Equivariant Frames and the Impossibility of Continuous Canonicalization

Nadav Dym*, Hannah Lawrence*, Jonathan W. Siegel*
Technion, MIT, Texas A&M University

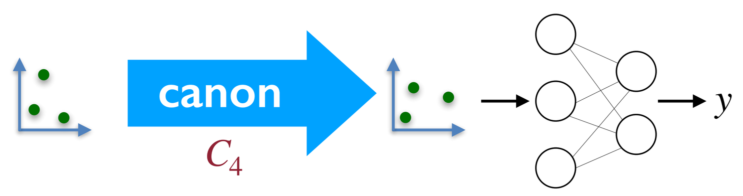
ICML International Conference On Machine Learning

Equivariant Canonicalization

Methods for making an *arbitrary* function equivariant ($f(gv) = gf(v)$):

Canonicalization: a function $h : V \to G$, $h(gv) = gh(v)$, applied as

$f(v) \mapsto \tilde{f}(v) := f(h_v^{-1} v)$



*Definition can be generalized to picking out orbit representatives when v is self-symmetric

Frame: a function $\mathcal{F} : V \to 2^G$, $\mathcal{F}(gv) = g\mathcal{F}(v)$, applied via "averaging"

$$f(v) \mapsto \tilde{f}(v) := \frac{1}{|\mathcal{F}(v)|} \sum_{g \in \mathcal{F}(v)} f(g^{-1} v)$$

Special case: group averaging $f(v) \mapsto \tilde{f}(v) := \frac{1}{|G|} \sum_{g \in G} f(g^{-1} v)$

Advantages over equivariant architectures: flexible, simpler to implement

Canonicalization | Group Averaging | Frames | **Weighted Frames**

Seemingly modest goal:

Continuous function f → Canonicalization → **Continuous** invariant function $\tilde{f}$

For some group actions, a canonicalization that preserves continuity can seem hard to construct. The intuitive options below are **discontinuous**:

Lexicographic Sort

S_n acting on $\mathbb{R}^{2 \times n}$

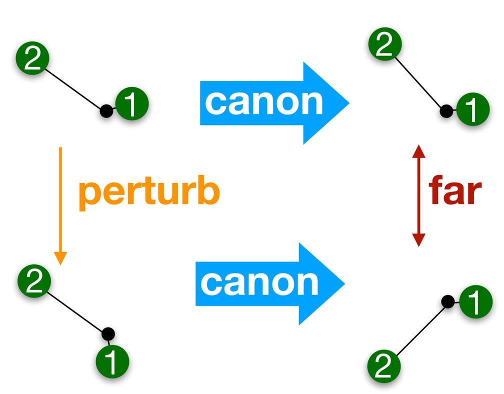
$$\begin{bmatrix} 3 & 3 & 4 \\ 8 & 7 & 2 \end{bmatrix} \xrightarrow{\text{canon}} \begin{bmatrix} 3 & 3 & 4 \\ 7 & 8 & 2 \end{bmatrix}$$

perturb ↓ ↕ far

$$\begin{bmatrix} 3 & 3.01 & 4 \\ 8 & 7 & 2 \end{bmatrix} \xrightarrow{\text{canon}} \begin{bmatrix} 3 & 3.01 & 4 \\ 8 & 7 & 2 \end{bmatrix}$$

Rotate by Max r

$SO(2)$ acting on $\mathbb{R}^{2 \times n}$



*Is there **any canonicalization** that preserves continuity for these actions?*

Theorem: Often, there is **no** continuous canonicalization

*Is there **any frame** that preserves continuity for these actions?*

Theorem: Often, there is **no** small frame

There is **no** continuity-preserving frame other than group-averaging for S_n acting on $\mathbb{R}^{d \times n}$ $(d, n > 1)$, and **no** *finite* continuity-preserving frame for $SO(2)$ acting on $\mathbb{R}^{2 \times n}$ $(n \geq 2)$

