Theoretical Analysis of HyperCube Objective for Group Representation Learning

Dongsung Huh DONGSUNGHUH@GMAIL.COM
IBM Research

Halyun Jeong HJEONG2@ALBANY.EDU
Dept of Mathematics & Statistics, University at Albany, State University of New York

Editors: List of editors' names

Abstract

The HyperCube model is a promising tensor factorization framework for discovering latent group structures in data. A foundational conjecture posits that its global minima correspond to unitary group representations, but a proof has remained elusive. We make significant theoretical progress by decomposing the HyperCube objective into a base term dependent on matrix norms and a *misalignment* term. We introduce the *Perfect Alignment Conjecture*, which states that this misalignment must vanish at any stationary point for the optimization to capture a true group. Under this condition, we prove that all local minima are in fact global and are unitarily equivalent to the group's regular representation, thus conditionally resolving the original conjecture. Our analysis reveals HyperCube's unique inductive bias for full-rank, unitary solutions, distinguishing it from typical low-rank models.

1. Introduction

Discovering hidden symmetry groups from data has long been central in physics, mathematics, and computer science. In modern deep learning, symmetry is typically *built in* through equivariant architectures (Bronstein et al., 2021), such as CNNs. However, directly *learning* group structure from data has remained challenging, since the discrete axioms of groups are incompatible with the continuous, gradient-based learning paradigm.

A recent work by Huh (2025) introduced a tensor factorization framework with a regularizer, called *HyperCube*, and showed empirically that it can efficiently recover hidden group structure. That study conjectured that the global minima of the model's loss landscape correspond to unitary representations of the group, but left this central question open.

In this paper, we make significant theoretical progress on this conjecture. We prove that, under a newly proposed *perfect alignment condition*, all local minima of the HyperCube objective are global and unitarily equivalent to the regular representation. The central open problem, which we formalize as the *Perfect Alignment Conjecture*, is whether every minimizer necessarily satisfies this condition.

2. Setup and notation

We follow the task setup from (Power et al., 2022) of learning finite binary operation tables.

Let $(G, \circ)$ be a finite set with $|G| = n$ and a binary operation $\circ$. We define the Cayley structure tensor as

$$\delta_{abc} := \mathbf{1}\{a \circ b = c\}, \qquad a, b, c \in G.$$

We assume $(G, \circ)$ is a quasi-group, meaning its operation table is a Latin square. $(G, \circ)$ is a group if $\circ$ is also associative, which will be our primary focus.

The HyperCube model learns a factorized parameterization of the Cayley tensor, subject to the constraint:[1]

$$\forall a, b, c \in G : \qquad \tfrac{1}{n} \operatorname{Tr}(A_a B_b C_c) = \langle C_c^\dagger, A_a B_b \rangle = \delta_{abc}. \tag{1}$$

Here, A, B, C are $n \times n \times n$ tensors and A_a, B_b, C_c are their respective $n \times n$ matrix slices. The objective is to minimize:

$$\mathcal{H}(A, B, C) := \sum_{b,c \in G} \|B_b C_c\|^2 + \sum_{c,a \in G} \|C_c A_a\|^2 + \sum_{a,b \in G} \|A_a B_b\|^2 \tag{2}$$

$$= \sum_{a,b,c \in G} \delta_{abc} \Big(\|B_b C_c\|^2 + \|C_c A_a\|^2 + \|A_a B_b\|^2 \Big), \tag{3}$$

where the second equality holds due to the Latin-square property. We aim to prove the following conjecture from Huh (2025), conditional on our new conjecture in Section 3.2.

Conjecture 1 (Main conjecture of Huh (2025)) *Subject to (1), the global minimizers of $\mathcal{H}$ consist of unitary matrices that are unitarily equivalent to the left-regular representation of G. The minimum value is $\mathcal{H}^* = 3n^2$.*

3. Decomposition and Bounds

3.1. Decomposition into Base and Misalignment

Let $\alpha_a := \|A_a\|$, $\beta_b := \|B_b\|$, and $\gamma_c := \|C_c\|$ denote the norms of the matrix slices. We define the **misalignment matrices** as

$$\Delta_{abc}^{(A)} := B_b C_c - \alpha_a^{-2} A_a^\dagger, \qquad \Delta_{abc}^{(B)} := C_c A_a - \beta_b^{-2} B_b^\dagger, \qquad \Delta_{abc}^{(C)} := A_a B_b - \gamma_c^{-2} C_c^\dagger. \tag{4}$$

Lemma 2 (Decomposition of $\mathcal{H}$) *The objective function (2) can be decomposed as*

$$\mathcal{H} = \mathcal{B} + \mathcal{R},$$

1. For matrices $X, Y \in M_n(\mathbb{C})$, we define the normalized Frobenius inner product as

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

where $\dagger$ denotes the conjugate transpose.

