The Geometry of Algorithmic Stability: A Hodge Theoretic View on Structural vs. Statistical Instability

Anonymous authors
Paper under double-blind review

Abstract

Algorithmic stability—the robustness of predictions to training data perturbations—is fundamental to reliable machine learning. We propose a unified mathematical framework that rigorously distinguishes between two fundamental sources of instability: structural inconsistency and statistical variance. We formalize structural inconsistency using Combinatorial Hodge Theory, characterizing it as cyclical flows (Condorcet cycles) on a graph of hypotheses. This framework reveals that methods like inflated operators and regularization specifically target these structural obstructions, while methods like bagging primarily address statistical variance. We provide direct empirical validation for this distinction. First, we demonstrate that in the presence of engineered Condorcet cycles (pure structural instability), inflated operators achieve perfect stability while bagging fails. Second, using Hodge decomposition as a diagnostic tool, we show that instability in standard classification tasks is primarily statistical (cyclical flows are near zero), explaining the effectiveness of bagging. Crucially, we demonstrate that significant structural obstructions (non-zero cyclical flows) emerge in fairness-constrained model selection on real-world data, providing a topological characterization of the instability arising from incompatible objectives. Third, we derive and validate a novel technique, obstruction-aware regularization, which directly enforces mathematical consistency and improves stability, along with a scalable stochastic implementation.

1 Introduction

The generalization performance of machine learning models is deeply connected to their stability—the ability to produce consistent predictions when trained on slightly different datasets (Bousquet & Elisseeff, 2002). This has motivated substantial research into stability enhancement methods, including the canonical approach of bagging (Breiman, 1996; Soloff et al., 2024), regularization (Bousquet & Elisseeff, 2002), and more recently, "inflated" operators (Adrian et al., 2024; Liang et al., 2025).

While effective, these methods appear disconnected, and it is often unclear which method is appropriate for a given source of instability. This fragmentation raises a fundamental question: Can we rigorously distinguish the sources of instability and unify the approaches to address them?

We argue the answer is yes, provided by the tools of Combinatorial Hodge Theory applied to the geometry of preferences. We propose that instability arises from two distinct sources: statisti-

cal variance (sensitivity of the learning process to noise or near-ties) and structural inconsistency (fundamental conflicts in the data preferences).

We can conceptualize structural inconsistency by viewing subsets of the data as "voters" and the potential models as "candidates."

Example 1.1 (Running Example: The Ranking Cycle)**.** Consider selecting the best model among three candidates {A,B,C}. Suppose one subset of the data (voter group 1) strongly suggests A is better than B ($A > B$). A second subset suggests $B > C$, and a third suggests $C > A$. This creates a cycle: $A > B > C > A$. No single global ranking can satisfy all local preferences. An algorithm forced to choose a single winner will be highly unstable in this scenario.

This scenario is the well-known Condorcet Paradox. We demonstrate that these paradoxes, which can be rigorously characterized by the cyclical component in a Hodge decomposition (Lim, 2020; Jiang et al., 2008), represent structural instability. This leverages established connections between social choice theory and algebraic topology (Chichilnisky, 1980; Baryshnikov, 1993).

While we find that structural inconsistencies are rare in standard supervised learning tasks—explaining the empirical success of methods like bagging—we demonstrate that they arise naturally in domains with conflicting objectives, such as fairness-aware machine learning. Our framework provides a mathematical characterization of the inherent instability in these scenarios.

Our framework formalizes the stability problem as one of preference aggregation under metric constraints. We show that:

1. The Hodge decomposition of ordinal preferences characterizes the mathematical conditions that determine the source of instability. A non-zero cyclical component ($C_{cycle} > 0$) indicates structural inconsistency, while $C_{cycle} \approx 0$ indicates instability is primarily due to statistical variance.
2. Structural inconsistency requires structural solutions. Inflated operators provide obstruction resolution (target space enlargement), while regularization can provide obstruction prevention.
3. Statistical variance is effectively addressed by variance reduction techniques like bagging. Crucially, we demonstrate that bagging is ineffective against pure structural instability.

This framework provides mathematical unification, revealing deep connections between algorithmic stability, social choice theory, and algebraic topology, and provides theoretical insight into selecting appropriate stabilization methods.

2 Related Work

Our work sits at the intersection of algorithmic stability, topological data analysis, and social choice theory.

2.1 Foundations of Algorithmic Stability

The foundational work of Bousquet & Elisseeff (2002) established the link between algorithmic stability and generalization. The standard paradigm often addresses stability through statistical

variance reduction, canonically exemplified by bagging (Breiman, 1996), whose assumption-free stability properties have been recently formalized (Soloff et al., 2024).

More recently, research has focused on the instability inherent in selection procedures like the argmax. Inflated operators (Adrian et al., 2024; Liang et al., 2025) address this by returning a stable set of near-optimal solutions, motivated largely by the metric sensitivity of the loss function under near-ties. Our work provides a unifying theoretical framework that reinterprets these methods as strategies for handling topological obstructions in the underlying data preferences.

2.2 Topological Data Analysis in Machine Learning

The application of Topological Data Analysis (TDA) to machine learning is a growing field ((Wasserman, 2018; Hensel et al., 2021; Zia et al., 2024)). Much of this work focuses on the geometry of the parameter space or the loss landscape. For instance, persistent homology and Morse theory have been used to identify obstructions to optimization, such as the structure of local minima and saddle points (Barannikov et al., 2022).

Crucially, our work analyzes the geometry of the preference space—how data induces rankings over the hypothesis space. The obstructions we identify are not features of the optimization landscape, but rather fundamental inconsistencies in the problem definition itself (conflicting data preferences), which exist independently of the optimization procedure.