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

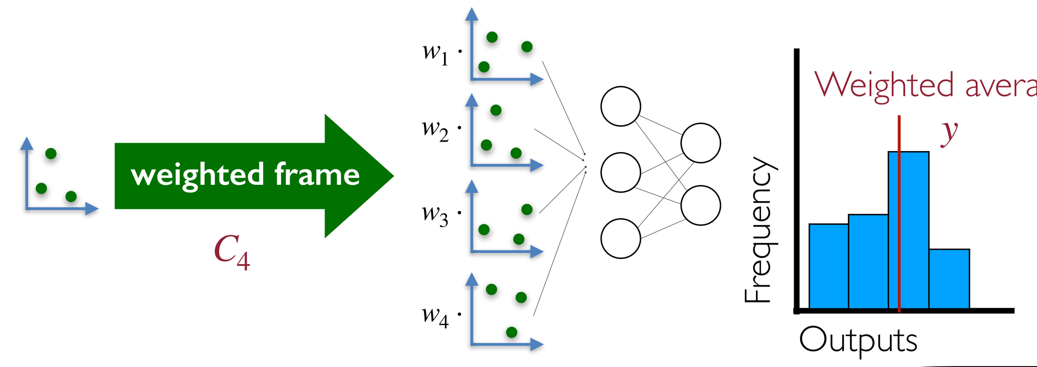
Table 1. Summary of main results. For various group actions, we show lower and upper bounds on the minimal cardinality N for which a continuity-preserving frame or weighted frame exists, and whether a continuous canonicalization exists (in which case $N = 1$). The $N = \infty$ result for unweighted $SO(d)$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

Can the ideas of frame averaging still be applied efficiently to such groups?

Solution: Robust Frames

Goal: Construct continuity-preserving frames that are efficient to apply

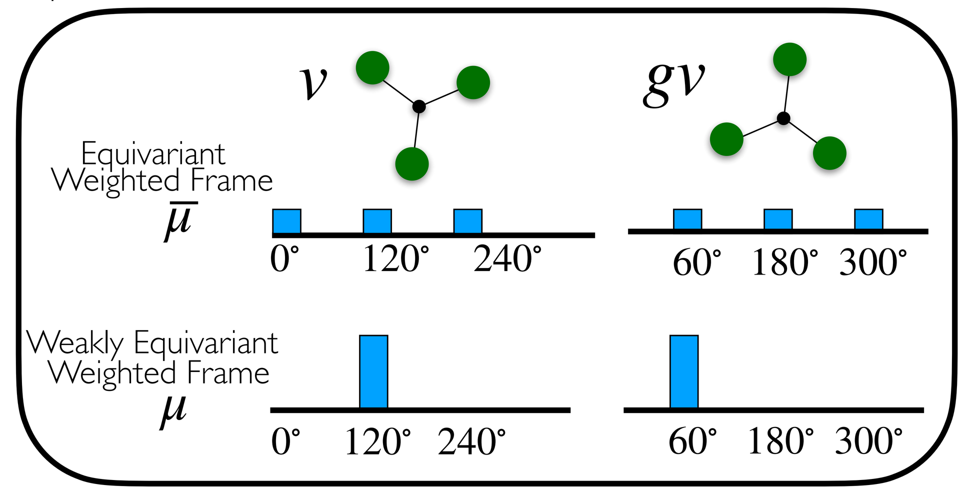
An **equivariant weighted frame** $\mu : V \to \mathcal{P}(G)$ is an equivariant map to the space $\mathcal{P}(G)$ of probability measures on G, i.e. $\mu(gv) =$ the push forward of $\mu(v)$ by g



$f(v) \mapsto \tilde{f}(v) := \mathbb{E}_{g \sim \mu(v)}[f(gv)]$

size of frame $= \max_v \mathrm{Supp}(\mu(v))$

We relax equivariance to *weak equivariance* at self-symmetric inputs v, which suffices to ensure output invariance



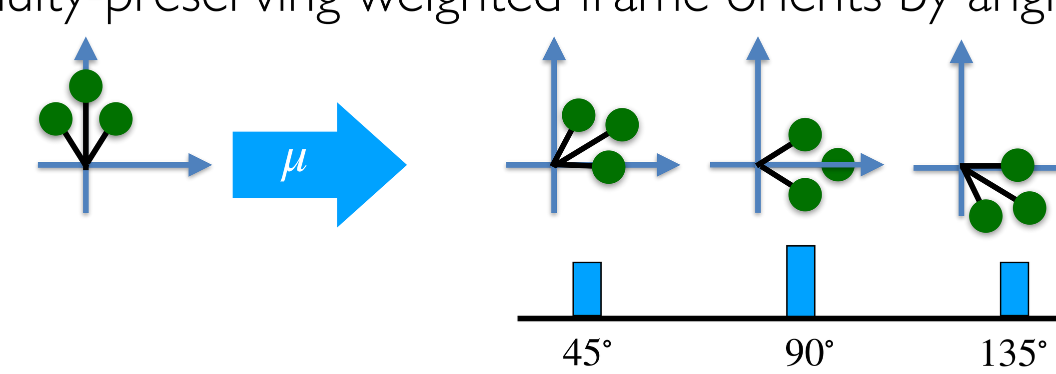
Proposition: A weighted frame μ preserves continuity, i.e. is a ***robust frame***, iff $v \mapsto \mu(v)$ itself is *continuous* (up to stabilizers).

