y). Thus, it induces a map $\tilde{\mathcal{F}}_N : Y^o/S^1 \to S^1_{N,u}$.

Observe that $Y^o/S^1 \cong (0,1) \times S^1$ via the parameterization map

$$(t, \theta) \to \left(\begin{pmatrix}\sqrt{1-t^2}\\ 0\end{pmatrix}, \begin{pmatrix}t\cos(\theta)\\ t\sin(\theta)\end{pmatrix}\right), \tag{65}$$

where t denotes the length of x_2 and θ denotes the counterclockwise angle from x_1 to x_2. Using this parameterization, we define a map $\mathcal{G}_N : (0,1) \times S^1 \to S^1_{N,u}$ via

$$\mathcal{G}_N(t,\theta) = \tilde{F}_N\left(\begin{pmatrix}\sqrt{1-t^2}\\ 0\end{pmatrix}, \begin{pmatrix}t\cos(\theta)\\ t\sin(\theta)\end{pmatrix}\right). \tag{66}$$

Next, we claim that the map $\mathcal{G}_N$ extends to a continuous map $[0,1] \times S^1 \to S^1_{N,u}$. Indeed, we define

$$\mathcal{G}_N(0,\theta) = \mathcal{F}_N\left(\begin{pmatrix}1\\ 0\end{pmatrix}, \begin{pmatrix}0\\ 0\end{pmatrix}\right),\ \mathcal{G}_N(1,\theta) = \theta \cdot \mathcal{F}_N\left(\begin{pmatrix}0\\ 0\end{pmatrix}, \begin{pmatrix}1\\ 0\end{pmatrix}\right), \tag{67}$$

and check that the resulting map is continuous on $[0,1] \times S^1$. Clearly, continuity must only be checked at points of the form $(0,\theta)$ and $(1,\theta)$.

Suppose first that $(t_n, \theta_n) \to (0, \theta)$. Continuity in θ is clear when $t = 0$, so we may assume that $0 < t_n < 1$. We then see that

$$\mathcal{G}_N(t_n, \theta_n) = \tilde{\mathcal{F}}_N\left(\begin{pmatrix}\sqrt{1-t_n^2}\\ 0\end{pmatrix}, \begin{pmatrix}t_n\cos(\theta_n)\\ t_n\sin(\theta_n)\end{pmatrix}\right) = \mathcal{F}_N\left(\begin{pmatrix}\sqrt{1-t_n^2}\\ 0\end{pmatrix}, \begin{pmatrix}t_n\cos(\theta_n)\\ t_n\sin(\theta_n)\end{pmatrix}\right) \to \mathcal{F}_N\left(\begin{pmatrix}1\\ 0\end{pmatrix}, \begin{pmatrix}0\\ 0\end{pmatrix}\right). \quad (68)$$

Here we have used the continuity of $\mathcal{F}_N$ on all of Y to get the final convergence, and that

$$\arg_1\begin{pmatrix}\sqrt{1-t_n^2}\\ 0\end{pmatrix} = e_1$$

corresponds to $\theta = 0$ to get the middle equality.

Next, suppose that $(t_n, \theta_n) \to (1, \theta)$. We may again assume that $0 < t_n < 1$ since $\mathcal{G}_N$ is continuous in θ when $t = 1$. We similarly calculate

$$\mathcal{G}_N(t_n, \theta_n)\mathcal{F}_N\left(\begin{pmatrix}\sqrt{1-t_n^2}\\ 0\end{pmatrix}, \begin{pmatrix}t_n\cos(\theta_n)\\ t_n\sin(\theta_n)\end{pmatrix}\right) \to \mathcal{F}_N\left(\begin{pmatrix}0\\ 0\end{pmatrix}, \begin{pmatrix}\cos(\theta)\\ \sin(\theta)\end{pmatrix}\right). \quad (69)$$

Now the equivariance of $\mathcal{F}_N$ implies that

$$\mathcal{F}_N\left(\begin{pmatrix}0\\ 0\end{pmatrix}, \begin{pmatrix}\cos(\theta)\\ \sin(\theta)\end{pmatrix}\right) = \theta \cdot \mathcal{F}_N\left(\begin{pmatrix}0\\ 0\end{pmatrix}, \begin{pmatrix}1\\ 0\end{pmatrix}\right) = \mathcal{G}_N(1, \theta). \quad (70)$$

This verifies the continuity of $\mathcal{G}_N$ on all of $[0, 1] \times S^1$. Thus $\mathcal{G}_N$ defines a homotopy between the constant loop $\mathcal{G}(0, \cdot)$ and the loop $\mathcal{G}(1, \cdot)$ defined in (67) in the space $S^1_{N,u}$.

We complete the proof by showing that these loops cannot be homotopic. For this, we use the multiplication map defined in (62). Composing $\mathcal{G}_N$ with this map gives a continuous map $\mathcal{H}_N : [0, 1] \times S^1 \to S^1$. From (67) we see that $\mathcal{H}_N(0, \cdot)$ is a constant map, while $\mathcal{H}_N(1, \cdot) : S^1 \to S^1$ loops around the circle N times. Thus $\mathcal{H}_N$ gives a homotopy between these two loops. This is impossible, since the fundamental group $\pi_1(S^1) \cong \mathbb{Z}$ and $\mathcal{H}_N(0, \cdot)$ represents the zero element while $\mathcal{H}_N(1, \cdot)$ represents the element N (see (Hatcher, 2002), Chapter 1). □

Proposition 3.3. (Invariant frame averaging) Let $\mu_{[\cdot]}$ be a weighted frame. If $\mu_{[\cdot]}$ is *weakly equivariant*, then $\mathcal{I}_{\text{weighted}}$ is a bounded, invariant projection operator.

