Harvard John A. Paulson School of Engineering and Applied Sciences

Achieving Approximate Symmetry Is Exponentially Easier than Exact Symmetry

Behrooz Tahmasebi Melanie Weber

Harvard John A. Paulson School of Engineering and Applied Sciences (SEAS)

ICLR

Preface

- Our goal is to understand how to enforce symmetry in machine learning models efficiently when exact symmetry is computationally expensive.
- Exact symmetry is typically enforced via *group averaging* over a symmetry group G. For large groups, $|G|$ can be exponential, making exact averaging computationally infeasible; in practice, alternative methods are often used.
- Empirically, approximate symmetry has shown greater flexibility and robustness in many practical scenarios.
- This raises a fundamental question: is approximate symmetry inherently easier and more scalable than exact symmetry?
- We study *exponentially sparse group averaging*, where a small subset $S \subset G$, with $|S| = \log|G|$, is used instead of the full group. We show that sparse group averages are computationally efficient while accurately approximating full group averaging.
- This leads to scalable methods for symmetry-aware learning, applicable across a range of models and settings.
- Our main result shows that achieving approximate symmetry is exponentially easier than exact symmetry; when exact symmetry is infeasible, relaxing symmetry is provably efficient and accurate.

Group Averaging and Approximate Symmetry

- A fundamental approach for enforcing symmetry in machine learning models: given a function $f : \mathcal{X} \to \mathbb{R}$ and a symmetry group G, the goal is to construct a G-invariant functions.
- A classical approach is group averaging:

$$f_G(x) = \frac{1}{|G|} \sum_{g \in G} f(gx) \implies f_G(gx) = f_G(x), \quad \forall g \in G. \tag{1}$$

- However, for many groups of interest (e.g., permutations, rotations, transformations), the size $|G|$ can be exponential, making exact averaging computationally infeasible.
- **Example.** For permutation invariance on d coordinates, the group size is $|G| = d!$, which grows super-exponentially, for sign invariance we have $|G| = 2^d$.
- In practice, one often resorts to approximate symmetry, using only a subset of transformations.
- We study sparse group averaging, where

$$f_S(x) = \frac{1}{|S|} \sum_{g \in S} f(gx), \quad S \subset G, \quad |S| \ll |G| \tag{2}$$

- A key question is whether such approximations can still capture the effect of full group averaging. Our results show that choosing $|S| = \mathcal{O}(\log|G|)$ is sufficient to accurately approximate f_G.
- This provides a scalable alternative to exact symmetry enforcement, applicable to large-scale and structured learning problems.

Our Contributions

- This paper studies the problem of enforcing symmetry in machine learning, and asks: how much averaging is needed to obtain its full benefits?
- Exact symmetry is hard! Enforcing exact invariance requires averaging over the entire group, which is computationally infeasible even for simple function classes.
- Sparse averaging is enough! Instead of averaging over the full group (exponentially large), using a small subset (of size $\log|G|$) achieves essentially near invariance.

Main Result: Sparse Averaging

Can we approximate full group averaging with only a small subset?

Theorem: For a random subset $S \subset G$, the below uniform bound holds with probability at least $1 - \delta$:

$$\boxed{\sup_{\|f\|_{L^2(\mathcal{X})} \leqslant 1} \left| \frac{1}{|G|} \sum_{g \in G} f(gx) - \frac{1}{|S|} \sum_{g \in S} f(gx) \right| \leqslant 2.67 \cdot \frac{\log(|G|/\delta)}{|S|}}$$

Main Result: Why It Matters

- Exponential savings. Full averaging requires $|G|$ evaluations, while sparse averaging needs only $|S| = \mathcal{O}(\log|G|)$.
- Scalable symmetry. This makes symmetry enforcement feasible even for very large groups.
- Dimension dependence. For permutation groups, $|G| = d!$, giving $|S| = \mathcal{O}(d \log d)$ instead of $\exp(\Theta(d \log d))$. Similarly, for sign invariances, we achieve $|S| = \mathcal{O}(d)$ instead of $\exp(\Theta(d))$.
- General applicability. The result holds uniformly over the function class.

Exact Symmetry is Hard

What if we want exact invariance?

$$\hat{f}(gx) = \hat{f}(x), \quad \forall g \in G. \tag{3}$$

- Enforcing exact symmetry requires matching the full group average.
- For moderately structured function classes (e.g., polynomials), this imposes strong constraints.