Permutations: Consider the action of S_n on point clouds $X \in \mathbb{R}^{d \times n}$.

$$\mu^{S_n}(X) := \sum_{g \in S_n} \mathbb{P}_{a \sim S^{d-1}}(g^{-1} = \operatorname{argsort}(a^T X))\delta_g$$

μ^{S_n} is a continuity-preserving weighted frame with $O(n^{2(d-1)})$ elements.

Rotations: Consider the action of $SO(2)$ on point clouds $X \in \mathbb{R}^{2 \times n}$. A continuity-preserving weighted frame orients by angle, and weights by radius:



We generalize this construction to higher dimensions d.

From Invariant to Equivariant

Weighted frame averaging as $\mathbb{E}_{g \sim \mu(v)}[g^{-1} f(gv)]$ produces *equivariant* functions when μ is equivariant, but not when μ is only weakly equivariant.

However, weak equivariance is desirable computationally!

Defn. *Stabilizer* of v: $G_v = \{g : gv = v\}$. *Stable function* satisfies $G_v \subseteq G_{f(v)} \forall v$.

Stable functions: special function class for which weak equivariance suffices

Weakly equivariant frame + Stable function ⟹ Equivariant function

How do we parametrize stable functions as learnable networks?

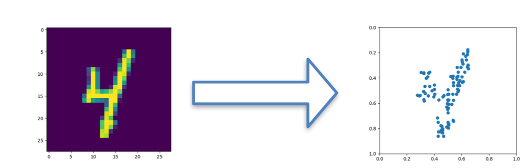
Example: For $SO(2)$ acting on $\mathbb{R}^{2 \times n}$, $f : \mathbb{R}^{2 \times n} \to \mathbb{R}^{2 \times n}$ is stable iff $f(\mathbf{0}) = \mathbf{0}$.

Stable frames: alternative that offloads stability to the frame. E.g. for $SO(2)$, weights rotations θ and $-\theta$ equally in each $\mu(v)$

Experiments

For permutations acting on $\mathbb{R}^{d \times n}$, our weighted frame works best:

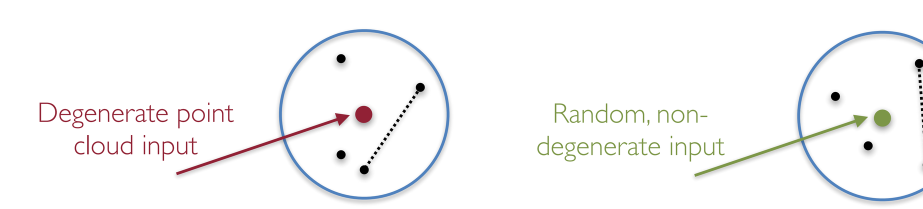
Task: Classify MNIST digit based on unordered point cloud



Invariance Method	Test Accuracy (%)
No Invariance	25.5
Discontinuous Canonicalization	85.6
Robust Frames	88.7
Reynolds Operator	22.6

Methods: no canonicalization, a discontinuous canonicalization (sorting along one axis), sampling 25 elements from the robust frame μ^{S_n}, and sampling 25 elements from S_n uniformly

We verify a discontinuity in a trained point cloud network:



Metric: average pairwise distance in neural network value between point clouds in a neighborhood of the hypothesized discontinuous input vs random input

We train the O(3)-equivariant network from the equiadapt library [2] on ModelNet40, and verify that the end-to-end function is discontinuous near a degenerate (rank 1) point cloud b

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

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

→ Empirically, there isn't a valid limiting value at b!

Future Work

- Stronger notions of smoothness, e.g. Lipschitz continuity
- Lower bounds on the size of weighted $SO(d)$ frames
- Analysis of efficiency of *sampling* from weighted frames
- Practical implementation and empirical analysis

References

1. Puny, O. et al, Frame averaging for invariant and equivariant network design. 2021
2. Kaba, S.-O. et al, Equivariance with learned canonicalization functions. 2022
3. Kim, J. et al, Learning probabilistic symmetrization for architecture agnostic equivariance. 2023
4. Pozdnyakov, S. et al, Smooth, exact rotational symmetrization for deep learning on point clouds. 2023

* equal contribution