Proof. Let $f : V \to \mathbb{R}$ be a function. We need to show that $\mathcal{I}_{\text{weighted}}[f]$ is an invariant function. We first show that in general μ_v can be replace with $\bar{\mu}_v$ in the defininition of $\mathcal{I}_{\text{weighted}}[f]$. That is

$$\begin{aligned}
\mathcal{I}_{\text{weighted}}[f](v) &= \int_G f(g^{-1}v)d\mu_v(g) \\
&= \int_{G_v}\int_G f((sg)^{-1}v)d\mu_v(g)ds \\
&= \int_{G_v}\int_G f(g^{-1}v)ds_*\mu_v(g)ds \\
&= \int_G f(g^{-1}v)d\bar{\mu}_v(g) \quad (71)
\end{aligned}$$

Using this we can now show that $\mathcal{I}_{\text{weighted}}[f]$ is an invariant function because for every $h \in G$ and $v \in V$,

$$\begin{aligned}\mathcal{I}_{\text{weighted}}[f](hv) &= \int_G f(g^{-1}hv)d\mu_{hv}(g) \\ &= \int_G f(g^{-1}hv)d\bar{\mu}_{hv}(g) \\ &= \int_G f(g^{-1}hv)dh_*\bar{\mu}_v(g) \\ &= \int_G f((hg)^{-1}hv)d\bar{\mu}_v(g) \\ &= \int_G f(g^{-1}v)d\bar{\mu}_v(g) \\ &= \int_G f(g^{-1}v)d\mu_v(g) \\ &= \mathcal{I}_{\text{weighted}}[f](v).\end{aligned}$$

Boundedness also follows easily since $d\mu_v$ is a probability measure, so that

$$|\mathcal{I}_{\text{weighted}}[f](v)| \leq \left|\int_G f(g^{-1}v)d\mu_v(g)\right| \leq \sup_w |f(w)|. \tag{72}$$

□

Proposition 3.6. *$\mathcal{I}_{\text{weighted}}[f]$ is a BEC operator iff μ is a robust frame.*

Proof. Let $f : V \to \mathbb{R}$ be a continuous functions. Let v_k be a sequence converging to $v \in V$. For the 'if' direction, we need to show that $\mathcal{I}_{\text{weighted}}[f](v_k)$ converges to $\mathcal{I}_{\text{weighted}}[f](v)$. We observe that

$$\begin{aligned}&|\mathcal{I}_{\text{weighted}}[f](v_k) - \mathcal{I}_{\text{weighted}}[f](v)| \\ &= \left|\int f(g^{-1}v_k)d\mu_{v_k}(g) - \int f(g^{-1}v)d\mu_v(g)\right| \\ &\leq \left|\int f(g^{-1}v_k)d\mu_{v_k}(g) - \int f(g^{-1}v_k)d\langle\mu_{v_k}\rangle_v(g)\right| + \left|\int f(g^{-1}v_k)d\langle\mu_{v_k}\rangle_v(g) - \int f(g^{-1}v)d\bar{\mu}_v(g)\right|.\end{aligned}$$

Now, by assumption $\langle\mu_{v_k}\rangle_v \to \bar{\mu}_v$ weakly so that the second term tends to 0 (since f and the action of G are continuous). Moreover, since v is by definition invariant under the action of G_v, we have

$$\int f(g^{-1}v)d\mu_{v_k}(g) - \int f(g^{-1}v)d\langle\mu_{v_k}\rangle_v(g) = 0. \tag{73}$$

Hence, we get

$$\begin{aligned}&\left|\int f(g^{-1}v_k)d\mu_{v_k}(g) - \int f(g^{-1}v_k)d\langle\mu_{v_k}\rangle_v(g)\right| \\ &\leq \int |f(g^{-1}v_k) - f(g^{-1}v)|d\mu_{v_k}(g) + \int |f(g^{-1}v_k) - f(g^{-1}v)|d\langle\mu_{v_k}\rangle_v(g)\end{aligned}$$

Since $v_k \to v$, the group G is compact, f is continuous, and the integrals are both against probability measures, we finally get

$$\left|\int f(g^{-1}v_k)d\mu_{v_k}(g) - \int f(g^{-1}v_k)d\langle\mu_{v_k}\rangle_v(g)\right| \to 0, \tag{74}$$

as desired.

