ASSOCIATIVITY AS OCCAM'S RAZOR: DIFFERENTIABLE MEASURE OF ALGEBRAIC COMPLEXITY

Anonymous authors
Paper under double-blind review

ABSTRACT

Discovering discrete algebraic structures from data is a fundamental challenge in machine learning. However, standard deep learning models often struggle to learn generalizable algebraic rules, suggesting that their intrinsic inductive biases may be ill-suited for such tasks. Formalizing this intuition reveals a deeper theoretical issue: the literature currently lacks a precise, quantitative definition of algebraic simplicity. In this work, we resolve this gap by analyzing the global objective landscape of a canonical model for recovering discrete operation tables. We establish a lower bound for this objective and prove that attaining this bound entails a geometric alignment (collinearity) between the model's latent matrix representations.

Furthermore, this lower bound implements an inverse ℓ_2 penalty, which induces a full-rank unitarity bias—fundamentally opposing the standard low-rank bias in deep learning models. We establish that the objective over the feasible set is bounded below by an absolute floor determined solely by the target size. Attaining this absolute floor is structurally rigid, requiring the latent factorization to be perfectly unitary and collinear, which imposes a rigid algebraic homomorphism. Crucially, this homomorphism requirement transfers the associativity of matrix multiplication directly to the target operation. The global optimum $\mathcal{H}_{\inf}(\delta)$ exhibits a strict associativity gap, attaining the absolute minimum floor if and only if the target is a group isotope. Every global minimizer uniquely encodes the underlying group's left-regular representation (up to isotopy and unitary gauge). We empirically confirm that the minimum objective monotonically increases with the target's latent associativity violation ratio. These results formalize a novel differentiable inductive bias for deep learning—structural simplicity via latent associativity up to isotopy—offering a provably distinct alternative to low-dimensional representation. All foundational theorems are formalized in Lean 4.

1 INTRODUCTION

Discovering algebraic rules from data—especially group structure—is a fundamental challenge in machine learning. This capability is critical for advancing algorithmic reasoning, mathematical problem-solving, and logic-based tasks (Power et al., 2022; Veličković & Blundell, 2021; Lample & Charton, 2019), which fundamentally rely on discovering and applying exact discrete rules. Furthermore, it holds profound implications for Geometric Deep Learning (Bronstein et al., 2021): automating the discovery of underlying symmetry groups (Benton et al., 2020; Dehmamy et al., 2021; Finzi et al., 2021; Chau et al., 2020; Sanborn et al., 2023) can enable systematic generalization in domains where the true symmetry group is unknown.

We consider algebraic completion tasks——the recovery of an unknown binary operation from partial observations(Power et al., 2022)—as the canonical formalization for learning discrete algebraic structures. Standard deep learning models equipped with usual ℓ_2 regularization typically struggle to efficiently discover generalizing solutions in this regime, frequently exhibiting delayed generalization phenomena known as "grokking" (Power et al., 2022; Nanda et al., 2023; Gromov, 2023), suggesting that their intrinsic inductive biases may be ill-suited for such tasks.

While previous literature posits a heuristic notion that "simple" operations (e.g., modular addition) are easier to learn than complex ones (e.g., modular polynomials), a fundamental theoretical ques-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

tion remains unanswered: How do we quantitatively define, and actively optimize for, algebraic simplicity among symbolic, discrete operations?

A recent work demonstrated that a canonical factorization model equipped with a gradient-based regularizer, called HyperCube, induces a strong empirical bias toward recovering group operations (Huh, 2025). This bias yields significant improvements in sample efficiency and convergence speed over standard deep learning models in algebraic completion. However, the exact mathematical mechanism by which this continuous regularizer natively favors discrete group structure has remained entirely unknown.

Our Contributions. In this work, we close this gap by characterizing the global landscape optimum of this canonical model in the fully-observed regime. We establish the following:

- **Intrinsic Geometric-Algebraic Equivalence:** We derive an orthogonal decomposition of HyperCube's objective: $\mathcal{H} = \mathcal{B}_\delta + \mathcal{R}_\delta$ (Lemma 1), where $\mathcal{B}_\delta$ provides a lower-bound and a non-negative residual $\mathcal{R}_\delta \geq 0$ which penalizes geometric misalignment, driving the latent factors toward collinearity.
- **Rank-Maximizing Pressure:** Within the collinear set, the inverse ℓ_2 term $\mathcal{B}_\delta$ induces to an inverse-rank penalty, favoring full-rank unitary embeddings (Lemma 4). This fundamentally opposes the standard low-rank simplicity bias prevalent in deep learning models.
- **Absolute Feasible Bound and Homomorphism Rigidity:** We establish that the objective over the feasible set is bounded below by an absolute floor of $3\,|\delta|$—determined solely by the target size $|\delta|$ and invariant to its internal algebraic structure. Attaining this floor is structurally *rigid*, requiring the latent factorization to be perfectly unitary and collinear. Crucially, this geometric rigidity imposes an algebraic homomorphism, transferring the associativity of matrix multiplication directly to the target operation.
- **Global Optimality and Associativity Gap:** We establish that the absolute floor is attained *if and only if* δ is a group isotope, yielding the left-regular representation of the underlying group up to relabeling and unitary gauge. For non-group-isotope targets, the global optimum is strictly separated from the floor by a quantifiable associativity gap (Theorem 9). Furthermore, we empirically confirm that the converged minimum objective monotonically increases alongside the target's latent associativity violation ratio.
- **Mechanized Verification:** To ensure rigorous correctness across complex tensor trace equivalences and high-degree nonconvex constraints, all foundational theorems are formally verified in Lean 4.[1]

2 Related Work

Algebraic Generalization and Grokking. Standard neural networks typically struggle to learn discrete algorithmic and algebraic rules from raw data. In algebraic completion tasks, models often memorize the training data perfectly before experiencing delayed generalization, a phenomenon termed "grokking" (Power et al., 2022; Nanda et al., 2023; Gromov, 2023). While recent work attributes this delayed generalization to the slow emergence of structured, frequency-based representations under weight decay (Barak et al., 2022; Chughtai et al., 2023), the precise mathematical definition of a "simple" generalizing algebraic operation has remained ambiguous. Our work formally defines this simplicity via the associativity gap, proving that group isotopes occupy the absolute global minimum of our structural objective.

Symmetry Discovery in Geometric Deep Learning. Geometric Deep Learning (Bronstein et al., 2021) has traditionally relied on known symmetries (e.g., translation, rotation) hardcoded into equivariant network architectures. Recent efforts aim to automatically discover these underlying symmetry groups directly from data (Benton et al., 2020; Dehmamy et al., 2021; Finzi et al., 2021; Chau et al., 2020). However, enforcing discrete, non-differentiable group axioms (such as associativity) via continuous gradient descent remains a significant technical hurdle. By characterizing a differentiable measure that natively penalizes associativity violations, our theoretical landscape provides a

[1] https://anonymous.4open.science/r/HyperCube_Lean-4664/README.md

foundational mechanism for automated symmetry discovery, unconstrained by predefined architectural priors.

Matrix Completion and Implicit Regularization. Our setting is structurally analogous to classical matrix completion (Candès & Recht, 2009; Candes & Tao, 2010), which formalizes the recovery of missing entries via structural priors. In deep learning, gradient descent on overparameterized factorizations is famously known to exhibit an implicit regularization toward low-rank solutions (Gunasekar et al., 2017; Arora et al., 2019). However, Cayley tables of quasigroups are Latin squares, which are intrinsically full-rank combinatorial designs. Consequently, the standard implicit low-rank bias of deep matrix factorization fundamentally opposes the structural requirements of algebraic completion, necessitating an entirely different bounding mechanism.

Simplicity Biases: Low-Rank vs. Full-Rank. A central tenet of deep learning theory is that models generalize by favoring low-complexity solutions, traditionally measured via linear rank or norm bounds (Huh et al., 2023; Aghajanyan et al., 2021). However, as established above, low-rank representations are fundamentally ill-suited for encoding exact, bijective discrete algebraic structures. We prove that HyperCube's regularizer induces an inverse ℓ_2 penalty (Lemma 4) that operates as a rank-maximizing, *full-rank bias*, favoring balanced, unitary representations. To our knowledge, this provides one of the first formal characterizations of a native, gradient-based structural simplicity bias that actively maximizes representation rank to capture discrete algebraic axioms.

Tensor Factorization and the HyperCube Model. This work provides the rigorous theoretical foundation for the empirical phenomena observed in the HyperCube tensor factorization architecture (Huh, 2025). While Huh (2025) introduced the objective and empirically demonstrated its sample efficiency and tendency to recover group isotopes, the underlying mathematical mechanics determining *why* it succeeds were unsolved. We resolve the core open conjectures of the original framework, formally proving the $3|\delta|$ global lower bound and structurally characterizing its equality cases via the left-regular representation.

3 Theoretical Framework

3.1 Algebraic Structure Tensor

Let $(Q, \circ)$ denote a finite set of n elements equipped with a binary operation $\circ : Q \times Q \to Q$. Its operation table is captured by the structure tensor $\delta \in \{0,1\}^{n\times n\times n}$, where $\delta_{abc} \coloneqq \mathbb{I}_{\{a\circ b=c\}}$ for $a, b, c \in Q$. We denote the table size by $|\delta| \coloneqq \sum_{abc} \delta_{abc} = n^2$.

Groups and quasigroups To formalize the discovery of associative structures, we restrict our scope to quasigroups, where, the equation $a \circ x = c$ (and also $x \circ b = c$) has a unique solution x for every $a, b, c \in Q$. A group can be defined as a quasigroup that additionally satisfies associativity $(a \circ b) \circ c = a \circ (b \circ c)$.

A triple (a, b, c) is a supported triple if $\delta_{abc} = 1$. Geometrically, these triples form the hyperedges of a support hypergraph. For quasigroups, this hypergraph is connected, ensuring that local algebraic constraints propagate globally.

