A Theoretical Framework for Discovering Groups and Unitary Representations via Tensor Factorization

Dongsung Huh[1] **Halyun Jeong**[2]
[1] IBM Research [2] University at Albany, SUNY

UNIVERSITY AT ALBANY
STATE UNIVERSITY OF NEW YORK

Motivation: Symmetry Discovery

The Status Quo: Hard-Coded Symmetry. Geometric Deep Learning (e.g., Equivariant NN) relies on *a priori* knowledge of symmetries to constrain the model. This guarantees generalization but limits applicability when symmetries are unknown.

The Challenge: Automated Discovery. Can a model *discover* the underlying symmetry group directly from data (e.g., an operation table), without hard-coding it?

HyperCube is a tensor factorization model that exhibits a strong inductive bias toward **Groups** and **Unitary Representations**, effectively learning symmetry from data.

Problem Setup

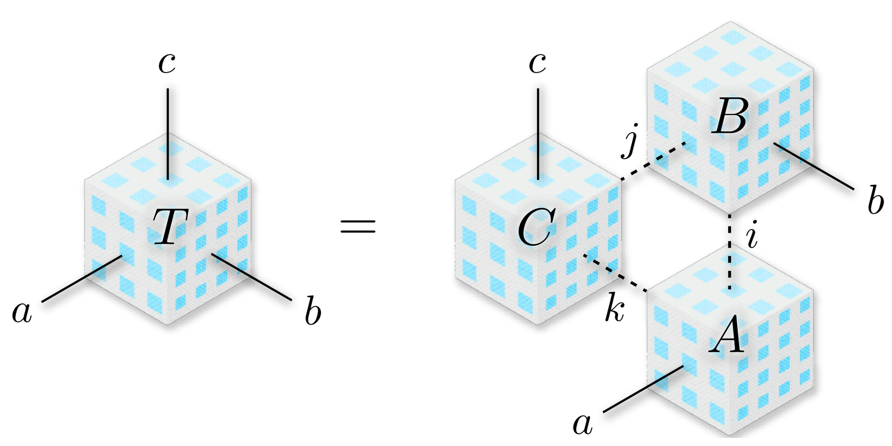
$(Q, \circ)$: a binary operation. $|Q| = n$.
Cayley Tensor: $\delta_{abc} = \mathbb{I}_{\{a \circ b = c\}}$.
Quasigroup: δ is a Latin-square.
Group: A quasigroup that is associative.
$|\delta| := \sum_{a,b,c} \delta_{abc} = n^2$.

HyperCube Model

Approximates δ via a trilinear product of parameter tensors $\Theta = (A, B, C)$:

$$T_{abc}(\Theta) = \frac{1}{n} \mathrm{Tr}(A_a B_b C_c) \qquad (1)$$

Uses matrix product of **linear operators** ($A_a \in \mathbb{C}^{n \times n}$) to capture interactions.



Regularization Objective.
Jacobian Penalty: *e.g.*, $\partial T_{abc}/\partial A_a = (B_b C_c)^\dagger$

$$\mathcal{H} = \sum_{b,c} \|B_b C_c\|^2 + \sum_{c,a} \|C_c A_a\|^2 + \sum_{a,b} \|A_a B_b\|^2$$
$$= \sum_{a,b,c} \delta_{abc} \left(\|B_b C_c\|^2 + \|C_c A_a\|^2 + \|A_a B_b\|^2 \right).$$

How does this induce bias towards groups?

Notation

Normalized inner product and norm

$$\langle X, Y \rangle := \frac{1}{n} \mathrm{Tr}(X^\dagger Y), \qquad \|X\|^2 := \langle X, X \rangle$$

so that $\|U\|^2 = 1$ for any unitary U.

Geometric Mechanism

1. C–S Lower Bound.
From Cauchy-Schwarz inequality:

$$\left\| \frac{\partial T_{abc}}{\partial A_a} \right\|^2 \geq \frac{|T_{abc}|^2}{\|A_a\|^2}$$

Derive a lower bound $\mathcal{H} \geq \mathcal{B}$:

$$\mathcal{B} := \sum \delta_{abc} |T_{abc}|^2 \left(\frac{1}{\|A_a\|^2} + \frac{1}{\|B_b\|^2} + \frac{1}{\|C_c\|^2} \right)$$

2. Orthogonal Decomposition.
$\mathcal{H}(\Theta) = \mathcal{B}(\Theta; \delta) + \mathcal{R}(\Theta; \delta)$, where

$$\mathcal{R} := \sum \delta_{abc} \left(\|\Delta^{(A)}_{abc}\|^2 + \|\Delta^{(B)}_{abc}\|^2 + \|\Delta^{(C)}_{abc}\|^2 \right)$$

with $\Delta^{(A)}_{abc} := \frac{T^*_{abc}}{\|A_a\|^2} A_a - (B_b C_c)^\dagger$
Orthogonality: $\langle A_a, \Delta^{(A)}_{abc} \rangle = T^*_{abc} - T^*_{abc} = 0$.