For the converse, assume that μ is not robust. This means that there exists a sequence v_k converging to v such that $\langle\mu_{v_k}\rangle_v$ does not converge to $\bar{\mu}_v$ weakly. Thus there is a continuous function ϕ defined on the group G, such that

$$\int_G \phi(g)d\langle\mu_{v_k}\rangle_v(g) \nrightarrow \int_G \phi(g)\bar{\mu}_v(g). \tag{75}$$

Further, we observe that by definition of the averaged measure $\langle\cdot\rangle_v$, for any measure μ we have

$$\int_G \phi(g)d\langle\mu\rangle_v(g) = \int_{G_v}\int_G \phi(g)d(s^*\mu)(g)ds = \int_{G_v}\int_G \phi(sg)d\mu(g)ds = \int_G \bar{\phi}(g)d\mu(g), \tag{76}$$

where the averaged function $\bar{\phi}$ is defined by

$$\bar{\phi}(g) := \int_{G_v} \phi(sg)ds \tag{77}$$

and is evidently G_v invariant. Thus $\bar{\phi}$ defines a function on G/G_v and (75) gives

$$\lim_{k\to\infty} \int_{G/G_v} \bar{\phi}(g)d\langle\mu_{v_k}\rangle_v(g) \neq \int_{G/G_v} \bar{\phi}(g)\bar{\mu}_v(g), \tag{78}$$

where $\langle\mu_{v_k}\rangle_v$ and $\bar{\mu}_v$ are viewed as measures on G/G_v.

Observe that since the action of G is continuous, the orbit $Gv = \{gv :\ g \in G\}$ is homeomorphic to G/G_v. Thus, we can view $\bar{\phi}$ as a function on the orbit Gv and extend it to a continuous function f on the whole space V by the Tietze extension theorem (Urysohn, 1925) (since G and thus Gv is a compact set). For this function f we have (using compactness of Gv and that $v_k \to v$)

$$\lim_{k\to\infty} \mathcal{I}_{\text{weighted}}[f](v_k) = \lim_{k\to\infty}\int_G f(gv_k)d\mu_{v_k} = \lim_{k\to\infty}\int_G f(gv)d\mu_{v_k} = \lim_{k\to\infty}\int_{G/G_v} \bar{\phi}(g)d\langle\mu_{v_k}\rangle_v(g). \tag{79}$$

On the other hand, we have

$$\mathcal{I}_{\text{weighted}}[f](v) = \int_G f(gv)d\mu_v = \int_{G/G_v} \bar{\phi}(g)\bar{\mu}_v(g). \tag{80}$$

Then (78) implies that $\lim_{k\to\infty} \mathcal{I}_{\text{weighted}}[f](v_k) \neq \mathcal{I}_{\text{weighted}}[f](v)$ so that $\mathcal{I}_{\text{weighted}}[f]$ is not continous and so $\mathcal{I}_{\text{weighted}}$ is not a BEC operator. □

Theorem 4.1. *Let $n, d > 1$ be natural numbers. Then Lebesgue almost every $a_1, \ldots, a_{n(d-1)} \in \mathbb{R}^d$ form a globally separated collection. Conversely, every globally separated collection must contain at least $n(d-1)$ vectors.*

Proof. **Proof of generic global monotonicity** First note that $X = (x_1, \ldots, x_n)$ is a separated if and only if $\bar{X} = (0, x_2 - x_1, \ldots, x_n - x_1)$ is a separated, which in turn will be a separated if and only if $\|\bar{X}\|^{-1}\bar{X}$ is a separated. Thus is it sufficient to consider X in the set

$$\mathbb{M} = \{X = (0, x_2, \ldots, x_n) \in \mathbb{R}^{d\times n}_{\text{distinct}} | \quad \|X\| = 1\}$$

which can be identified with the unit circle $S^{n(d-1)-1}$ which is an $n(d-1)-1$ dimensional semi-algebraic set.

Note that X is a separated unless X, a is a zero of the polynomial

$$F(X; a) = \prod_{i\neq j} a^T(x_i - x_j).$$

The proof is based on the finite witness theorem from (Amir et al., 2023). Namely

Theorem G.3 (Special case of Theorem A.2 in (Amir et al., 2023))**.** *Let $\mathbb{M} \subseteq \mathbb{R}^p$ be a semi-algebraic set of dimension D. Let $F : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ be a polynomial. Define*

$$\mathcal{N} = \{X \in \mathbb{M} | \quad F(X; a) = 0, \forall a \in \mathbb{R}^q\}$$

Then for generic $a_1, \ldots, a_{D+1}$,

$$\mathcal{N} = \{X \in \mathbb{M} | \quad F(X; a_i = 0, \forall i = 1, \ldots, D+1\}$$

Applying to this theorem to our setting with $D = n(d-1) - 1$, since for every $X \in \mathbb{M}$ there exists a direction a for which X is separated, we see that $\mathcal{N}$ is the empty set. According to the theorem we have for generic $a_1, \dots, a_{D+1} \in \mathbb{R}^d$ that the set

$$\{X \in \mathbb{M} | \quad F(X; a_i = 0, \forall i = 1, \dots, D+1\}$$

is equal to $\mathcal{N}$ and thus is empty. Equivalently, every X will be separated with respect to at least one of the a_i.