3.2 HyperCube Model

Notations. For $X, Y \in \mathbb{C}^{n\times n}$, we define the *normalized* Frobenius inner product as $\langle X, Y\rangle \coloneqq \frac{1}{n}\operatorname{Tr}(X^\dagger Y)$ and the associated norm as $\|X\|^2 \coloneqq \langle X, X\rangle$. This normalization ensures $\|U\|^2 = 1$ for any unitary U. Throughout, † denotes the conjugate transpose.

Operator-Valued Factorization. HyperCube defines a model[2] $T : Q^3 \to \mathbb{C}$ as a trilinear product of three factor tensors $\Theta \coloneqq (A, B, C) \in (\mathbb{C}^{n\times n\times n})^3$:

$$T_{abc}(\Theta) \coloneqq \frac{1}{n}\operatorname{Tr}(A_a B_b C_c), \tag{1}$$

[2]Alternatively, this model can be interpreted as a mapping $T : Q^2 \to \mathbb{C}^n$ that produces a vector of logits over the output index c.

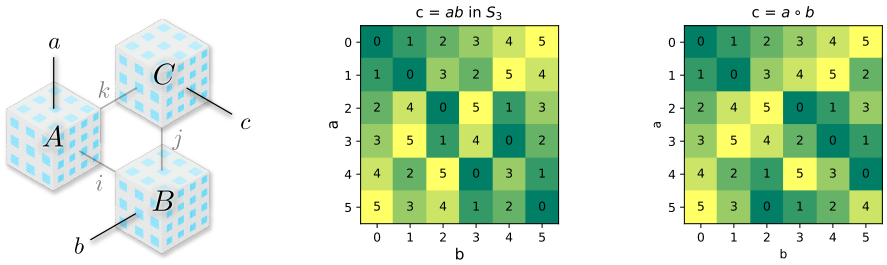


Figure 1: (Left) Illustration of HyperCube product (Reproduced from Huh (2025)). (Middle) S_3 group table. (Right) A non-group isotope table.

where A_a, B_b, and C_c denote the $n \times n$ matrix-embeddings (linear operators) of $a, b, c \in Q$.

Optimization Objective. HyperCube minimizes an objective that penalizes the squared norm of the output gradient with respect to the parameters (*e.g.*, $\partial T_{abc}/\partial A_a = (B_b C_c)^\dagger$):

$$\mathcal{H}(\Theta) = \sum_{a,b,c \in Q} \delta_{abc} \big(\|B_b C_c\|^2 + \|C_c A_a\|^2 + \|A_a B_b\|^2 \big). \tag{2}$$

We characterize its global minimum under the exact target reconstruction constraint, $T(\Theta) = \delta$,

$$\mathcal{H}_{\inf}(\delta) = \inf_{\Theta \in \mathcal{F}_\delta} \mathcal{H}(\Theta), \tag{3}$$

which defines the model's intrinsic **measure of algebraic complexity** for a given target δ.

3.3 Formal Definitions

Definition 1 (Feasible Set)**.** *The feasible set $\mathcal{F}_\delta$ consists of all parameter triples Θ that exactly reconstruct the target tensor δ: $\mathcal{F}_\delta := \{\Theta \mid T(\Theta) = \delta\}$.*

Definition 2 (Unitary Set)**.** *The unitary set $\mathcal{U}$ is the collection of parameters Θ where every constituent matrix slice is unitary, satisfying $A_a A_a^\dagger = B_b B_b^\dagger = C_c C_c^\dagger = I_n$ for all $a, b, c \in Q$.*

Definition 3 (Isotopy)**.** *Two operation tensors δ and δ' are isotopic if they are equivalent up to independent permutations i.e. $\exists (\phi, \psi, \chi) \in S_n^3$: such that $\delta'_{abc} = \delta_{\phi(a)\psi(b)\chi(c)}$.*

3.4 Intrinsic Model Symmetries

HyperCube exhibits symmetries that govern its optimization landscape and algebraic equivalence.

Gauge Invariance. The HyperCube product (1) is invariant under the continuous group of invertible *gauge transformations*. For any invertible matrices $U, V, W \in \mathrm{GL}(n, \mathbb{C})$, the transformation

$$\Theta' = (U A_a V^{-1},\ V B_b W^{-1},\ W C_c U^{-1}) \tag{4}$$

preserves the product exactly: $T_{abc}(\Theta') = T_{abc}(\Theta)$. The objective (2) is invariant under *unitary gauge transformations* only: i.e. where U, V, W are unitary matrices.

Isotopy Equivariance. The HyperCube product (1) is equivariant under *isotopy* (Definition 3). For any bijections $(\phi, \psi, \chi) \in S_n^3$, permuting the parameters as $\Theta' = (A_{\phi(\cdot)}, B_{\psi(\cdot)}, C_{\chi(\cdot)})$ implies $T_{abc}(\Theta') = T_{\phi(a)\psi(b)\chi(c)}(\Theta)$. Consequently, the scalar objective (2) is *isotopy-invariant*. This symmetry greatly simplifies the subsequent analysis by allowing us to reduce any quasigroup target to its loop isotope (a quasigroup with a designated identity e) without loss of generality.

4 Orthogonal Decomposition and Geometric Alignment

HyperCube's quartic objective under trilinear reconstruction constraints presents a formidable non-convex optimization problem. Direct stationarity analysis is complicated by the high-degree non-convex constraints. Instead, we employ a direct bounding strategy: Decomposing the objective into a lower bound and a non-negative residual identifies the algebraic optimality conditions.

Lemma 1 (Decomposition of $\mathcal{H}$)**.** *For any Θ and target δ, the objective* (2) *decomposes as:*

$$\mathcal{H}(\Theta) = \mathcal{B}_\delta(\Theta) + \mathcal{R}_\delta(\Theta), \tag{5}$$

where $\mathcal{B}_\delta$ is the inverse-scale term (i.e. inverse ℓ_2 penalty):

$$\mathcal{B}_\delta(\Theta) \coloneqq \sum_{a,b,c} \delta_{abc} \, |T_{abc}(\Theta)|^2 \left(\frac{1}{\|A_a\|^2} + \frac{1}{\|B_b\|^2} + \frac{1}{\|C_c\|^2} \right), \tag{6}$$

and $\mathcal{R}_\delta(\Theta) \coloneqq \mathcal{H}(\Theta) - \mathcal{B}_\delta(\Theta) \geq 0$ is the misalignment residual. Consequently, $\mathcal{H}(\Theta) \geq \mathcal{B}_\delta(\Theta)$, with equality if and only if *the factors satisfy the **collinear identities** for every supported triple:*

$$B_b C_c = T_{abc} \frac{A_a^\dagger}{\|A_a\|^2}, \quad C_c A_a = T_{abc} \frac{B_b^\dagger}{\|B_b\|^2}, \quad A_a B_b = T_{abc} \frac{C_c^\dagger}{\|C_c\|^2}. \tag{7}$$

Proof Sketch. By the Cauchy–Schwarz inequality, the inner product $T_{abc} = \langle A_a^\dagger, B_b C_c \rangle$ yields the lower bound $\|B_b C_c\|^2 \geq |T_{abc}|^2 / \|A_a\|^2$. Summing this symmetric bound across all three gradient terms and over the target support directly establishes $\mathcal{H}(\Theta) \geq \mathcal{B}_\delta(\Theta)$, guaranteeing that the residual $\mathcal{R}_\delta(\Theta)$ is non-negative. Equality holds *if and only if* the Cauchy–Schwarz bound saturates for every supported triple, which requires the matrices in each inner product to be collinearly aligned, yielding exactly (7). The full proof is provided in Appendix B. □

Definition 4 (Collinear Set)**.** *The collinear set is defined as the zero-set of the misalignment residual:*

$$\mathcal{M}_\delta \coloneqq \{\Theta \mid \mathcal{R}_\delta(\Theta) = 0\}.$$

By Lemma 1, $\Theta \in \mathcal{M}_\delta$ if and only if its matrix-embeddings satisfy (7) *for every supported triple.*

We analyze the spectral structure of the collinear set $\mathcal{M}_\delta$. This set exhibits structural rigidity arising from the propagation of local alignment constraints across the target's support. Throughout this section, we assume $T_{abc} \neq 0$ for all supported triples.

Lemma 2 (Shared Gram Matrices)**.** *For any $\Theta \in \mathcal{M}_\delta$, there exist index-independent, normalized positive semi-definite (PSD) Gram matrices $X, Y, Z \in \mathbb{C}^{n \times n}$ such that for all supported triples*

$$X = \frac{A_a A_a^\dagger}{\|A_a\|^2} = \frac{C_c^\dagger C_c}{\|C_c\|^2}, \quad Y = \frac{B_b B_b^\dagger}{\|B_b\|^2} = \frac{A_a^\dagger A_a}{\|A_a\|^2}, \quad Z = \frac{C_c C_c^\dagger}{\|C_c\|^2} = \frac{B_b^\dagger B_b}{\|B_b\|^2}. \tag{8}$$

The trace of these matrices is constant: $\mathrm{Tr}(X) = \mathrm{Tr}(Y) = \mathrm{Tr}(Z) = n$.