Theorem: For polynomial features up to degree K, exact invariance requires averaging over the full group $S = G$ once

$$K \geqslant \frac{d(d-1)}{2} \quad \text{(for permutations on } d \text{ elements).}$$

- More generally, the required degree K admits bounds in terms of the spectral properties (eigenvalue structure) of the group representation matrices, capturing how symmetry constraints propagate across features.
- Even for moderate degree, exact symmetry requires full group averaging.
- This shows a sharp contrast: exact symmetry is computationally hard, while approximate symmetry is easy.
- There is an exponential separation between exact and approximate symmetry enforcement via averaging.

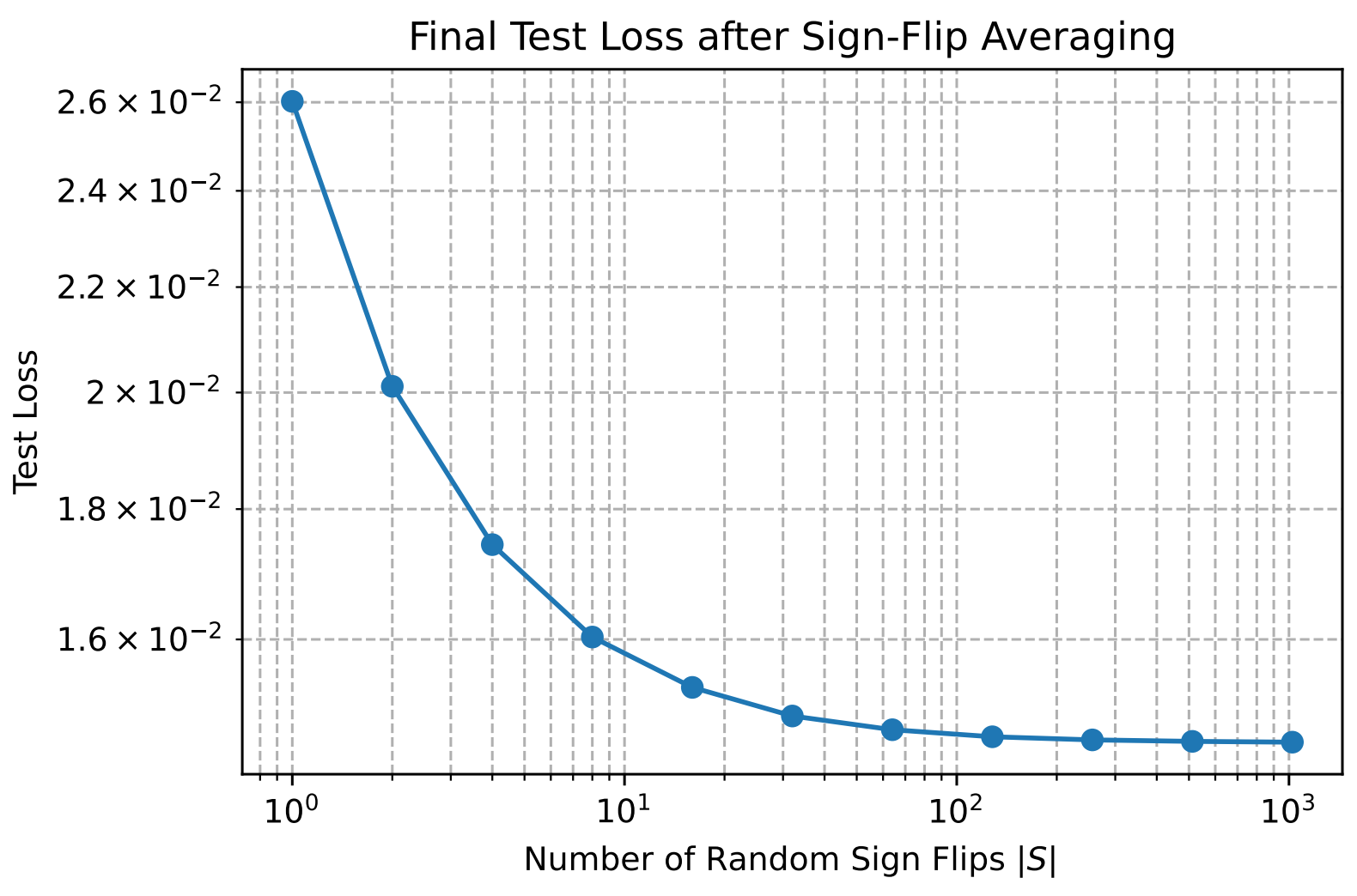
Beyond Uniform Averaging and Generalization