Remark: We add a more elementary proof that a finite number of a_i is sufficient, but this proof has a worse cardinality, quadratic in n: Assume that we are given m vectors $a_i, i = 1, \dots, m$ in $\mathbb{R}^d$ and assume that they are "full spark", meaning that any $d \times d$ matrix formed by choosing d of these vectors is full rank. Note that Lebesgue almost every choice of $a_i, i = 1, \dots, m$ in $\mathbb{R}^d$ will be full spark. Assume that the number of vectors m is strictly larger than $\binom{n}{2}(d-1)$. Now let $X \in \mathbb{R}^{d \times n}$ and assume that $g(X; a_i)$ is not uniquely defined for any $i = 1, \dots, m$. This means the for each such i there exists $s < t$ such that $a_i^T(x_s - x_t) = 0$. Since the number of (s,t) pairs is $\binom{n}{2}$ and $m > (d-1)\binom{n}{2}$, by the pigeon hole principle there exists a pair (s,t) for which $a_i^T(x_s - x_t) = 0$ for $\geq d$ different indices i.

By the full spark assumption it follows that $x_s - x_t = 0$ so X has a non-trivial stabilizer.

Optimality We now show that, if $a_1, \dots, a_k \in \mathbb{R}^d$ and $k < n(d-1)$, then $a_1, \dots, a_k$ are not globally separated.

Note that by adding zeros to the sequence $a_1, ..., a_k$ we may assume without loss of generality that $k = d(n-1) - 1$.

We partition the $a_i \in \mathbb{R}^d$ as $\{a_1, ..., a_s\}, \{a_1^1, ..., a_d^1\}, ..., \{a_1^t, ..., a_d^t\}$, where each block $\{a_1^q, ..., a_d^q\}$ is linearly dependent (i.e. contained in a subspace of dimension $d-1$), and the vectors $a_1, ..., a_s$ have the property that any subset of size d is linearly independent (note this is vacuous if $s < d$). We can do this inductively by starting with all of the a_i and removing linearly dependent subsets of size d until there are none left. We note that $s + td = k$ and thus $s \equiv d - 1 \pmod{d}$ (so that in particular $s \geq d - 1$).

We now construct the $x_1, ..., x_n \in \mathbb{R}^d$ as follows. First set $x_1 = 0$. Next, set $x_2 \neq 0$ satisfying the condition that $a_1 \cdot x_2 = a_2 \cdot x_2 = \cdots = a_{d-1} \cdot x_2$. Such an x_2 can be found since $a_1, ..., a_{d-1}$ cannot span the whole space $\mathbb{R}^d$.

Now, for $p = 1, ..., v$, where $s = d - 1 + dv$, we set x_{p+2} to be the unique vector satisfying

$$a_{pd} \cdot x_{p+2} = 0, \ \cdots \ , a_{pd+d-2} \cdot x_{p+2} = 0, \ a_{pd+d-1} \cdot (x_{p+2} - x_2) = 0. \tag{81}$$

This vector exists and is unique since by construction $a_{pd}, ..., a_{pd+d-1}$ are linearly independent so that the linear system in (81) is non-singular.

At this stage, we verify that for each $l = 1, ..., s$ there exist $1 \leq i \neq j \leq v + 2$ such that $a_l \cdot x_i = a_l \cdot x_j$. This follows easily from the construction. Indeed, if $l = 1, ..., d-1$ we can take $i = 1, j = 2$. If $l = pd + k$ with $0 \leq k < d - 1$ and $p \geq 1$ we can take $i = 1$ and $j = p + 2$. Finally, if $l = pd + d - 1$ with $p \geq 1$ we can take $i = 2$ and $j = p + 2$.

Next, we verify that all of the vectors $x_1, ..., x_{v+2}$ are distinct. We first claim that $x_i \neq x_1 = 0$ if $i > 1$. This follows by construction for $i = 2$, while if the solution to (81) were 0 for some p, then $a_{pd+d-1} \cdot x_2 = 0$. However, this would imply that a_{pd+d-1} lies in the same plane as $a_1, ..., a_{d-1}$, which contradicts the linear independence of any size d subset of $a_1, ..., a_s$. Next, we claim that $x_i \neq x_j$ if $i \neq j \geq 2$. Indeed, note that (81) implies that x_{p+2} is orthogonal to $a_{pd}, ..., a_{pd+d-2}$, while x_2 is orthogonal to $a_1, ..., a_{d-1}$. If any of these vectors were equal, then the planes spanned by their corresponding sets of $d - 1$ vectors would be the same, again contradicting the linear independence property of $a_1, ..., a_s$.

To complete the construction, we now choose for each block $\{a_1^q, ..., a_d^q\}$ a vector x_{v+2+q} which is orthogonal to $a_1^q, ..., a_d^q$ and is not equal to any of the previously chosen vectors. Such a vector can be chosen since the block $a_1^q, ..., a_d^q$ is linearly dependent (and thus lies in a $(d-1)$-dimensional subspace) by construction. It is clear that for an element a_k^q in block q we can now choose $i = 1$ and $j = v + 2 + q$ to ensure that $a_k^q \cdot x_i = a_k^q \cdot x_j$.

To complete the proof, we count the number of vectors produced by this construction, which is $v+2+t$. Since $s = d-1+dv$ and $s+td = k = d(n-1)-1$, we see that $(v+t+1)d-1 = d(n-1)-1$, from which it easily follows that $n = v+2+t$ as required.