Proof Sketch. Substituting the collinear identities (7) into $A_a(B_b C_c) = (A_a B_b) C_c$ (i.e., associativity of matrix multiplication) and canceling the nonzero scalar T_{abc} isolates the relation $A_a A_a^\dagger / \|A_a\|^2 = C_c^\dagger C_c / \|C_c\|^2$. The left side depends only on a, while the right side depends only on c. Because the quasigroup support hypergraph is connected, this local equality propagates globally across all indices, implying both sides equal an index-independent constant matrix X. □

Lemma 3 (Normalized Rank κ)**.** *For any $\Theta \in \mathcal{M}_\delta$ with $T_{abc} \neq 0$ on the support, the dimensionless ratio $\kappa_{abc} \coloneqq \|A_a\|^2 \|B_b\|^2 \|C_c\|^2 / |T_{abc}|^2$ is constant across the support. This constant κ enforces the strict projection identity $X = \kappa X^2$, thereby identifying κ as the **normalized rank**:*

$$\kappa = \frac{\mathrm{rank}(X)}{n} \leq 1 \tag{9}$$

Moreover, $\kappa = 1$ if and only if *the Gram matrices are identity matrices ($X = Y = Z = I_n$).*

Proof Sketch. Substituting the collinear identities (7) into the definition of the Gram matrix X yields the relation $X = \kappa_{abc} X^2$ for any supported triple. Since X is already proven to be index-independent (Lemma 2), the coefficient κ_{abc} must be a constant κ. Defining $P \coloneqq \kappa X$ yields $P^2 = \kappa^2 X^2 = \kappa X = P$, making P an orthogonal projection. This forces its eigenvalues into the **discrete set** $\{0, 1\}$. Thus, $\text{rank}(X) = \text{rank}(P) = \text{Tr}(P) = \kappa \,\text{Tr}(X) = \kappa n$. Rearranging yields $\kappa = \text{rank}(X)/n \leq 1$. If $\kappa = 1$, X is a trace-n projection, strictly implying $X = I_n$. □

Lemma 4 (Scalar AM-GM Bound and Inverse-rank Penalty (Collinear Set))**.** *For any $\Theta \in \mathcal{M}_\delta$, the objective $\mathcal{H}(\Theta)$ is bounded below by*

$$\mathcal{H}(\Theta) = \mathcal{B}_\delta(\Theta) \geq \kappa^{-1/3} \cdot 3 \sum_{a,b,c} \delta_{abc} |T_{abc}|^{4/3}, \tag{10}$$

where the inequality is tight if and only if *all matrix-embedding norms are equal. ($\|A_a\|^2 = \|B_b\|^2 = \|C_c\|^2$). Equality with the following term additionally requires $\kappa = 1$ (full rank).*

$$\mathcal{H}(\Theta) \geq \mathcal{H}^*_\delta(T(\Theta)) \coloneqq 3 \sum_{a,b,c} \delta_{abc} |T_{abc}|^{4/3}. \tag{11}$$

Proof Sketch. For $\Theta \in \mathcal{M}_\delta$, $\mathcal{H}(\Theta) = \mathcal{B}_\delta(\Theta)$. Let $\alpha_a \coloneqq \|A_a\|^{-2}$, $\beta_b \coloneqq \|B_b\|^{-2}$, $\gamma_c \coloneqq \|C_c\|^{-2}$. Note that $\alpha_a \beta_b \gamma_c = \kappa^{-1} |T_{abc}|^{-2}$ (Lemma 3). Applying the ***scalar* AM–GM** inequality ($\alpha + \beta + \gamma \geq 3(\alpha\beta\gamma)^{1/3}$) to the definition of $\mathcal{B}_\delta(\Theta)$ (6) yields the lower bound, which saturates *if and only if* $\alpha_a = \beta_b = \gamma_c$ is satisfied for every supported triple (a, b, c). Due to the connectivity of the quasigroup support graph (Appendix A.1), these local equalities propagate globally, ensuring all matrix-embedding norms are balanced. □

While Lemma 4 establishes an inverse-rank penalty, and a lower bound, they are restricted within the collinear set $\mathcal{M}_\delta$, where perfect geometric alignment allows for a reduction to scalar inequalities. Evaluating the broader optimization landscape—and identifying when collinear states attain the global lower bound—require a bounding mechanism that makes no assumptions about alignment.

5 Global Optimality and Associativity Gap

The following matrix inequality extends the scalar AM–GM inequality. Crucially, it applies directly to arbitrary matrices without assuming commutativity or collinearity.

Lemma 5 (Matrix AM-GM)**.** *For all $X, Y, Z \in \mathbb{C}^{n \times n}$,*

$$\|XY\|^2 + \|YZ\|^2 + \|ZX\|^2 \geq 3 \left| \frac{1}{n} \text{Tr}(XYZ) \right|^{4/3}, \tag{12}$$

which saturates if and only if X, Y, Z are scaled unitarities that satisfy $XX^\dagger = YY^\dagger = ZZ^\dagger = \tau^2 I_n$ with $\tau \neq 0$ and $XYZ = \tau^3 I$.

Proof Sketch. Define a block-cyclic matrix $M \in \mathbb{C}^{3n \times 3n}$ and evaluate its products:

$$M \coloneqq \begin{pmatrix} 0 & X & 0 \\ 0 & 0 & Y \\ Z & 0 & 0 \end{pmatrix}, \qquad M^2 = \begin{pmatrix} 0 & 0 & XY \\ YZ & 0 & 0 \\ 0 & ZX & 0 \end{pmatrix}, \qquad M^3 = \begin{pmatrix} XYZ & 0 & 0 \\ 0 & YZX & 0 \\ 0 & 0 & ZXY \end{pmatrix}.$$

Evaluating the normalized norm of M^2 and the normalized trace of M^3 yields (via trace cyclicity):

$$\frac{1}{3n} \text{Tr}\left(M^{2\dagger} M^2\right) = \frac{1}{3} \left(\|XY\|^2 + \|YZ\|^2 + \|ZX\|^2\right), \qquad \frac{1}{3n} \text{Tr}(M^3) = \frac{1}{n} \text{Tr}(XYZ).$$

Applying the Spectral Trace-Frobenius Bound (Lemma 14 in Appendix C) to M yields the inequality (12) and its equality condition. □

5.1 Absolute Lower Bound

We establish that the *feasible* optimization landscape is governed by an absolute lower bound determined solely by the target table size $|\delta|$, invariant to the target's internal algebraic structure. However, attaining this floor is structurally *rigid*, requiring a unitary collinear factorization. To show this, we first generalize Lemma 4 beyond the collinear set $\mathcal{M}_\delta$ to establish the following bound.

Theorem 6 (Dynamic Unconstrained Bound)**.** *For any quasigroup target δ and any Θ, the objective $\mathcal{H}(\Theta)$ is bounded below by the dynamic floor $\mathcal{H}^*_\delta(T(\Theta))$:*

$$\mathcal{H}(\Theta) \geq \mathcal{H}^*_\delta(T(\Theta)). \tag{13}$$

The bound holds even when some supported outputs vanish. Its equality characterization on the feasible set is given in Theorem 7.

Theorem 7 (Absolute Feasible Bound)**.** *For any quasigroup target δ and any feasible factorization $\Theta \in \mathcal{F}_\delta$, the objective $\mathcal{H}(\Theta)$ is bounded below by the absolute floor $3|\delta|$:*

$$\mathcal{H}(\Theta) \geq \mathcal{H}_{\inf}(\delta) \geq 3\,|\delta|. \tag{14}$$

Equality holds throughout if and only if Θ *is a unitary collinear factorization of δ ($\mathcal{F}_\delta \cap \mathcal{M}_\delta \cap \mathcal{U}$).*

Proof of Theorems 6 and 7. Lemma 5 establishes a local lower bound for any triple (a, b, c):

$$\|A_a B_b\|^2 + \|B_b C_c\|^2 + \|C_c A_a\|^2 \geq 3\,|T_{abc}|^{4/3}. \tag{15}$$

Since $\delta_{abc} \in \{0, 1\}$, multiplying (15) by δ_{abc} and summing over all entries yields the unconditional global lower bound (13). Equality holds *if and only if* (15) saturates for every supported triple (where $\delta_{abc} = 1$). On the feasible set, $T_{abc} = 1$ for each supported triple, so the equality conditions of Lemma 5 apply: the matrix-embeddings are unitary and $A_a B_b C_c = I_n$.

Restricting the landscape to the feasible set $\mathcal{F}_\delta$ (i.e., $T = \delta$) reduces the global bound to $\mathcal{H}^*_\delta(\delta) = 3\,|\delta|$ and the equality condition to $A_a B_b C_c = I_n$. This constrains the matrix-embeddings to be unitary ($\mathcal{U}$) that perfectly satisfy collinearity ($\mathcal{M}_\delta$). Thus, the absolute floor is attained *if and only if* $\Theta \in \mathcal{F}_\delta \cap \mathcal{M}_\delta \cap \mathcal{U}$. □

5.2 Collinearity–Associativity Equivalence

Here, we establish the Collinearity–Associativity Equivalence for the case of unitary factorizations.

Theorem 8 (Unitary Collinearity $\iff$ Group Isotope)**.** *For any finite quasigroup target δ, the **unitary** feasible collinear set ($\mathcal{F}_\delta \cap \mathcal{M}_\delta \cap \mathcal{U}$) is non-empty* if and only if *δ is isotopic to a group.*

Proof Sketch. (**Necessity** $\implies$): By isotopy equivariance, we can assume without loss of generality that δ is a loop (a quasigroup with an identity element e). A unitary collinear factorization admits a unitary gauge transformation $U, V, W \in U(n)$ that synchronizes the matrix-embeddings: $A'_g = B'_g = (C'_g)^\dagger := \rho(g)$ (Lemma 10), which enforces homomorphism $\rho(a)\rho(b) = \rho(a \circ b)$ via collinear identities (7): $A'_a B'_b = (C'_{a \circ b})^\dagger$. Also, feasibility ($T = \delta$) forces this map to be injective (Lemma 11). Consequently, the quasigroup operation $\circ$ directly inherits the associativity of matrix multiplication

$$\rho((x \circ y) \circ z) = \rho(x)\rho(y)\rho(z) = \rho(x \circ (y \circ z)).$$

(**Sufficiency** $\impliedby$): If δ is a group isotope, its left-regular representation defines a unitary collinear factorization that achieves $T = \delta$ and $\mathcal{R} = 0$ (Lemma 13). (See Appendix B.2 for the full proof). □

5.3 Global Optimality for Group and Non-Group Targets

The global bound identifies the geometry required at equality. Combining it with the characterization of feasible collinearity now determines which targets realize the floor and which representations occur there.