*where $\mathcal{B}$ is a **base term** that depends only on the norms (α, β, γ) and $\mathcal{R}$ is a **misalignment term** that quantifies the geometric alignment of the matrix products. Explicitly,*

$$\mathcal{B} = \sum_{a,b,c} \delta_{abc} \Big(\alpha_a^{-2} + \beta_b^{-2} + \gamma_c^{-2} \Big), \tag{5}$$

$$\mathcal{R} = \sum_{a,b,c} \delta_{abc} \Big(\|\Delta_{abc}^{(A)}\|^2 + \|\Delta_{abc}^{(B)}\|^2 + \|\Delta_{abc}^{(C)}\|^2 \Big). \tag{6}$$

Proof Expanding $\|\Delta_{abc}^{(C)}\|^2$ yields $\|A_a B_b - \gamma_c^{-2} C_c^\dagger\|^2 = \|A_a B_b\|^2 - 2\Re\langle C_c^\dagger, A_a B_b\rangle/\gamma_c^2 + \|C_c\|^2/\gamma_c^4$. Under the constraint (1), this becomes $\|A_a B_b\|^2 - 2\delta_{abc}/\gamma_c^2 + 1/\gamma_c^2$. Summing this over all triples (a, b, c) and cyclic permutations directly yields $\mathcal{R} = \mathcal{H} - \mathcal{B}$. ■

Lemma 3 (Lower bound) *The objective function is bounded below by*

$$\mathcal{H} \geq 3 \sum_{a,b,c} \delta_{abc} \left(\kappa_{abc}\right)^{-2/3}, \tag{7}$$

where $\kappa_{abc} := \alpha_a \beta_b \gamma_c$. Equality holds if $\alpha_a = \beta_b = \gamma_c$ for all (a, b, c) with $\delta_{abc} = 1$.

Proof Immediate from Lemma 2, $\mathcal{H} \geq \mathcal{B}$. Applying the AM–GM inequality $\alpha_a^{-2} + \beta_b^{-2} + \gamma_c^{-2} \geq 3 \left(\alpha_a^{-2} \beta_b^{-2} \gamma_c^{-2}\right)^{1/3}$ yields the desired result. ■

3.2. The Perfect Alignment Conjecture

Minimizing $\mathcal{H}$ requires simultaneously reducing both the base term $\mathcal{B}$ and the misalignment term $\mathcal{R}$. Our analysis suggests that for a true group G, any stationary point of the optimization must satisfy $\mathcal{R} = 0$. We formalize this as a new conjecture.

Conjecture 4 (Perfect Alignment Conjecture) *If $(G, \circ)$ is a group, then any stationary point (A, B, C) of the constrained optimization problem must satisfy the **perfect alignment condition** $\mathcal{R} = 0$.*

The remainder of this paper proves that, if this conjecture holds, then the original HyperCube conjecture follows.

4. Structure under Perfect Alignment

In this section, we explore the profound structural consequences of the perfect alignment condition ($\mathcal{R} = 0$). Our goal is to show that this condition forces the matrix slices $\{A_a\}$, $\{B_b\}$, and $\{C_c\}$ to be unitary and mutually aligned.