□

Lemma 4.2. *$\mu^{\text{separated}}$ is a robust frame for $\mathbb{R}^{d \times n}_{\text{distinct}}$.*

Proof. We begin by noting that by definition the stabilizer of any $X \in \mathbb{R}^{d \times n}_{\text{distinct}}$ is trivial. In addition it is clear by construction that the frame $\mu^{\text{separated}}$ is equivariant.

To verify continuity, we let $X^k \to X$ be a convergent sequence in $\mathbb{R}^{d\times n}_{\text{distinct}}$, with limit $X \in \mathbb{R}^{d\times n}_{\text{distinct}}$. Since the stabilizer of X is trivial, we must show that $\mu^{\text{separated}}_{X^k} \to \mu^{\text{separated}}_X$ in the weak topology. Since these are probability distributions on a finite set, this means that for each permutation $\tau \in S_n$ we have $\mu^{\text{separated}}_{X^k}(\tau) \to \mu^{\text{separated}}_X(\tau)$.

First, we note that the denominator in (4.1) satisfies

$$\sum_{j=1}^{m} \tilde{w}_j(X) > 0 \tag{82}$$

since the collection a_i is globally separated. It is clear that $\tilde{w}_i(X)$ are continuous functions of X, so we have that in a sufficiently small neighborhood of X

$$\sum_{j=1}^{m} \tilde{w}_j(X) > \epsilon \tag{83}$$

for some $\epsilon > 0$.

Let $\tau \in S_n$ be arbitrary and consider the set

$$I(\tau, X) = \{i : \ a_i^T x_{\tau(1)} < a_i^T x_{\tau(2)} < \ldots < a_i^T x_{\tau(n)}\} \tag{84}$$

where x_i are the columns of X. $I(\tau, X)$ is the set of indices i such that $\tau = g_i(X)$ and X is strictly separated in the direction a_i.

From the definition of $\mu^{\text{separated}}$ it follows that

$$\mu^{\text{separated}}_X(\tau^{-1}) = \sum_{i\in I(\tau,X)} w_i(X) = \frac{\sum_{i\in I(\tau,X)} \tilde{w}_i(X)}{\sum_{j=1}^{m} \tilde{w}_j(X)}. \tag{85}$$

Since $X^k \to X$, there exists a sufficiently large N such that for all $k > N$ and all $i \in I(\tau, X)$ we have

$$a_i^T x^k_{\tau(1)} < a_i^T x^n_{\tau(2)} < \ldots < a_i^T x^k_{\tau(n)}, \tag{86}$$

where x_i^k are the columns of X^k, i.e. $I(\tau, X) \subset I(\tau, X^k)$. Thus, for $n > N$ we have

$$\mu^{\text{separated}}_{X^k}(\tau^{-1}) = \sum_{i\in I(\tau,X^k)} w_i(X^k) \geq \sum_{i\in I(\tau,X)} w_i(X^k) = \frac{\sum_{i\in I(\tau,X)} \tilde{w}_i(X^k)}{\sum_{j=1}^{m} \tilde{w}_j(X^k)} \tag{87}$$

Combined with the continuity of $\tilde{w}_i$ and (83), this implies that we have

$$\liminf_{k\to\infty} \mu^{\text{separated}}_{X^k}(\tau^{-1}) \geq \lim_{k\to\infty} \frac{\sum_{i\in I(\tau,X)} \tilde{w}_i(X^k)}{\sum_{j=1}^{m} \tilde{w}_j(X^k)} = \frac{\sum_{i\in I(\tau,X)} \tilde{w}_i(X)}{\sum_{j=1}^{m} \tilde{w}_j(X)} = \mu^{\text{separated}}_X(\tau^{-1}). \tag{88}$$

Since $\mu^{\text{separated}}_{X^k}$ and $\mu^{\text{separated}}_X$ are probability distributions, i.e. we have

$$\sum_{\tau\in S_n} \mu^{\text{separated}}_{X^k}(\tau) = 1 = \sum_{\tau\in S_n} \mu^{\text{separated}}_X(\tau)$$

it follows that we must actually have $\lim_{k\to\infty} \mu^{\text{separated}}_{X^k}(\tau) = \mu^{\text{separated}}_X(\tau)$ for all τ. □

Proposition 4.3. *$\mu^{\mathrm{S_n}}$ is a robust frame for S_n acting on $\mathbb{R}^{d\times n}$, with cardinality bounded by* $2 \cdot \sum_{k=0}^{d-1} \binom{\frac{n^2-n-2}{2}}{k}$.

Proof. We first prove the stated bound on the cardinality of the frame μ^{S_n}. Let $X \in \mathbb{R}^{d\times n}$. We can remove the measure 0 set of directions a for which $a^T x_i = a^T x_j$ for some pair of distinct columns x_i and x_j of X . Thus, we need to bound the cardinality of the set

$$\{\text{argsort}(a^T X) : \ a \in S(X)\}, \tag{89}$$

where the set $S(X)$ is given by

$$S(X) = \{a \in S^{d-1},\ a^T x_i = a^T x_j \implies x_i = x_j\}.$$

Note that argsort$(a, X) \neq$ argsort(a', X) for $a, a' \in S(X)$ implies that there are two distinct columns $x_i \neq x_j$ of X such that

$$a \cdot (x_i - x_j) < 0 < a' \cdot (x_i - x_j).$$

Thus, the cardinality of the set in (89) is equal to the number of regions that the hyperplanes $(x_i - x_j) \cdot a = 0$ divide the space $\mathbb{R}^d$ into (where x_i, x_j run over all distinct pairs of columns $x_i \neq x_j$ of X).