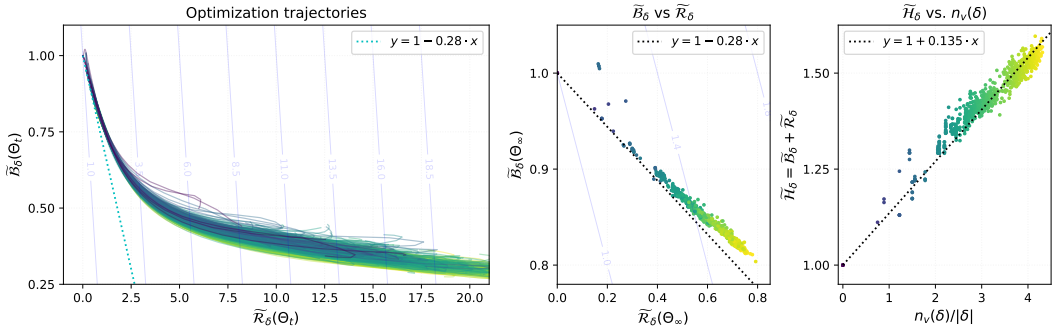


Figure 2: **(Left)** Optimization trajectories of $\tilde{\mathcal{B}}_\delta(\Theta_t)$ vs $\tilde{\mathcal{R}}_\delta(\Theta_t)$ from random initialization. **(Middle)** Numerical minima components ($\tilde{\mathcal{B}}_\delta(\Theta_\infty)$ vs $\tilde{\mathcal{R}}_\delta(\Theta_\infty)$) for each target. **(Right)** Total converged minimum objective $\tilde{\mathcal{H}}_\delta(\Theta_\infty)$ plotted against the target's associativity violation. **Optimization trajectory of objective components.** ($\tilde{\mathcal{B}}_\delta \coloneqq \mathcal{B}_\delta/\mathcal{H}^*_\delta$, $\tilde{\mathcal{R}}_\delta \coloneqq \mathcal{R}_\delta/\mathcal{H}^*_\delta$) **Left:** Optimization trajectories of $\widetilde{\mathcal{B}}_\delta$ against $\widetilde{\mathcal{R}}_\delta$. **Middle:** Converged points, denoted Θ_∞. **Right:** attained $\widetilde{\mathcal{H}}_\delta = \widetilde{\mathcal{B}}_\delta + \widetilde{\mathcal{R}}_\delta$ against $\nu(\delta)$. All objective components are normalized by the dynamic floor $\mathcal{H}^*_\delta(T(\Theta))$. Points and trajectories are color-coded by the target's normalized associativity violation counts ($\nu(\delta)$). Contours (Left, Middle) show the sum of the two normalized components visualizing the underlying objective landscape: $\tilde{\mathcal{H}}_\delta \coloneqq \tilde{\mathcal{B}}_\delta + \tilde{\mathcal{R}}_\delta$; dotted lines are empirical references, not theoretical bounds. Training details appear in Appendix D.1.

Theorem 9 (Global Optimality and Associativity Gap)**.** *For any quasigroup target δ, the global optimality of the objective $\mathcal{H}$ exhibit a strict dichotomy:*

1. ***Group Isotopes:*** *If δ is isotopic to a group, then its global minimum attains the absolute floor: $\mathcal{H}_{\inf}(\delta) = 3\,|\delta|$. Furthermore, every global minimizer is equivalent to the left-regular representation of the underlying group, up to isotopy and unitary gauge.*

2. ***Non-Group Isotopes:*** *Otherwise, the global optimum is separated from the absolute floor by a positive associativity gap: $\mathcal{H}_{\inf}(\delta) > 3\,|\delta|$.*

Proof. The absolute bound and its equality characterization follow from Theorem 7. For a group isotope, Lemma 13 supplies a unitary collinear factorization at the floor, and Lemma 12 identifies it up to the stated equivalences. For any other target, Theorem 8 rules out feasible unitary collinear factorizations. Since attaining the absolute floor requires such a factorization, the global optimum is strictly separated from the floor: $\mathcal{H}_{\inf}(\delta) > 3|\delta|$. □

6 Empirical Analysis: Learning Dynamics and the Associativity Gap

To empirically validate our theoretical framework, we investigate the numerical optimum solutions of the model objective across a diverse population of finite quasigroups. Specifically, we optimize the parameters Θ over the unconstrained space by minimizing the regularized reconstruction loss, $\mathcal{L}(\Theta) = \|T(\Theta) - \delta\|_F^2 + \lambda\mathcal{H}(\Theta)$, using full-batch gradient descent. The target population comprises all reduced loop tables of orders 5 through 8 (see Appendix D.3 for complete details). To encourage structural discovery without artificially constraining the final representation, we apply a two-phase regularization schedule: $\lambda = 0.05$ for the initial 500 epochs, followed by $\lambda = 0$ to allow unhindered convergence to the exact reconstruction manifold.

Tracking the Objective Landscape. To isolate the structural bias from transient fluctuations in approximation fidelity throughout training, we track the dimensionless objective components: the inverse-scale penalty $\widetilde{\mathcal{B}}_\delta \coloneqq \mathcal{B}_\delta/\mathcal{H}^*_\delta(T(\Theta))$ and the misalignment residual $\widetilde{\mathcal{R}}_\delta \coloneqq \mathcal{R}_\delta/\mathcal{H}^*_\delta(T(\Theta))$. Dividing by the dynamic floor (Theorem 6) provides stable normalization for these terms.

Latent Associativity Violation. Because the objective $\mathcal{H}_{\inf}$ is invariant under isotopy, we evaluate targets based on their most associative isotopic representative. Let $\mathfrak{L}(\delta)$ denote the finite set of reduced loop tables isotopic to a target δ. We define the *latent associativity violation ratio* as:

$$\nu(\delta) \coloneqq \min_{(Q,\star)\in\mathfrak{L}(\delta)} \frac{1}{|\delta|} \sum_{a,b,c\in Q} \mathbb{I}_{\{(a\star b)\star c \neq a\star(b\star c)\}}. \tag{16}$$

This metric is isotopy-invariant and vanishes if and only if δ is a group isotope.

Optimization Dynamics and Geometric Obstruction. Empirically, the optimization trajectories exhibit smooth, stable convergence (Figure 2, Left). During training, the models navigate a direct geometric trade-off: reducing the misalignment residual ($\widetilde{\mathcal{R}}_\delta$) structurally forces an increase in the inverse-scale penalty ($\widetilde{\mathcal{B}}_\delta$). Because the slope of this observed trade-off is approximately -0.28—which is strictly shallower than -1—gradient updates that minimize geometric misalignment consistently guarantee a net decrease in the total normalized objective $\widetilde{\mathcal{H}}_\delta = \widetilde{\mathcal{B}}_\delta + \widetilde{\mathcal{R}}_\delta$. However, for non-associative targets, the mathematical impossibility of a unitary collinear factorization yields a strict geometric obstruction for the misalignment term. Rather than diverging, the optimizer is forced to halt at a stationary point where this structural barrier prevents any further reduction in $\widetilde{\mathcal{R}}_\delta$.

Convergence and the Associativity Gap. Upon convergence, the attained optima strictly obey the dichotomy established in Theorem 9 (Figure 2, Middle and Right). For group-isotope targets ($\nu(\delta) = 0$), the trajectories successfully eliminate misalignment, converging to the theoretical floor $(\widetilde{\mathcal{R}}_\delta, \widetilde{\mathcal{B}}_\delta) = (0, 1)$. Conversely, for non-group isotopes, the unitary collinear set is mathematically empty; consequently, the optimizer halts at Pareto-stationary points characterized by strictly positive residual misalignment ($\widetilde{\mathcal{R}}_\delta > 0$). Furthermore, across the experimental population, the converged total objective $\widetilde{\mathcal{H}}_\delta$ monotonically increases with the latent associativity violation ratio $\nu(\delta)$. This confirms that the continuous optimization landscape successfully operationalizes the associativity gap, inherently penalizing non-associative algebraic structures.

7 Conclusion

By characterizing the global optimum landscape of the HyperCube architecture, this work establishes a rigorous mathematical framework for algebraic completion and formally resolves the open conjecture of Huh (2025). Crucially, our equality characterization identifies the **associativity gap** as a strict mathematical boundary separating group isotopes from non-associative structures. Rather than merely memorizing arbitrary tables, the continuous objective actively repels non-groups by forcing them to absorb a permanent, theoretically bounded geometric misalignment penalty.

More broadly, this establishes a foundational bridge between continuous optimization and the discovery of discrete algebraic structures. While traditional neurosymbolic and program induction frameworks often rely on discrete search strategies that can be combinatorially brittle, our results demonstrate that a continuous optimization landscape can natively encode exact algebraic axioms through its equality cases. Furthermore, our empirical validations enrich this theoretical dichotomy: across the tested targets, the attained minima strictly increase alongside the latent associativity violation ratio. This association indicates that the objective provides a graded, empirical measure of algebraic complexity extending beyond our exact equality characterization.

Limitations and Future Work. While our theoretical landscape guarantees are mathematically rigid, the current analysis and empirical validations are focused on exact reconstruction of finite, fully observed discrete targets. Extending this theoretical characterization to partial or noisy observations, establishing formal recovery rates, and analyzing the convergence guarantees of gradient-based optimizers within this landscape remain important open challenges. Furthermore, scaling this differentiable measure to high-dimensional, real-world continuous data is a critical next step. Ultimately, as the machine learning community increasingly seeks models capable of reliable symbolic abstraction, these findings illustrate how the geometry of a continuous objective can successfully represent and enforce formal algorithmic structure.

AI USE STATEMENT

All research ideas, mathematical claims, and proofs in this paper were developed by the authors without generative AI. We used generative AI tools for two purposes. First, to help translate our proofs into the Lean 4 formalization in `https://anonymous.4open.science/r/HyperCube_Lean-4664/README.md`. Second, to polish the language, editing, and formatting of parts of the manuscript; these edits were reviewed by the authors and introduced no technical content.

REFERENCES

Armen Aghajanyan, Sonal Gupta, and Luke Zettlemoyer. Intrinsic dimensionality explains the effectiveness of language model fine-tuning. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: long papers)*, pp. 7319–7328, 2021.

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].

Boaz Barak, Benjamin L. Edelman, Surbhi Goel, Sham Kakade, Eran Malach, and Cyril Zhang. Hidden progress in deep learning: SGD learns parities near the computational limit. In *Advances in Neural Information Processing Systems*, volume 35, pp. 21750–21764. Curran Associates, Inc., 2022. URL `https://proceedings.neurips.cc/paper_files/paper/2022/hash/884baf65392170763b27c914087bde01-Abstract-Conference.html`.