Lemma 5 *The misalignment vanishes ($\mathcal{R} = 0$) if and only if the following **alignment equations** hold for all $(a, b, c) \in G^3$ with $\delta_{abc} = 1$:*

$$A_a B_b = \gamma_c^{-2} C_c^\dagger, \qquad B_b C_c = \alpha_a^{-2} A_a^\dagger, \qquad C_c A_a = \beta_b^{-2} B_b^\dagger. \tag{8}$$

Lemma 6 (Index-Independent Gram Matrices) *If $\mathcal{R} = 0$ and $(G, \circ)$ is a group, there exist positive semidefinite matrices X, Y, Z, independent of the indices a, b, c, such that:*

$$X := A_a A_a^\dagger/\alpha_a^2 = C_c^\dagger C_c/\gamma_c^2, \qquad Y := B_b B_b^\dagger/\beta_b^2 = A_a^\dagger A_a/\alpha_a^2, \qquad Z := C_c C_c^\dagger/\gamma_c^2 = B_b^\dagger B_b/\beta_b^2.$$

Proof For any triple (a, b, c) with $\delta_{abc} = 1$, (8) yields: $C_c^\dagger C_c/\gamma_c^2 = (A_a B_b) C_c = A_a(B_b C_c) = A_a A_a^\dagger/\alpha_a^2$. Since this must hold for any b such that $c = a \circ b$, the resulting matrix X must be independent of the indices a and c. Cyclic permutations yield the other identities. ∎

Lemma 7 (Projection Equation) *If $\mathcal{R} = 0$, the product $\kappa := \alpha_a\beta_b\gamma_c$ is constant for all triples with $\delta_{abc} = 1$, and the matrix $P := \kappa^2 X$ is an orthogonal projection. Moreover,*

$$\kappa^2 = \frac{\operatorname{rank}(X)}{n} \leq 1.$$

Proof From Lemma 6, $X = C_c^\dagger C_c/\gamma_c^2 = \gamma_c^2 A_a(B_b B_b^\dagger)A_a^\dagger = \beta_b^2\gamma_c^2(A_a A_a^\dagger)(A_a A_a^\dagger)/\alpha_a^2 = \alpha_a^2\beta_b^2\gamma_c^2(A_a A_a^\dagger/\alpha_a^2)(A_a A_a^\dagger/\alpha_a^2) = \kappa_{abc}^2 X^2$. This yields $X = \kappa_{abc}^2 X^2$. Since X is index-independent, $\kappa := \kappa_{abc}$ must be a constant. Then, $P := \kappa^2 X$ is an orthogonal projection, since $P^2 = \kappa^4 X^2 = \kappa^2 X = P$. Then, $n = \operatorname{Tr}(X) = \operatorname{Tr}(P)/\kappa^2 = \operatorname{rank}(X)/\kappa^2$. ∎

We can now assemble these results into our main theorem, which characterizes the structure of the optimizers.

Theorem 8 (Structure of Global Minimizers) *Assume the Perfect Alignment Conjecture holds. Then any global minimizer of $\mathcal{H}$ consists of unitary matrices where $\alpha_g = \beta_g = \gamma_g = 1$ for all $g \in G$. The minimum value is $\mathcal{H}_{\min} = 3n^2$.*

Proof By the Perfect Alignment Conjecture, a global minimum must have $\mathcal{R} = 0$. From Lemma 3, the lower bound on $\mathcal{H}$ is minimized when the term $(\alpha_a\beta_b\gamma_c)^{-2/3} = \kappa^{-2/3}$ is maximized. Lemma 7 shows that any solution with $\mathcal{R} = 0$ must have $\kappa \leq 1$, so the maximum possible value is $\kappa = 1$. Such a point saturates the Cauchy–Schwarz bound and thus represents a minimum for the objective among all perfectly aligned solutions and is therefore a global minimizer.

With $\kappa = 1$, the lower bound becomes $\mathcal{H} \geq 3n^2$. For this bound to be achieved, the AM–GM inequality in Lemma 3 must be an equality, which requires $\alpha_a^2 = \beta_b^2 = \gamma_c^2$ for all valid triples (a, b, c). As $\kappa = \alpha_a\beta_b\gamma_c = 1$, this implies $\alpha_a = \beta_b = \gamma_c = 1$ for all $g \in G$.

From Lemma 7, $\kappa^2 = 1$ implies that X is a projection with full rank ($\operatorname{rank}(X) = n$), which is the identity matrix. Thus, $X = (Y = Z) = I$. Finally, the definition of X, Y, Z in Lemma 6 implies that all matrices $\{A_a\}$, $\{B_b\}$, $\{C_c\}$ are unitary. This achieves the bound $\mathcal{H} = \sum_{a,b,c} \delta_{abc}(1 + 1 + 1) = 3n^2$. ∎

This theorem shows that any global minimizer must be composed of three families of aligned unitary matrices. As shown in Appendix A, such solutions exist and are all unitarily equivalent to the left-regular representation of the group G.