The total number of these hyperplanes is at most $\binom{n}{2}$. Next, we claim that given N hyperplanes in $\mathbb{R}^d$ passing through the origin, the number of regions that the space $\mathbb{R}^d$ is divided into is bounded by

$$2\sum_{k=0}^{d-1}\binom{N-1}{k}. \tag{90}$$

We prove this by induction on both N and d. It is clearly true for $N = 1$, since in this case the sum in (90) is 2 (we interpret $\binom{n}{k} = 0$ if $k > n$) and a single hyperplane divides the space into two pieces. Also, if $d = 1$, then the sum in (90) is also always 2, and in one dimension there is only one 'hyperplane' through the origin (i.e. the points 0 itself), which divides the space into two pieces.

Now, consider the case $N > 1$ and $d > 1$. Suppose that the first $N - 1$ hyperplanes divide $\mathbb{R}^d$ into M pieces. The number of these pieces which is divided into two by the N-th hyperplane is equal to the number of pieces that the N-th hyperplane (which is a $d - 1$ dimensional space) is divided into by its intersections with the first $N - 1$-hyperplanes. Thus, letting $M(N, d)$ denote the maximum number of pieces that N hyperplanes can divide $\mathbb{R}^d$ into, we obtain the recursive relation

$$M(N, d) \leq M(N-1, d) + M(N-1, d-1).$$

Combining this with the base cases discussed above gives the bound (90). Plugging in $N = \binom{n}{2}$ gives the claimed bound on the cardinality of μ^{S_n}.

Finally, we prove that μ^{S_n} is a robust frame. We need to show that this frame is weakly equivariant and continuous. For both of these computations we will first prove that the averaged measure $\bar{\mu}_X^{S_n}$ is given by

$$\bar{\mu}_X^{S_n}(g^{-1}) = \frac{1}{|G_X|}\mathbb{P}_{a\sim S^{d-1}}\left(a^T x_{g(1)} \leq \cdots \leq a^T x_{g(n)}\right). \tag{91}$$

Indeed, for any $g \in S_n$ we have

$$\mathbb{P}_{a\sim S^{d-1}}\left(a^T x_{g(1)} \leq \cdots \leq a^T x_{g(n)}\right) = \mathbb{P}(\exists h \in G_X,\ gh = \text{argsort}(a^T X)).$$

Moreover, an $h \in G_X$ such that $gh = \text{argsort}(a^T X)$), if it exists, must be unique by construction. This means that

$$\bar{\mu}_X^{S_n}(g^{-1}) = \frac{1}{|G_X|}\sum_{h\in G_X}\mu_X^{S_n}(h^{-1}g^{-1}) = \frac{1}{|G_X|}\mathbb{P}(\exists h \in G_X,\ gh = \text{argsort}(a^T X)), \tag{92}$$

which proves (91).

It is clear from (91) that $\bar{\mu}_{hX}^{S_n} = h^*\bar{\mu}_X^{S_n}$ for all $h \in S_n$, since

$$\bar{\mu}_{hX}^{S_n}(g^{-1}) = \frac{1}{|G_{hX}|}\mathbb{P}_{a\sim S^{d-1}}\left(a^T x_{gh(1)} \leq \cdots \leq a^T x_{gh(n)}\right),$$

while

$$h^*\bar{\mu}_X^{S_n}(g^{-1}) = \bar{\mu}_X^{S_n}(h^{-1}g^{-1}) = \frac{1}{|G_X|}\mathbb{P}_{a\sim S^{d-1}}\left(a^T x_{gh(1)} \leq \cdots \leq a^T x_{gh(n)}\right),$$

and G_X and G_{hX} are conjugate and thus have the same order. All that remains is to verify the continuity.

Let $X \in \mathbb{R}^{d\times n}$ and suppose that $X^k \to X$. Let G_X denote the stabilizer of X. By definition, we need to show that as $k \to \infty$

$$\left\langle \mu^{S_n}_{X^k} \right\rangle_X \to \bar{\mu}^{S_n}_X$$

weakly. Since S_n is a finite group, this means that we must show that for every $g \in S_n$, we have

$$\left\langle \mu^{S_n}_{X^k} \right\rangle_X (g) \to \bar{\mu}^{S_n}_X(g) \tag{93}$$

as $k \to \infty$.

Let $\delta_X > 0$ denote the smallest distance between two non-equal columns of X. We observe that if $|Y - X| < \delta_X/2$, then any two columns which are non-equal in X must also be non-equal in Y, so that $G_Y \subset G_X$. Thus, for sufficiently large k, we have $G_{X^k} \subset G_X$. This implies that

$$\left\langle \mu^{S_n}_{X^k} \right\rangle_X = \left\langle \bar{\mu}^{S_n}_{X^k} \right\rangle_X$$

for sufficiently large k, so that in (93) we can replace $\langle \mu^{S_n}_{X^k} \rangle_X$ by $\langle \bar{\mu}^{S_n}_{X^k} \rangle_X$.