Gregory Benton, Marc Finzi, Pavel Izmailov, and Andrew Gordon Wilson. Learning invariances in neural networks from training data. In *Advances in neural information processing systems*, volume 33, pp. 17605–17616, 2020.

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].

Emmanuel J. Candes and Terence Tao. The Power of Convex Relaxation: Near-Optimal Matrix Completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, May 2010. ISSN 0018-9448, 1557-9654. doi: 10.1109/TIT.2010.2044061. URL `http://ieeexplore.ieee.org/document/5452187/`.

Emmanuel J. Candès and Benjamin Recht. Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics*, 9(6):717–772, December 2009. ISSN 1615-3383. doi: 10.1007/s10208-009-9045-5. URL `https://doi.org/10.1007/s10208-009-9045-5`.

Ho Yin Chau, Frank Qiu, Yubei Chen, and Bruno Olshausen. Disentangling images with lie group transformations and sparse coding. *arXiv preprint arXiv:2012.12071*, 2020.

Bilal Chughtai, Lawrence Chan, and Neel Nanda. Neural Networks Learn Representation Theory: Reverse Engineering how Networks Perform Group Operations. In *ICLR 2023 Workshop on Physics for Machine Learning*, March 2023.

Nima Dehmamy, Robin Walters, Yanchen Liu, Dinghuai Wang, and Rose Yu. Automatic symmetry discovery with lie algebra convolutional network. In *Advances in Neural Information Processing Systems*, volume 34, pp. 2503–2515, 2021.

Marc Finzi, Max Welling, and Andrew Gordon Wilson. A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups. In *International conference on machine learning*, pp. 3318–3328. PMLR, 2021.

Andrey Gromov. Grokking modular arithmetic. *arXiv preprint arXiv:2301.02679*, 2023.

Suriya Gunasekar, Blake E Woodworth, Srinadh Bhojanapalli, Behnam Neyshabur, and Nati Srebro. Implicit Regularization in Matrix Factorization. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://papers.nips.cc/paper/2017/hash/58191d2a914c6dae66371c9dcdc91b41-Abstract.html.

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].

Guillaume Lample and François Charton. Deep learning for symbolic mathematics. In *International Conference on Learning Representations*, 2019.

Neel Nanda, Lawrence Chan, Tom Lieberum, Jess Smith, and Jacob Steinhardt. Progress measures for grokking via mechanistic interpretability. In *International Conference on Learning Representations*, 2023.

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

Sophia Sanborn, Christian Shewmake, Bruno Olshausen, and Christopher Hillar. Bispectral Neural Networks, May 2023. URL http://arxiv.org/abs/2209.03416. arXiv:2209.03416 [cs].

Petar Veličković and Charles Blundell. Neural algorithmic reasoning. *Patterns*, 2(7):100273, 2021.

A Formal Definitions and Auxiliary Setup

A.1 Support Hypergraphs and Connectivity

To facilitate the understanding of the topological structure of algebraic operations within the HyperCube framework, we provide the following formal definitions.

Support of a 3rd-Order Tensor Given a 3rd-order binary target tensor $\delta \in \{0,1\}^{n\times n\times n}$ representing a binary operation $a \circ b = c$, the **support** of δ is defined as the set of indices corresponding to non-zero entries:

$$\text{supp}(\delta) = \{(a,b,c) \in [n]^3 \mid \delta_{abc} = 1\}$$

Support Hypergraph The support hypergraph $H_\delta = (V, E)$ is a 3-uniform hypergraph constructed from $\text{supp}(\delta)$ such that:

- $V = \{1, 2, \ldots, n\}$ is the set of vertices, representing the elements of the algebra.
- $E = \{(a,b,c) \mid (a,b,c) \in \text{supp}(\delta)\}$ is the set of hyperedges, recorded as ordered triples representing the relationship $a \circ b = c$.

Hypergraph Connectedness Connectedness in H_δ determines the global reachability of elements through the binary operation.

- **Adjacency:** Two hyperedges $e_i, e_j \in E$ are adjacent if they share at least one vertex ($e_i \cap e_j \neq \emptyset$).
- **Path:** A path between two vertices $u, v \in V$ is a sequence of vertices and hyperedges $v_0, e_1, v_1, e_2, \ldots, e_k, v_k$ such that $v_0 = u$, $v_k = v$, and for each i, $\{v_{i-1}, v_i\} \subseteq e_i$.
- **Connectedness:** H_δ is connected if there exists a path between every pair of vertices in V.

In the case of a quasigroup, the Latin-square property ensures that the support hypergraph is connected. This property allows local algebraic constraints, such as the collinear constraints (7), to propagate across all indices.

A.2 Enumeration and Symmetry Classes

While the main text formulates the problem for general quasigroups, our experimental analysis (Appendix D) relies on the enumeration of loops to track the search space size.

Loops and Reduced Latin Squares. A loop is a quasigroup with a two-sided identity element $e \in Q$. If we label the elements of Q as $\{0, 1, \ldots, n-1\}$ with $0 = e$, the operation table is a *reduced* (or normalized) Latin square, where the first row and first column are in natural order $(0, 1, \ldots, n-1)$. Classifying loops of order n is equivalent to classifying reduced Latin squares of order n.

Isomorphism vs. Isotopy. The enumeration of distinct structures depends on the chosen notion of equivalence:

- **Loop Isomorphism (Simultaneous Relabeling):** Two loops are isomorphic if they differ by a bijection $\phi : Q \to Q'$ satisfying $\phi(a) \circ' \phi(b) = \phi(a \circ b)$. In terms of Latin squares, this corresponds to applying the *same* permutation to rows, columns, and symbol labels simultaneously. We use this relation to identify unique loops.
- **Quasigroup Isotopy (Independent Relabeling):** The HyperCube model, which treats the operation as a 3-way tensor without a privileged identity, naturally respects a larger internal symmetry group. Two Latin squares are *isotopic* if they differ by an arbitrary triple of permutations $(\phi, \psi, \chi) \in S_n \times S_n \times S_n$ acting independently on rows, columns, and symbols.

Since the HyperCube objective is exactly invariant under this larger group (as detailed in Section 3.4), counting distinct structures for the optimization landscape corresponds to counting quasigroup isotopy classes (orbits of ${S_n}^3$).

B Deferred Proofs

B.1 Proofs of Auxiliary Lemmas

Proof of Lemma 1: Orthogonal Decomposition. Let the geometric misalignment matrix (for matrix A) be

$$\Delta_{abc}^{(A)} \coloneqq (B_b C_c)^\dagger - T_{abc}^* A_a / \|A_a\|^2$$

Expanding its squared Frobenius norm yields:

$$\begin{aligned}
\|\Delta_{abc}^{(A)}\|^2 &= \left\langle (B_b C_c)^\dagger - \frac{T_{abc}^*}{\|A_a\|^2} A_a,\ (B_b C_c)^\dagger - \frac{T_{abc}^*}{\|A_a\|^2} A_a \right\rangle \\
&= \|B_b C_c\|^2 - \frac{T_{abc}}{\|A_a\|^2} \langle (B_b C_c)^\dagger, A_a \rangle - \frac{T_{abc}^*}{\|A_a\|^2} \langle A_a, (B_b C_c)^\dagger \rangle + \frac{|T_{abc}|^2}{\|A_a\|^4} \|A_a\|^2 \\
&= \|B_b C_c\|^2 - \frac{T_{abc} T_{abc}^*}{\|A_a\|^2} - \frac{T_{abc}^* T_{abc}}{\|A_a\|^2} + \frac{|T_{abc}|^2}{\|A_a\|^2} \\
&= \|B_b C_c\|^2 - \frac{|T_{abc}|^2}{\|A_a\|^2}.
\end{aligned}$$

Applying this expansion symmetrically to $\Delta_{abc}^{(B)}$ and $\Delta_{abc}^{(C)}$, and summing over all supported triples (a, b, c) weighted by the target δ_{abc}:

$$\begin{aligned}
\mathcal{R}_\delta(\Theta) &= \sum_{a,b,c} \delta_{abc} \left(\|B_b C_c\|^2 + \|C_c A_a\|^2 + \|A_a B_b\|^2 - |T_{abc}|^2 \left(\frac{1}{\|A_a\|^2} + \frac{1}{\|B_b\|^2} + \frac{1}{\|C_c\|^2} \right) \right) \\
&= \mathcal{H}(\Theta) - \mathcal{B}_\delta(\Theta).
\end{aligned}$$

Rearranging this equation yields the exact decomposition $\mathcal{H}(\Theta) = \mathcal{B}_\delta(\Theta) + \mathcal{R}_\delta(\Theta)$. Since $\mathcal{R}_\delta$ is defined as a sum of squared norms, $\mathcal{R} \geq 0$, enforcing the lower bound $\mathcal{H} \geq \mathcal{B}$. □

Proof of Lemma 2: Shared Gram Matrices. Substituting the alignment identities (7) into the associativity condition $A_a(B_b C_c) = (A_a B_b) C_c$ and cancelling the nonzero scalar T_{abc} yields $A_a A_a^\dagger / \|A_a\|^2 = C_c^\dagger C_c / \|C_c\|^2$. The Latin-square property ensures that for any fixed a, as we vary b, the index c varies uniquely to maintain support. The left side depends only on a, while the right side depends only on c. Since the support graph of a quasigroup is connected, this equality propagates across all indices, implying both sides equal a constant matrix X. Taking the trace yields $\mathrm{Tr}(X) = \mathrm{Tr}(A_a A_a^\dagger)/\|A_a\|^2 = n\|A_a\|^2/\|A_a\|^2 = n$. The remaining equalities for Y and Z are obtained analogously. □