5. Conclusion and Future Work

Our work provides the first rigorous theoretical analysis of the HyperCube model and shows that it correctly recovers group structure through representation learning, conditional on our Perfect Alignment Conjecture. This analysis reveals a distinctive property of HyperCube: a bias toward full-rank, unitary solutions. This stands in contrast to the common low-rank bias in deep learning models, which is often associated with smooth interpolation between training examples (Arora et al., 2019; Jacot, 2023; Huh et al., 2023; Balzano et al., 2025). The full-rank, unitary nature of HyperCube solutions may indicate a novel form of inductive bias, potentially enabling more robust extrapolation beyond the training data.

The central open question is whether, when the data represents a group, all stationary points of the HyperCube objective necessarily satisfy the *perfect alignment condition* ($\mathcal{R} = 0$). Resolving this conjecture would close the theoretical gap, explaining how perfect alignment is the key mechanism that enforces the associativity axiom, thereby allowing the model to distinguish true group structures from other quasi-groups. Future work will pursue this question and explore broader links to nonconvex optimization, symmetry discovery, and representation learning.

References

Sanjeev Arora, Nadav Cohen, Wei Hu, and Yuping Luo. Implicit Regularization in Deep Matrix Factorization, October 2019. URL http://arxiv.org/abs/1905.13655. arXiv:1905.13655 [cs, stat].

Laura Balzano, Tianjiao Ding, Benjamin D Haeffele, Soo Min Kwon, Qing Qu, Peng Wang, Zhangyang Wang, and Can Yaras. An overview of low-rank structures in the training and adaptation of large models. *arXiv preprint arXiv:2503.19859*, 2025.

Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges, May 2021. URL http://arxiv.org/abs/2104.13478. arXiv:2104.13478 [cs, stat].

Dongsung Huh. Discovering group structures via unitary representation learning. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=Tz8Li6G2xU&noteId=Tz8Li6G2xU.

Minyoung Huh, Hossein Mobahi, Richard Zhang, Brian Cheung, Pulkit Agrawal, and Phillip Isola. The Low-Rank Simplicity Bias in Deep Networks, March 2023. URL http://arxiv.org/abs/2103.10427. arXiv:2103.10427 [cs].

Arthur Jacot. Implicit Bias of Large Depth Networks: a Notion of Rank for Nonlinear Functions, March 2023. URL http://arxiv.org/abs/2209.15055. arXiv:2209.15055 [cs, stat].

Alethea Power, Yuri Burda, Harri Edwards, Igor Babuschkin, and Vedant Misra. Grokking: Generalization Beyond Overfitting on Small Algorithmic Datasets, January 2022.

Appendix A. Equivalence to the Regular Representation

In the main text, we established that under the Perfect Alignment Conjecture, any global minimizer must consist of three families of unitary matrices. This appendix answers the next logical question: what is the precise nature of these unitary solutions? We show that they must form a representation of the group G, and furthermore, that all such solutions are equivalent to one canonical solution: the *left-regular representation*.

First, we provide a constructive proof that the left-regular representation is a valid solution that achieves the global minimum, certifying that the value $\mathcal{H}_{\min} = 3n^2$ is attainable.

Lemma 9 (The Regular Representation is a Minimizer) *Let $\rho_r(g)$ be the left-regular representation of G. The triple*

$$A_a = \rho_r(a), \qquad B_b = \rho_r(b), \qquad C_c = \rho_r(c)^\dagger \tag{9}$$

is a global minimizer of $\mathcal{H}$, achieving the value $3n^2$.