3. Collinearity Manifold.
$\mathcal{R} = 0$ iff A_a and $(B_b C_c)^\dagger$ are collinear.

On the Manifold ($\mathcal{R} = 0$):
4. Constant Gram matrix:

$$X = \frac{A_a A_a^\dagger}{\|A_a\|^2} = \frac{C_c^\dagger C_c}{\|C_c\|^2}$$

$$\kappa := \frac{\|A_a\|^2 \|B_b\|^2 \|C_c\|^2}{|T_{abc}|^2} = \frac{\mathrm{rank}(X)}{n} \leq 1$$

5. AM-GM bound within the manifold:

$$\mathcal{B} \geq 3\kappa^{-1/3} |\delta|$$

$\mathcal{B}$ forces $\kappa \to 1$ (Full Rank) $\implies$ **Unitarity**.

Theoretical Results

Regular Representation Certificate.
For any group, the *Regular Representation* (ρ_r) defines a *unitary collinear factorization*:

$$A_g = B_g = C_g^\dagger = \rho_r(g) \implies \mathcal{H} = 3|\delta|, \mathcal{R} = 0.$$

Uniqueness Theorem.
Any unitary collinear factorization ($\mathcal{R} = 0$) is unitarily equivalent to ρ_r.

Implication: A quasigroup admitting $\mathcal{R} = 0$ *must* be a group isotope.

Collinearity Dominance Conjecture.
Misalignment penalty $\mathcal{R}$ dominates $\mathcal{B}$. This ensures optimization stays near $\mathcal{R} = 0$, ruling out spurious minima.

Under the conjecture:
Global minimum for Group isotopes.

$$\mathcal{H}_{\min}(\delta) = \mathcal{H}(\Theta^*) = 3|\delta|$$

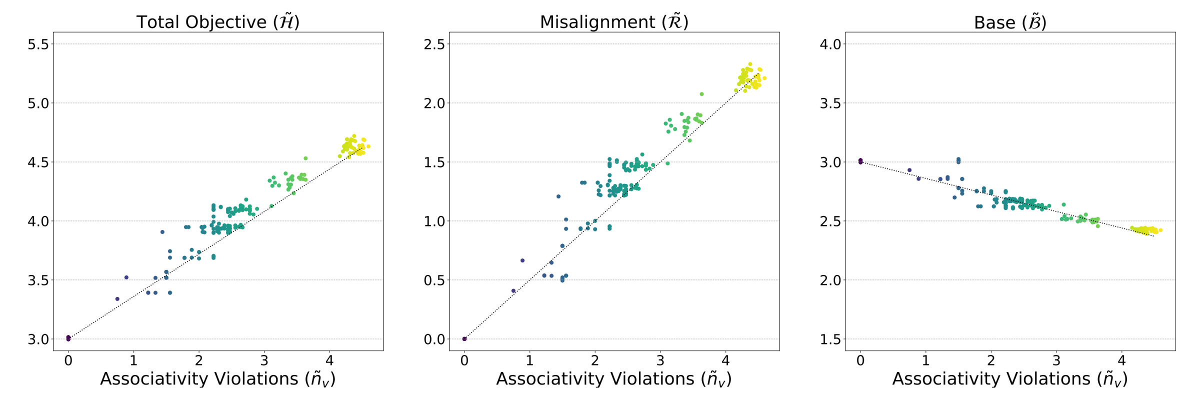
Θ^* is given by unitary collinear factors: ρ_r.

Non-group isotopes must incur **strictly higher loss**:

$$\mathcal{H}_{\min}(\text{Non-Group}) > \mathcal{H}_{\min}(\text{Group})$$

Empirical Verification

We optimized $\mathcal{H}$ for Latin-squares of orders $n \in \{5, 6, 7, 8\}$. We compare the minimum loss against the **Associativity Violation** n_v.



Linear Scaling Laws.
The results reveal a precise tradeoff:

- **Misalignment Penalty:** $\mathcal{R}_{\min} \approx 0.50 \cdot n_v$.
- **Rank "Savings" (Right):** $\mathcal{B}$ decreases with non-associativity, but only by $\mathcal{B}_{\min} \approx 3|\delta| - 0.14 \cdot n_v$.

Collinearity Dominance.
The penalty slope (0.50) strictly dominates the savings (0.14).

$$\mathcal{H}_{\min}(\delta) \approx 3|\delta| + 0.36 \cdot n_v.$$

This confirms the global landscape forces optimization to **Group Isotopes** ($\tilde{n}_v = 0$).

Conclusion: $\mathcal{H}$ provides a **differentiable proxy** for groupness (associativity violation).

Takeaways & Outlook

1. Novel Inductive Bias: Unitarity.
Standard DL models exhibit *Low-Rank* bias (interpolation). In contrast, HyperCube actively **maximizes rank** to achieve unitarity and enables **discrete structure discovery** beyond the limits of interpolation.

2. Mechanism of Discovery.
$\mathcal{H}$ decomposes into geometric forces — Collinearity + Rank Maximization — that structurally mandate Group theory.

3. Differentiable Associativity.
$\mathcal{H}$ serves as a differentiable proxy for associativity, allowing gradient-based discovery of discrete symmetries.

Future Work.
Analytic proof of the Dominance Conjecture and extension to partial observation (Tensor Completion).

NeuReps Workshop @ NeurIPS 2025