Proof of Lemma 3: Normalized Rank κ. Fix a supported triple (a, b, c). Starting from $X = C_c^\dagger C_c / \|C_c\|^2$ and substituting the collinear relation $C_c^\dagger \propto A_a B_b$ from (7) and the Gram identity for Y from (8):

$$X = \frac{\|C_c\|^2}{|T_{abc}|^2} (A_a B_b)(B_b^\dagger A_a^\dagger) = \frac{\|B_b\|^2 \|C_c\|^2}{|T_{abc}|^2} A_a \left(\frac{B_b B_b^\dagger}{\|B_b\|^2} \right) A_a^\dagger = \frac{\|B_b\|^2 \|C_c\|^2}{|T_{abc}|^2} \frac{A_a (A_a^\dagger A_a) A_a^\dagger}{\|A_a\|^2} = \kappa_{abc} X^2.$$

Since X is index-independent (Lemma 2), the coefficient κ_{abc} must be a constant κ. Define $P \coloneqq \kappa X$. Then $P^2 = \kappa^2 X^2 = \kappa X = P$, making P an orthogonal projection. Thus, $\mathrm{rank}(X) = \mathrm{rank}(P) = \mathrm{Tr}(P) = \kappa\ \mathrm{Tr}(X) = \kappa n$. Rearranging yields $\kappa = \mathrm{rank}(X)/n \leq 1$. If $\kappa = 1$, then X is a trace-n projection, implying $X = I_n$. □

Proof of Lemma 4: AM-GM Lower Bound. Assume collinearity ($\mathcal{R} = 0$) and $T_{abc} \neq 0$ on the support. Let $\alpha_a \coloneqq \|A_a\|^{-2}$, $\beta_b \coloneqq \|B_b\|^{-2}$, and $\gamma_c \coloneqq \|C_c\|^{-2}$. Recall from the definition of the normalized rank κ (Lemma 3) that $\kappa = \|A_a\|^2 \|B_b\|^2 \|C_c\|^2 / |T_{abc}|^2$ on the support of δ. This implies the product identity: $\alpha_a \beta_b \gamma_c = \kappa^{-1} |T_{abc}|^{-2}$.

Applying the scalar AM-GM inequality ($\alpha + \beta + \gamma \geq 3(\alpha\beta\gamma)^{1/3}$) to the definition of the base penalty $\mathcal{B}_\delta$ yields:

$$\begin{aligned}\mathcal{B}_\delta(\Theta) &= \sum_{a,b,c} \delta_{abc}|T_{abc}|^2(\alpha_a + \beta_b + \gamma_c) \\ &\geq \sum_{a,b,c} \delta_{abc}|T_{abc}|^2 \cdot 3(\alpha_a\beta_b\gamma_c)^{1/3} \\ &= 3\kappa^{-1/3} \sum_{a,b,c} \delta_{abc}\,|T_{abc}|^{4/3}.\end{aligned}$$

Equality holds *if and only if* the AM-GM condition $\alpha_a = \beta_b = \gamma_c$ is satisfied for every supported triple (a, b, c). Due to the connectivity of the quasigroup support hypergraph (as detailed in Appendix A.1), these local equalities propagate globally, ensuring all factor norms are perfectly balanced: $\|A_a\|^2 = \|B_b\|^2 = \|C_c\|^2$ for all indices across the entire representation. □

B.2 Proof of Theorem 8: Unitary Collinearity–Associativity Equivalence

To formally prove Theorem 8, we first establish supporting lemmas regarding gauge synchronization, the resulting homomorphism, and the constructive sufficiency of the left-regular representation. By isotopy equivariance (Section 3.4), we assume without loss of generality that the target δ is a loop $(Q, \circ')$ (i.e. a quasigroup with an identity e).

Lemma 10 (Synchronization)**.** *Let $(Q, \circ')$ be a finite loop admitting a unitary collinear factorization Θ. There exists a unitary gauge transformation (U, V, W) such that the transformed parameter triple $\Theta' = (A', B', C')$ satisfy the synchronized condition: $A'_g = B'_g = (C'_g)^\dagger := \rho(g)$ for all $g \in Q$.*

Proof. Since the factorization is unitary, the slices A_e, B_e are unitary matrices. We explicitly define the unitary gauge as $(U, V, W) := (A_e^\dagger, I_n, B_e)$. Under collinearity and unitarity ($|T_{abc}| = \|A_a\| = \|B_b\| = \|C_c\| = 1$), the projection identity $A_aB_b \propto C_c^\dagger$ holds with unit scalars. Applying this to triples (g, e, g) and (e, g, g) yields $A_gB_e = C_g^\dagger$ and $A_eB_g = C_g^\dagger$, implying $A_gB_e = A_eB_g$. Applying the gauge transformation:

$$A'_g = A_e^\dagger A_g = A_e^\dagger(A_eB_gB_e^\dagger) = B_gB_e^\dagger = B'_g.$$

For the third factor:

$$C'_g = B_eC_gA_e = B_e(B_e^\dagger A_g^\dagger)A_e = A_g^\dagger A_e = (A'_g)^\dagger.$$

Thus, all three factors collapse to a single unitary map $\rho(g)$. □

Lemma 11 (Homomorphism and Injectivity)**.** *Let $\rho(g)$ be the synchronized map from Lemma 10. The collinearity condition enforces that ρ is a homomorphism. Furthermore, the feasibility condition $T(\Theta) = \delta$ ensures ρ is an injective map.*

Proof. By the collinear identities of the transformed factors, $A'_aB'_b = T'_{abc}(C'_c)^\dagger$ for supported triples ($a \circ' b = c$). Substituting the synchronized map yields $\rho(a)\rho(b) = T'_{abc}\rho(c)$. Because the factorization is unitary and feasible, the trace inner product evaluates to $T'_{abc} = 1$. Thus, $\rho(a)\rho(b) = \rho(c) = \rho(a \circ' b)$, establishing ρ as a true group homomorphism into $U(n)$.

To prove injectivity, suppose $\rho(x) = \rho(y)$ for some $x, y \in Q$. For the valid triple (x, e, x), feasibility requires $T'_{xex} = \frac{1}{n}\operatorname{Tr}(\rho(x)\rho(e)\rho(x)^\dagger) = 1$. Because $\rho(x) = \rho(y)$, substituting y yields: $T'_{yex} = \frac{1}{n}\operatorname{Tr}(\rho(y)\rho(e)\rho(x)^\dagger) = \frac{1}{n}\operatorname{Tr}(\rho(x)\rho(e)\rho(x)^\dagger) = 1$. Feasibility ($T'_{yex} = \delta_{yex}$) dictates that $\delta_{yex} = 1$, which means $y \circ' e = x$. Since e is the identity, $y \circ' e = y$, forcing $x = y$. Thus, ρ is injective. □

Proof of Theorem 8: Unitary Collinearity $\iff$ Group Isotope. (**Necessity**): As established via isotopy equivariance, if $(Q, \circ)$ admits a unitary collinear factorization, so does its loop isotope $(Q, \circ')$. By Lemma 10, this factorization induces a synchronized map $\rho : Q \to U(n)$. Since

ρ is a homomorphism (Lemma 11), invoking the associativity of continuous matrix multiplication yields:

$$\rho((x \circ' y) \circ' z) = \rho(x)\rho(y)\rho(z) = \rho(x \circ' (y \circ' z)).$$

Since ρ is injective (Lemma 11), this equality implies the associativity of the discrete operation $\circ'$: $(x \circ' y) \circ' z = x \circ' (y \circ' z)$. Thus, the loop $(Q, \circ')$ is an associative group, and the original quasigroup $(Q, \circ)$ is a group isotope.

(**Sufficiency**): This direction is established constructively in Lemma 13, which proves that the left-regular representation of the group isotope provides a valid unitary collinear factorization. □

Lemma 12 (Uniqueness of Representation)**.** *Let $(Q, \circ')$ be the group identified in the necessity proof. The induced map $\rho : Q \to U(n)$ is unitarily equivalent to the left-regular representation of $(Q, \circ')$.*

Proof. Since ρ is a true group homomorphism (Lemma 11), we can evaluate the feasibility constraint on the specific triple (g, e, e):

$$T_{gee} = \frac{1}{n} \operatorname{Tr}(\rho(g)\rho(e)\rho(e)^\dagger) = \frac{1}{n} \operatorname{Tr}(\rho(g \circ' e \circ' e^{-1})) = \frac{1}{n} \operatorname{Tr}(\rho(g)) = \delta_{gee}.$$

Thus, the character of ρ evaluates to: $\operatorname{Tr}(\rho(g)) = n \cdot \delta_{gee} = n \cdot \mathbb{I}_{\{g=e\}}$, which is precisely the character of the left-regular representation. By standard character theory, two representations with identical characters are unitarily equivalent. Thus, the factorization is uniquely determined as the left-regular representation up to unitary equivalence. □

Lemma 13 (Sufficiency: Group Isotope $\implies$ Unitary Collinear Factorization)**.** *Let $(Q, \circ)$ be a quasigroup isotopic to a group $(Q, \circ')$. Then, $(Q, \circ)$ admits a unitary collinear factorization ($\mathcal{F}_\delta \cap \mathcal{M}_\delta \cap \mathcal{U}$) and attains an objective value of $\mathcal{H}(\Theta) = 3\,|\delta|$.*

Proof. Let $\rho : Q \to U(n)$ be the left-regular representation of the group $(Q, \circ')$. By definition, $\rho(g)$ is a unitary matrix for all $g \in Q$. Define a synchronized unitary parameter triple $\Theta \in \mathcal{U}$ as: $A_g = B_g = C_g^\dagger = \rho(g)$.