Proof The matrices $\rho_r(g)$ are unitary, so all norms are 1. We check the alignment equations (8):

$$A_a B_b = \rho_r(a)\rho_r(b) = \rho_r(a \circ b).$$

For the other side of the equation, $C^\dagger_{a\circ b} = (\rho_r(a \circ b)^\dagger)^\dagger = \rho_r(a \circ b)$, confirming alignment. Next, we check the constraint (1):

$$\frac{1}{n}\operatorname{Tr}(A_a B_b C_c) = \frac{1}{n}\operatorname{Tr}(\rho_r(a)\rho_r(b)\rho_r(c)^\dagger) = \frac{1}{n}\operatorname{Tr}(\rho_r(a \circ b \circ c^{-1})).$$

The character of the regular representation is $\operatorname{Tr}(\rho_r(g)) = n \cdot \mathbf{1}\{g = e\}$. Therefore, the trace is n if $a \circ b \circ c^{-1} = e$ (i.e., $a \circ b = c$) and 0 otherwise. Dividing by n confirms the constraint holds. Since this solution is composed of unitary matrices, it achieves the minimum value $\mathcal{H}_{\min} = 3n^2$. ■

Finally, we prove that the regular representation is, in essence, the only minimizer. The following theorem shows that any global minimizer can be transformed into the regular representation via a change of basis.

Theorem 10 (Global Minimizers are Unitarily Equivalent to ρ_r) *Let (A, B, C) be a global minimizer. There exist unitary matrices $U, V, W \in U(n)$ such that the transformed triple*

$$A'_g := U A_g V^\dagger, \qquad B'_g := V B_g W^\dagger, \qquad C'_g := W C_g U^\dagger$$

defines a unitary representation $\rho(g)$ of G where $A'_g = B'_g = (C'_g)^\dagger =: \rho(g)$. This representation ρ is unitarily equivalent to the left-regular representation ρ_r.

Proof Since (A, B, C) is a global minimizer, its matrices are unitary and satisfy the alignment equations $A_a B_b = C^\dagger_{a\circ b}$, etc. Let e be the identity of G. Choose the unitary matrices $U = A_e^\dagger$, $V = I$, and $W = B_e$.

1. Synchronization: We first show that $A'_g = B'_g = (C'_g)^\dagger$.

- $(C'_g)^\dagger = (B_e C_g A_e)^\dagger = A_e^\dagger C_g^\dagger B_e^\dagger$.
- $A'_g = A_e^\dagger A_g$. From alignment (8), $A_g B_e = C_g^\dagger$, so $A_g = C_g^\dagger B_e^\dagger$. Thus $A'_g = A_e^\dagger C_g^\dagger B_e^\dagger$.
- $B'_g = B_g B_e^\dagger$. From alignment (8), $C_g A_e = B_g^\dagger$, so $B_g = A_e^\dagger C_g^\dagger$. Thus $B'_g = A_e^\dagger C_g^\dagger B_e^\dagger$.
- Thus, $A'_g = B'_g = (C'_g)^\dagger$. Let's call this common matrix $\rho(g)$.

2. Homomorphism: We show $\rho(g)$ is a group homomorphism.

$$\begin{aligned}\rho(a)\rho(b) &:= A'_a B'_b = (U A_a V^\dagger)(V B_b W^\dagger) = U(A_a B_b)W^\dagger \\ &= U(C_{a\circ b}^\dagger)W^\dagger = (W C_{a\circ b} U^\dagger)^\dagger = (C'_{a\circ b})^\dagger \\ &= \rho(a \circ b).\end{aligned}$$

This confirms ρ is a group representation. Since the transform preserves unitarity, ρ is a unitary representation.

3. Character: We show ρ has the same character as ρ_r. The character is $\chi_\rho(g) = \mathrm{Tr}(\rho(g))$.

$$\chi_\rho(g) = \mathrm{Tr}(A'_g) = \mathrm{Tr}(A_e^\dagger A_g).$$

From the alignment equations with $a = e, b = e, c = e$, we have $A_e B_e = C_e^\dagger$. Since matrices are unitary, $A_e^\dagger = B_e C_e$. Substituting this:

$$\chi_\rho(g) = \mathrm{Tr}((B_e C_e)A_g) = \mathrm{Tr}(A_g B_e C_e).$$

Now we can use the constraint (1):

$$\frac{1}{n}\chi_\rho(g) = \frac{1}{n}\,\mathrm{Tr}(A_g B_e C_e) = \delta_{gee} = \mathbf{1}\{g = e\}.$$

Thus, the character is $\chi_\rho(g) = n \cdot \mathbf{1}\{g = e\}$. This is precisely the character of the left-regular representation. Since two representations with the same character are unitarily equivalent, ρ is equivalent to ρ_r. ■