Finally, since both $\langle \bar{\mu}^{S_n}_{X^k} \rangle_X$ and $\bar{\mu}^{S_n}_X$ are probability distributions, it suffices to show that for every $g \in S_n$ and $\epsilon > 0$

$$\langle \bar{\mu}^{S_n}_{X^k} \rangle_X(g) \geq \bar{\mu}^{S_n}_X(g) - \epsilon \tag{94}$$

for sufficiently large k.

Using the definition of $\mu^{S_n}_{X^k}$ and averaging over the stabilizers G_X and G_{X_k} we get

$$\langle \bar{\mu}^{S_n}_{X^k} \rangle_X(g^{-1}) = \frac{1}{|G_{X^k}||G_X|} \sum_{h \in G_X} \mathbb{P}_{a\sim S^{d-1}} \left(a^T x^k_{gh(1)} \leq \cdots \leq a^T x^k_{gh(n)} \right), \tag{95}$$

while

$$\bar{\mu}^{S_n}_X(g^{-1}) = \frac{1}{|G_X|} \mathbb{P}_{a\sim S^{d-1}} \left(a^T x_{g(1)} \leq \cdots \leq a^T x_{g(n)} \right). \tag{96}$$

For any $\delta > 0$ consider the event

$$A(\delta, X, g) := \left\{ a \in S^{d-1} : \ a^T x_{g(i)} + \delta \leq a^T x_{g(j)}, \ \forall i > j \text{ and } x_{g(i)} \neq x_{g(j)} \right\}. \tag{97}$$

In words, $A(\delta, X, g)$ is the set of directions $a \in S^{d-1}$ for which $a^T(gX)$ is sorted and all non-equal columns of X differ by at least δ. Observe that

$$\left\{ a \in S^{d-1} : \ a^T x_{g(1)} \leq \cdots \leq a^T x_{g(n)} \right\} - \bigcup_{\delta > 0} A(\delta, X, g) \tag{98}$$

contains only directions a such that $a^T x_i = a^T x_j$ for non-equal columns $x_i \neq x_j$ of X, and thus is a set of measure 0. This means that

$$\lim_{\delta \to 0} \mathbb{P}(A(\delta, X, g)) = \mathbb{P}_{a\sim S^{d-1}} \left(a^T x_{g(1)} \leq \cdots \leq a^T x_{g(n)} \right). \tag{99}$$

Thus, in light of (95) and (96) it thus suffices to show that for any $\delta > 0$ we have

$$\frac{1}{|G_{X^k}|} \sum_{h \in G_X} \mathbb{P}_{a\sim S^{d-1}} \left(a^T x^k_{gh(1)} \leq \cdots \leq a^T x^k_{gh(n)} \right) \geq \mathbb{P}(A(\delta, X, g)) \tag{100}$$

for sufficiently large k. To show this, suppose that $a \in A(\delta, X, g)$, i.e. $a \in S^{d-1}$ satisfies

$$a^T x_{g(i)} + \delta \leq a^T x_{g(j)}, \ \forall i > j \text{ and } x_{g(i)} \neq x_{g(j)}. \tag{101}$$

Choose k large enough so that for all columns i, we have $|x^k_i - x_i| < \delta/4$. Then we will also have

$$a^T x^k_{g(i)} + \delta/2 \leq a^T x^k_{g(j)}, \ \forall i > j \text{ and } x_{g(i)} \neq x_{g(j)}. \tag{102}$$

This means that by permuting only columns in X^k which are equal in X, we can sort $a^T(gX^k)$ (since any columns which are not equal in X are already sorted in $a^T(gX^k)$ by (102)), so that there exists an $h \in G_X$ such that

$$a^T x^k_{gh(1)} \leq \cdots \leq a^T x^k_{gh(n)}. \tag{103}$$

Moreover, h is obviously only unique up to multiplication by an element of G_{X^k} (which is contained in G_X since by (102), $x_i \neq x_j$ implies $x^k_i \neq x^k_j$). So each $a \in A(\delta, X, g)$ is contained in at least $|G_{X^k}|$ of the sets

$$\left\{ a \in S^{d-1} : \; a^T x^k_{gh(1)} \leq \cdots \leq a^T x^k_{gh(n)} \right\} \text{ for } h \in G_X. \tag{104}$$

This implies (100) and completes the proof.

□

Proposition 5.1. *$\mu^{SO(2)}$ is a robust frame.*

Proof. To prove weak equivariance we need to show that $\mu = \mu^{SO(2)}$ satisfies

$$\bar{\mu}_{gZ} = g_* \bar{\mu}_Z, \forall g \in S^1, Z \in \mathbb{C}^n$$

For $Z = 0$ both sides of this equation are the zero measure. For $Z \neq 0$ we note that $w_i(gZ) = w_i(Z)$ for all $g \in S^1$, that $g_i(gZ) = g \cdot g_i(Z)$ when $z_i \neq 0$, while when $z_i = 0$ we have $w_i(Z) = 0$, so that overall we obtain

$$\begin{aligned} \mu_{gZ} &= \sum_{i: z_i \neq 0} w_i(gZ) \delta_{g_i(gZ)} \\ &= \sum_{i: z_i \neq 0} w_i(Z) \delta_{g \cdot g_i(Z)} \\ &= g_* \mu_Z \end{aligned}$$