For any supported triple $a \circ' b = c$, we have $A_a B_b = \rho(a)\rho(b) = \rho(a \circ' b) = \rho(c) = C_c^\dagger$, satisfying collinearity ($\mathcal{M}_\delta$) (7). Also, their product (1) satisfies feasibility ($\mathcal{F}_\delta$).

$$T_{abc}(\Theta) = \frac{1}{n} \operatorname{Tr}(\rho(a)\rho(b)\rho(c)^\dagger) = \frac{1}{n} \operatorname{Tr}(\rho(c)\rho(c)^\dagger) = \frac{1}{n} \operatorname{Tr}(I_n) = 1 = \delta_{abc}.$$

For any unitary parameters $\Theta \in \mathcal{U}$, the objective (2) evaluates to

$$\mathcal{H}(\Theta) = \sum_{a,b,c} \delta_{abc}(\|B_b C_c\|^2 + \|C_c A_a\|^2 + \|A_a B_b\|^2) = \sum_{a,b,c} \delta_{abc}(1^2 + 1^2 + 1^2) = 3\,|\delta|.$$

□

C Matrix AM-GM Inequality

Throughout this section, we generalize the normalized trace and the normalized Frobenius norm from the main text to square matrices of arbitrary dimension N: For $M \in \mathbb{C}^{N \times N}$, we define and the norm as $\|M\|^2 \coloneqq \frac{1}{N} \operatorname{Tr}(M^\dagger M)$.

C.1 Supporting Lemma

Lemma 14 (Spectral Trace–Frobenius Bound)**.** *For every $M \in \mathbb{C}^{N \times N}$, the normalized trace and Frobenius norm satisfy*

$$\|M^2\|^2 = \frac{1}{N} \operatorname{Tr}(M^{2\dagger} M^2) \geq \left| \frac{1}{N} \operatorname{Tr}(M^3) \right|^{4/3}. \tag{17}$$

Proof. Let $\lambda_1, \ldots, \lambda_N$ be the eigenvalues of M, counted with algebraic multiplicity, and set $r_i := |\lambda_i|$. Since the eigenvalues of M^3 are λ_i^3, we have $\mathrm{Tr}(M^3) = \sum_{i=1}^N \lambda_i^3$. Hence,

$$\begin{aligned}
\left|\frac{1}{N}\mathrm{Tr}(M^3)\right|^{4/3} &= \left|\frac{1}{N}\sum_{i=1}^N \lambda_i^3\right|^{4/3} \\
&\le \left(\frac{1}{N}\sum_{i=1}^N r_i^3\right)^{4/3} \\
&\le \frac{1}{N}\sum_{i=1}^N r_i^4 \qquad \text{(Jensen's inequality)} \qquad (18)\\
&= \frac{1}{N}\sum_{i=1}^N |\lambda_i^2|^2 \\
&\le \|M^2\|^2. \qquad \text{(Schur/Frobenius domination)} \qquad (19)
\end{aligned}$$

The last inequality follows by applying Schur triangularization to M^2. If $Q^\dagger M^2 Q = T$ is upper triangular with diagonal entries λ_i^2, then

$$\|M^2\|^2 = \frac{1}{N}\mathrm{Tr}(T^\dagger T) = \frac{1}{N}\left(\sum_i |\lambda_i^2|^2 + \sum_{i<j}|T_{ij}|^2\right) \ge \frac{1}{N}\sum_i |\lambda_i^2|^2.$$

Equality in this Schur/Frobenius step holds if and only if the Schur form has no strictly upper triangular part, equivalently if and only if M^2 is normal. □

C.2 A Noncommutative Trace–Frobenius Matrix AM-GM Inequality

The following matrix inequality replaces the scalar AM-GM argument. Crucially, it applies directly to arbitrary matrices without assuming commutativity.

[Matrix AM-GM Lemma 5] For all $X, Y, Z \in \mathbb{C}^{n\times n}$,

$$\|XY\|^2 + \|YZ\|^2 + \|ZX\|^2 \ge 3\,|\,\mathrm{tr}(XYZ)|^{4/3}. \qquad (20)$$

Moreover, equality holds if and only if X, Y, Z are scaled unitarities that satisfy $XX^\dagger = YY^\dagger = ZZ^\dagger = \tau^2 I_n$ with $\tau \neq 0$ and $XYZ = \tau^3 I$.

Proof. Define the block-cyclic matrix

$$M := \begin{pmatrix} 0 & X & 0 \\ 0 & 0 & Y \\ Z & 0 & 0 \end{pmatrix} \in \mathbb{C}^{3n\times 3n}.$$

Then

$$M^2 = \begin{pmatrix} 0 & 0 & XY \\ YZ & 0 & 0 \\ 0 & ZX & 0 \end{pmatrix}, \qquad M^3 = \begin{pmatrix} XYZ & 0 & 0 \\ 0 & YZX & 0 \\ 0 & 0 & ZXY \end{pmatrix}.$$

Evaluating the normalized norm of M^2 yields:

$$\|M^2\|^2 = \frac{1}{3n}\mathrm{Tr}\left((M^2)^\dagger M^2\right) = \frac{1}{3}\left(\|XY\|^2 + \|YZ\|^2 + \|ZX\|^2\right).$$

Similarly, by the cyclicity of the standard trace, $\mathrm{Tr}(YZX) = \mathrm{Tr}(ZXY) = \mathrm{Tr}(XYZ)$. Evaluating the normalized trace of M^3 yields:

$$\mathrm{tr}(M^3) = \frac{1}{3n}\mathrm{Tr}(M^3) = \frac{1}{3n}\left(3\,\mathrm{Tr}(XYZ)\right) = \mathrm{tr}(XYZ).$$

Substituting M into the spectral bound from Lemma 14 gives exactly (12).

It remains to prove the equality case under the assumption $\operatorname{tr}(XYZ) = 1$. Suppose equality holds in (12). Then $\operatorname{tr}(M^3) = 1$ and

$$\|M^2\|^2 = \frac{1}{3}\big(\|XY\|^2 + \|YZ\|^2 + \|ZX\|^2\big) = 1.$$

Thus equality holds at every step of the spectral chain above for M. Let $\lambda_1, \dots, \lambda_{3n}$ be the eigenvalues of M and put $r_i = |\lambda_i|$.

First, equality in the Jensen's step (18)

$$\left(\frac{1}{3n}\sum_i r_i^3\right)^{1/3} \le \left(\frac{1}{3n}\sum_i r_i^4\right)^{1/4}$$

forces all r_i to be equal, say $r_i = \rho$ for every i. Since equality also holds in the triangle step and $\operatorname{tr}(M^3) = 1$, we have

$$1 = \left|\frac{1}{3n}\sum_i \lambda_i^3\right| = \frac{1}{3n}\sum_i |\lambda_i|^3 = \rho^3.$$

Hence $\rho = 1$; all eigenvalues of M lie on the unit circle.

Second, equality in the triangle inequality implies that the complex numbers λ_i^3 all have the same argument. Since $|\lambda_i^3| = 1$ and their average is the real number 1,

$$\frac{1}{3n}\sum_i \lambda_i^3 = 1,$$

the common argument must be zero. Therefore

$$\lambda_i^3 = 1 \qquad \text{for every } i. \tag{21}$$

Third, equality in the Schur/Frobenius domination step (19) forces M^2 to be normal. The eigenvalues of M^2 are λ_i^2, and by the preceding paragraph they all have modulus one. A normal matrix with all eigenvalues on the unit circle is unitary; hence

$$M^2 \text{ is unitary.} \tag{22}$$

Indeed, after unitary diagonalization of M^2, the relation $(M^2)^\dagger M^2 = I$ follows entrywise from $|\lambda_i^2| = 1$.

We next show that $M^3 = I_{3n}$, not merely that the eigenvalues of M^3 are equal to 1. Since M^2 is normal and its eigenvalues λ_i^2 satisfy $(\lambda_i^2)^3 = 1$, functional calculus for normal matrices gives

$$(M^2)^3 = I_{3n},$$

so $M^6 = I_{3n}$. The polynomial $t^6 - 1$ has no repeated roots, so any matrix annihilated by it is diagonalizable. Thus M is diagonalizable. Combining diagonalizability with (21) gives

$$M^3 = I_{3n}. \tag{23}$$

Now inspect the block consequences of (22) and (23). From the displayed form of M^2, unitarity of M^2 implies that each block product XY, YZ, and ZX is unitary. From the displayed form of M^3, the identity (23) gives

$$XYZ = I, \qquad YZX = I, \qquad ZXY = I. \tag{24}$$

Finally, these facts force the individual factors to be unitary. Since XY is unitary and $XYZ = I$, we have

$$Z = (XY)^{-1} = (XY)^\dagger,$$

so Z is unitary. Since YZ is unitary and Z is unitary,

$$Y = (YZ)Z^\dagger$$

is unitary. Since XY is unitary and Y is unitary,

$$X = (XY)Y^\dagger$$

is unitary. Hence equality implies X, Y, Z are unitary and $XYZ = I$.

Conversely, if X, Y, Z are unitary and $XYZ = I$, then XY, YZ, and ZX are unitary, so each has normalized Frobenius norm squared equal to 1, while $|\operatorname{tr}(XYZ)|^{4/3} = |\operatorname{tr}(I)|^{4/3} = 1$. Thus both sides of (12) equal 3, proving equality. □

D EMPIRICAL VERIFICATION

D.1 TRAINING SETTING AND HYPERPARAMETERS

To ensure reproducibility and provide a granular view of the learning dynamics discussed in the main text, we detail the specific training hyperparameters used to generate the optimization trajectories. We optimize the model by minimizing the mean squared error (MSE) reconstruction loss augmented by the HyperCube objective:

$$\mathcal{L}(\Theta) = \|T(\Theta) - \delta\|_F^2 + \lambda\mathcal{H}(\Theta) \tag{25}$$

We track the objective components, *i.e.*, $\mathcal{H}, \mathcal{B}_\delta, \mathcal{R}_\delta, \mathcal{H}_\delta^*$, and the ℓ_2 norm ($\|\Theta\|^2$) for reference.

We use the standard Gradient Descent optimizer (learning rate 0.5, zero momentum) with a two-phase schedule for λ, adopting a standard **penalty continuation method**. Theoretically, minimizing $\mathcal{H}$ subject to strict feasibility ($T = \delta$) corresponds to the limit $\lambda \to 0$ in the unconstrained loss $\mathcal{L}(\Theta)$, which is impractical for training. To resolve this, we set $\lambda = 0.05$ for the first 500 epochs, sufficient for convergence, after which the coefficient is turned off ($\lambda = 0$) to yield perfect reconstruction (exact feasibility). Crucially, the normalized objective components $\tilde{\mathcal{B}}_\delta := \mathcal{B}_\delta/\mathcal{H}_\delta^*$ and $\tilde{\mathcal{R}}_\delta := \mathcal{R}_\delta/\mathcal{H}_\delta^*$ remain invariant upon this transition. This confirms the validity of the continuation technique: the finite λ reliably guides the optimizer toward the true constrained minimum without artificially distorting the geometric structure of the converged solution.