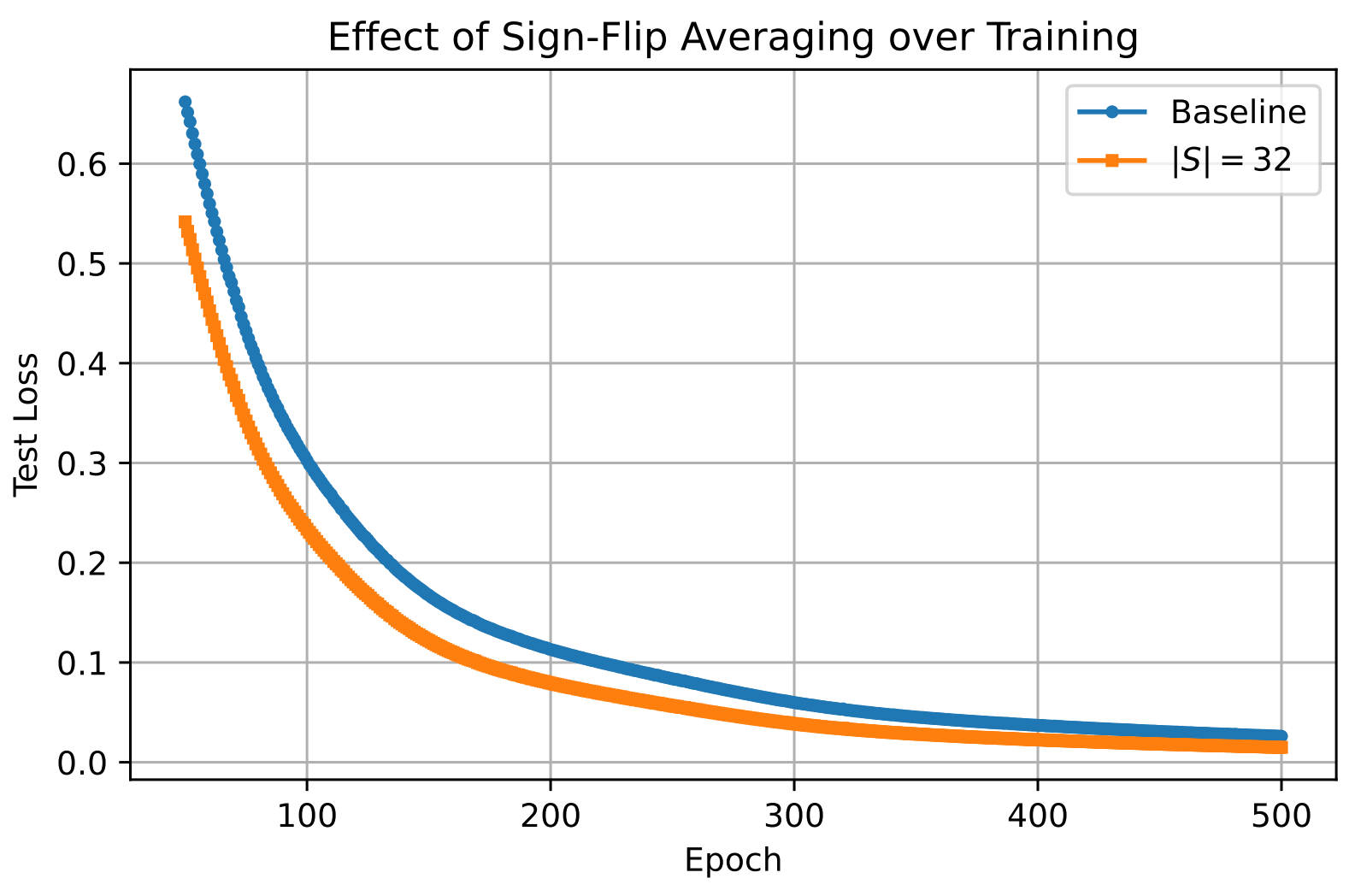
- Unlike optimal transport or uniform averaging methods, our approach avoids the curse of dimensionality by exploiting group structure.
- Rather than requiring uniform coverage of the space, we only sample group elements, leading to logarithmic dependence on $|G|$.
- We further establish generalization error bounds for approximate symmetry.
- In particular, sparse averaging achieves the full statistical benefits of exact symmetry with only logarithmic cost.

Proof Sketch

- We study the deviation between full group averaging and sparse averaging uniformly over functions with $\|f\|_{L^2(\mathcal{X})} \leqslant 1$.
- The key idea is to view sparse averaging as a random sampling process over the group G.
- We apply concentration inequalities to control the deviation between empirical (subset) and population (full group) averages.
- The analysis reduces to bounding a uniform deviation over the function class, leading to a logarithmic dependence on $|G|$.
- A careful representation-theoretic argument yields the final bound.
- For exact symmetry, we instead analyze when averaging operators annihilate non-invariant components, which requires access to the full group (Fourier analysis over groups).

Experiments





Complete Version of the Paper

behrooz_tahmasebi@seas.harvard.edu