Finally we need to show that $\mu = \mu^{SO(2)}$ is a continuous weakly equivariant weighted frame. If $Z_n \to Z$ and $Z \neq 0$ it is straightforward to see that $\mu_{Z_n} \to \mu_Z$. If $Z_n \to 0$ then we note that for any probability measure μ on G we have that $\langle \mu \rangle_0 = 0 = \langle \mu_0 \rangle_0$, and so in particular $\langle \mu_{Z_n} \rangle_0 = \langle \mu_0 \rangle_0$ and $\mu_{Z_n} - \mu_{Z_n} = 0$ converges weakly to 0 so we are done. □

Proposition 6.1. *Let $\mu_{[\cdot]}$ be a robust frame. The restriction of $\mathcal{E}_{\text{weighted}}$ to* stable *input functions is a continuity-preserving, bounded, equivariant projection.*

Proof. Let $f : V \to W$ be a stable function. The stability assumption essentially allows us to reconstruct the proof from the invariant case, replacing $\mathcal{I}_{\text{weighted}}$ with $\mathcal{E}_{\text{weighted}}$. The critical observation is that if f is stable, then s stabilizes $gf(g^{-1}v)$ which occurs in the definition of $\mathcal{E}_{\text{weighted}}$. This is because in general

$$G_{gx} \subseteq g G_x g^{-1}$$

so

$$G_{gf(g^{-1}v)} = g G_{f(g^{-1}v)} g^{-1} \supseteq g G_{g^{-1}v} g^{-1} = G_v.$$

as a result, we can replace μ_v with $\bar{\mu}_v$ in the definition of $\mathcal{E}_{\text{weighted}}[f]$, because

$$\begin{aligned} \mathcal{E}_{\text{weighted}}[f](v) &= \int_G g f(g^{-1}v) d\mu_v(g) \\ &= \int_{G_v} \int_G sgf((sg)^{-1}v) d\mu_v(g) ds \\ &= \int_{G_v} \int_G gf(g^{-1}v) ds_* \mu_v(g) ds \\ &= \int_G gf(g^{-1}v) d\bar{\mu}_v(g) \end{aligned}$$

Using this we can now show that $\mathcal{E}_{\text{weighted}}[f]$ is an equivariant function because for every $h \in G$ and $v \in V$,

$$\begin{aligned}
\mathcal{E}_{\text{weighted}}[f](hv) &= \int_G gf(g^{-1}hv)d\mu_{hv}(g) \\
&= \int_G gf(g^{-1}hv)d\bar{\mu}_{hv}(g) \\
&= \int_G gf(g^{-1}hv)dh_*\bar{\mu}_v(g) \\
&= \int_G hgf((hg)^{-1}hv)d\bar{\mu}_v(g) \\
&= h\left[\int_G gf(g^{-1}v)d\bar{\mu}_v(g)\right] \\
&= h\left[\int_G gf(g^{-1}v)d\mu_v(g)\right] \\
&= h\mathcal{E}_{\text{weighted}}[f](v).
\end{aligned}$$

If f is equivariant then $\mathcal{E}_{\text{weighted}}[f] = f$ since $gf(g^{-1}v) = f(v)$ from equivariance.

Boundedness also follows easily since $d\mu_v$ is a probability measure, so that on every compact $K \subseteq V$ which is also closed under the action of G, we have for every $v \in K$ that

$$|\mathcal{E}_{\text{weighted}}[f](v)| \leq \left|\int_G gf(g^{-1}v)d\mu_v(g)\right| \leq \max_{g\in G}\|g\| \max_{w\in K}|f(w)|, \tag{105}$$

where $\|g\|$ denotes the operator norm of the linear operator g which is bounded due to the fact that G is compact acting linearly and continuously.

Continuity Let $f : V \to \mathbb{R}$ be a continuous functions. Let v_k be a sequence converging to $v \in V$. We need to show that $\mathcal{E}_{\text{weighted}}[f](v_k)$ converges to $\mathcal{E}_{\text{weighted}}[f](v)$. We observe that

$$\begin{aligned}
&|\mathcal{E}_{\text{weighted}}[f](v_k) - \mathcal{E}_{\text{weighted}}[f](v)| \\
&= \left|\int gf(g^{-1}v_k)d\mu_{v_k}(g) - \int gf(g^{-1}v)d\mu_v(g)\right| \\
&\leq \left|\int gf(g^{-1}v_k)d\mu_{v_k}(g) - \int gf(g^{-1}v)d\mu_{v_k}(g)\right| + \left|\int gf(g^{-1}v)d\mu_{v_k}(g) - \int gf(g^{-1}v)d\bar{\mu}_v(g)\right|.
\end{aligned}$$

Since f is stable we have

$$\int gf(g^{-1}v)d\mu_{v_k}(g) - \int gf(g^{-1}v)d\langle\mu_{v_k}\rangle_v(g) = 0. \tag{106}$$

and by assumption $\langle\mu_{v_k}\rangle_v \to \bar{\mu}_v$ weakly the second term tends to 0 (since f and the action of G are continuous). The first term tends to zero because $gf(g^{-1}v_k)$ converges to $gf(g^{-1}v)$ uniformly in g as k tends to infinity.

□