D.2 DETAILED OPTIMIZATION DYNAMICS: ASSOCIATIVE VS. NON-ASSOCIATIVE TARGETS

To understand the localized optimization mechanism, Figure 3 plots the evolution of the objective components over training for two distinct $n = 6$ targets: a commutative group (Z_6) and a non-associative quasigroup (NG_2).

D.2.1 GROUP ISOTOPE DYNAMICS (Z_6):

As shown in Figure 3 (Top), the optimizer successfully discovers the collinear set. The geometric misalignment penalty $\mathcal{R}_\delta$ smoothly decays to exactly zero around epoch 400. The total objective $\mathcal{H}$, the base penalty $\mathcal{B}_\delta$, the dynamic tracking floor $\mathcal{H}_\delta^*$, and the ℓ_2 norm all perfectly converge to the universal lower bound of $3\,|\delta| = 3 \cdot 6^2 = 108$, with zero misalignment ($\mathcal{R}_\delta = 0$).

D.2.2 NON-GROUP DYNAMICS (NG_2):

Conversely, Figure 3 (Bottom) illustrates the structural obstruction faced by non-associative targets. Because the unitary collinear set is empty for NG_2, the misalignment penalty $\mathcal{R}_\delta$ plateaus and never decays to zero. Following the transition to strict feasibility at epoch 500, the dynamic floor $\mathcal{H}_\delta^*$ naturally stabilizes at 108; however, the trajectory retains a geometric obstruction. At convergence, the components are strictly separated: the total objective is elevated to $\mathcal{H} \approx 145$, comprising a base penalty of $\mathcal{B}_\delta \approx 95$ and a massive residual misalignment of $\mathcal{R}_\delta \approx 55$ (with an L_2 norm of 125). This observed separation illustrates the objective excess above the dynamic floor for this non-group target.

Table 1: Cayley Table for $Z_2 \times Z_3 \cong Z_6$ (Group Isotope, zero latent associativity violation)

$\circ$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	0	3	2	5	4
2	2	3	4	5	0	1
3	3	2	5	4	1	0
4	4	5	0	1	2	3
5	5	4	1	0	3	2

D.3 POPULATION RESULTS AND THE GLOBAL LANDSCAPE

While the individual trajectories above illustrate the optimization mechanics for specific targets, we must verify that this strict variational trade-off generalizes across the broader algebraic landscape.

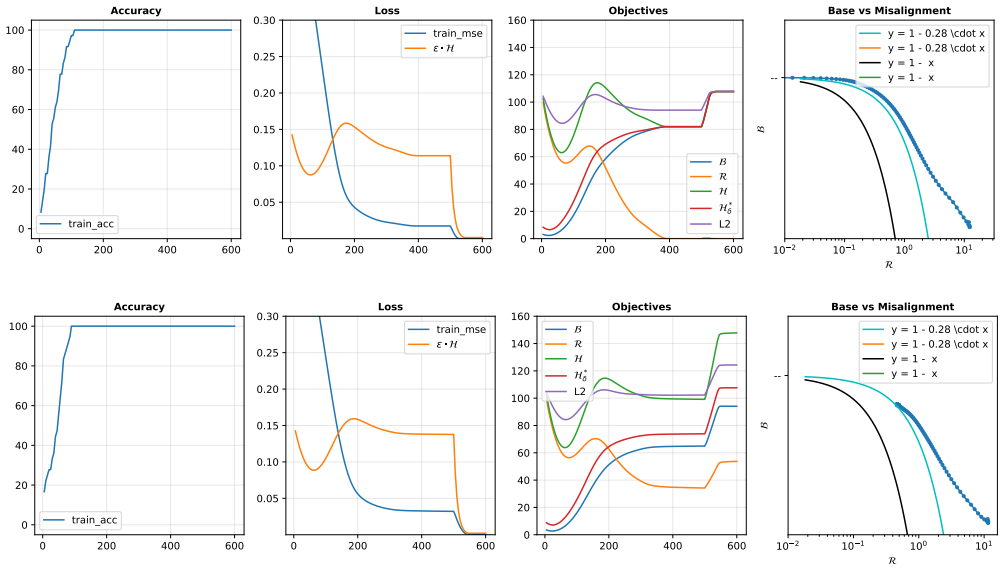


Figure 3: **Component-wise Optimization Dynamics.** The evolution of the objective components $(\mathcal{H}, \mathcal{B}_\delta, \mathcal{R}_\delta, \mathcal{H}^*_\delta)$ alongside the train loss and accuracy during optimization. The HyperCube penalty λ is held at 0.05 until epoch 500, and then dropped to 0 to ensure exact feasibility. **(Top)** For the Z_6 group isotope, the misalignment $\mathcal{R}_\delta$ reliably reaches zero. At convergence ($t > 500$), all relevant components $(\mathcal{H}, \mathcal{B}_\delta, \mathcal{H}^*_\delta)$ collapse exactly to the $3\,|\delta| = 108$ theoretical floor. **(Bottom)** For the NG_2 non-group target, structural obstruction prevents $\mathcal{R}_\delta$ from reaching zero. At convergence, the total objective $\mathcal{H}$ is strictly elevated above the $\mathcal{H}^*_\delta = 108$ floor, visualizing the permanent penalty exacted by the Associativity Gap.

Table 2: Cayley Table for a Non-Group Isotope: Latent associativity violation count = 100

$\circ$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	0	3	4	5	2
2	2	4	5	0	1	3
3	3	5	4	2	0	1
4	4	2	1	5	3	0
5	5	3	0	1	2	4

Specifically, we scale our analysis to quantify the interaction between the misalignment penalty $\mathcal{R}_\delta$ and the base term $\mathcal{B}_\delta$ over a comprehensive population of quasigroups (as visualized in Figure 2 of the main text).

Experimental Setup. We analyze reduced Latin squares δ corresponding to loops of orders $n \in \{5, 6, 7, 8\}$. For the smaller orders $n \in \{5, 6\}$, we exhaustively evaluate all unique loops up to isomorphism (6 and 109 cases, respectively). For $n \in \{7, 8\}$, we evaluate a random subset of 100 loops per order, due to the combinatorial explosion of the search space (e.g., 23,746 unique loops for $n = 7$).

For each δ, we record the numerical optimum value from a gradient-based optimizer with multiple random initial values to mitigate local minima. Let $\mathfrak{L}(\delta)$ be the set of reduced loop tables isotopic to δ. To correlate the landscape geometry with algebraic structure, we define the *latent associativity violation ratio*

$$\nu(\delta) := \min_{(Q,\star)\in\mathfrak{L}(\delta)} \frac{1}{|\delta|} \sum_{a,b,c\in Q} \mathbb{I}_{\{(a\star b)\star c \neq a\star(b\star c)\}}.$$

Every quasigroup is isotopic to a loop, so $\mathfrak{L}(\delta)$ is nonempty. If δ and δ' are isotopic, then $\mathfrak{L}(\delta) = \mathfrak{L}(\delta')$; consequently, $\nu(\delta) = \nu(\delta')$. Moreover, $\nu(\delta) = 0$ if and only if δ is a group isotope.

Results. We analyze the normalized objective terms ($\tilde{\mathcal{B}}_\delta \coloneqq \mathcal{B}_\delta/\mathcal{H}^*_\delta$ and $\tilde{\mathcal{R}}_\delta \coloneqq \mathcal{R}_\delta/\mathcal{H}^*_\delta$). At random initialization, models reliably occupy a state of high geometric misalignment ($\tilde{\mathcal{R}}_\delta$) and a comparably low base penalty ($\tilde{\mathcal{B}}_\delta$).

As optimization proceeds, the trajectories sweep right-to-left (Figure 2, Left), revealing an empirical trade-off between the objective terms: in the reported runs, the base penalty increases ($\tilde{\mathcal{B}}_\delta \to 1$) as the misalignment decreases ($\tilde{\mathcal{R}}_\delta \to 0$). The explored states lie near a curved empirical boundary. Along this boundary, the slope of the trade-off is bounded by $\Delta\tilde{\mathcal{B}}_\delta/\Delta\tilde{\mathcal{R}}_\delta \gtrsim -0.28$. The converged points cluster tightly along this boundary (Figure 2, Middle).

Empirical Pressure Toward Collinearity. These relations provide empirical evidence that the objective favors geometric alignment along the reported trajectories. Because the observed trade-off slope satisfies $\Delta\tilde{\mathcal{B}}_\delta/\Delta\tilde{\mathcal{R}}_\delta \gtrsim -0.28$, the corresponding rate of change for the total objective satisfies $\Delta\tilde{\mathcal{H}}_\delta/\Delta\tilde{\mathcal{R}}_\delta \approx 1 + \Delta\tilde{\mathcal{B}}_\delta/\Delta\tilde{\mathcal{R}}_\delta \gtrsim 0.72$ along these trajectories. This guarantees that any reduction in misalignment yields a strict net decrease in the total objective ($\Delta\tilde{\mathcal{H}}_\delta \lesssim 0.72\Delta\tilde{\mathcal{R}}_\delta < 0$).

A Differentiable Measure for Group Isotopy. Moreover, the empirical optima monotonically grows with the latent associativity violation ratio (Figure 2, Right):

$$\tilde{\mathcal{H}}(\delta) \approx 1 + c_H \cdot \nu(\delta)/|\delta| \quad (c_H \approx 0.12).$$

This positive fitted coefficient ($c_H > 0$) indicates that attained optima monomoincrease with the latent associativity violation ratio.

Crucially, both $\mathcal{H}_{\inf}$ and ν are **isotopy invariant**. The observed correlation therefore compares quantities defined on isotopy classes rather than on arbitrary displayed tables.