2.3 Cohomology and Hodge Theory in ML

The mathematical tools we employ have precursors in related ML domains, though their application and interpretation differ significantly from ours.

Fairness. The foundational theory of topological social choice (Chichilnisky, 1980; Baryshnikov, 1993) has been applied to diagnose impossibility results in fairness. Prior work demonstrates that simultaneously satisfying disparate fairness criteria can be impossible due to Condorcet-like cycles across different metrics. Our work generalizes this insight from fairness to the foundational property of algorithmic stability.

Rank Aggregation. Combinatorial Hodge theory was introduced to the problem of rank aggregation by Lim (2020); Jiang et al. (2008). They utilized the Hodge decomposition to extract a global ranking (the gradient component) from pairwise comparison data and quantify the remaining inconsistencies (the cyclical components). While this provides a powerful tool for analyzing ranking data, our work makes the crucial connection between this mathematical structure and algorithmic stability. We reinterpret the cyclical component as the obstruction that causes instability, thereby providing a unifying framework for understanding stability-enhancing methods like bagging and inflated operators.

3 Stability as Constrained Preference Aggregation

We formalize the learning problem and stability constraints, analyzing how local data preferences can conflict.

3.1 The Learning Setup and Stability

Let $\mathcal{X}$ be the data space. We define the dataset space $\mathcal{D}$ as the set of all training datasets of a fixed size N, drawn i.i.d. from an underlying distribution over $\mathcal{X}$. We view $\mathcal{D}$ as a graph where edges connect *adjacent* datasets, defined as datasets differing by exactly one data point.

Let $\mathcal{H}$ be the hypothesis space, equipped with a metric $d_{\mathcal{H}}$. Let $L : \mathcal{H} \times \mathcal{D} \to \mathbb{R}$ be the loss function. A learning algorithm $\mathcal{A} : \mathcal{D} \to \mathcal{H}$ aims to select $\mathcal{A}(S) \approx \arg\min_{h \in \mathcal{H}} L(h, S)$.

Definition 3.1 (Algorithmic Stability)**.** An algorithm $\mathcal{A}$ is ϵ-stable if for any two adjacent datasets $S, S' \in \mathcal{D}$: $d_{\mathcal{H}}(\mathcal{A}(S), \mathcal{A}(S')) \leq \epsilon$.

3.2 Local Preferences and the Accuracy-Stability Tradeoff

The loss function induces preferences over the hypothesis space for each dataset.

Definition 3.2 (Induced Preference Relation)**.** For a dataset S, we say h_1 is preferred over h_2 (denoted $h_1 >_S h_2$) if $L(h_1, S) < L(h_2, S)$.

To capture the set of viable hypotheses, we define the set of near-optimal solutions.

Definition 3.3 (Near-Optimal Hypothesis Set)**.** For accuracy tolerance $\delta \geq 0$, the δ-optimal hypothesis set for dataset S is: $O_\delta(S) := \{h \in \mathcal{H} | L(h, S) \leq \min_{h' \in \mathcal{H}} L(h', S) + \delta\}$.

The fundamental challenge is finding an algorithm that is simultaneously accurate and stable.

Definition 3.4 (The Stable Selection Problem)**.** Given $\epsilon, \delta \geq 0$, the Stable Selection Problem is to find a map $\mathcal{A} : \mathcal{D} \to \mathcal{H}$ such that for all S, $S' \in \mathcal{D}$:

1. $\mathcal{A}(S) \in O_\delta(S)$ (Accuracy)
2. If S, S' are adjacent, $d_{\mathcal{H}}(\mathcal{A}(S), \mathcal{A}(S')) \leq \epsilon$ (Stability)

To ensure the problem is well-posed, we assume standard regularity conditions.

Assumption 3.5. The dataset space $\mathcal{D}$ (as defined by adjacency) is connected.

This assumption is necessary to analyze the global implications of local stability constraints. If $\mathcal{D}$ were disconnected, an algorithm could be locally stable within each component but globally unstable across components. For the definition of $\mathcal{D}$ given above (fixed size N, differing by one point), this connectivity generally holds.

4 Cyclical Obstructions to Stability

We introduce the mathematical framework used to characterize inconsistencies in preferences. While motivated by deep connections between social choice theory and abstract algebraic topology (Chichilnisky, 1980; Baryshnikov, 1993, see Appendix A), we utilize the concrete machinery of Combinatorial Hodge Theory (Lim, 2020; Jiang et al., 2008). This approach analyzes flows on a graph, providing a rigorous and accessible way to quantify structural inconsistencies (cycles) in the data preferences.

4.1 The Graph of Hypotheses and Preference Flows

We focus on a finite subset of relevant hypotheses $\mathcal{H}_k = \{h_1, \ldots, h_k\}$ (e.g., models in an ensemble or candidates in model selection). We construct a complete graph K where the vertices are the hypotheses $\mathcal{H}_k$.

We analyze the preferences induced by the data over these hypotheses. As introduced earlier, we view the collection of datasets $\mathcal{D}$ as "voters." In practical contexts, such as analyzing an existing ensemble on a test set (as in Experiment 2), individual data points can also serve as the voters, providing an empirical estimate of the preferences.

Definition 4.1 (Preference Profile)**.** A preference profile P is a collection of preference relations $\{>_S\}_{S \in \mathcal{D}}$ over $\mathcal{H}_k$.

We are interested in how these preferences manifest as flows on the edges of the graph K.

Definition 4.2 (Cochains/Flows)**.**

- A 0-cochain is a function assigning values to vertices (e.g., a loss or utility function).
- A 1-cochain (or flow) is a function assigning values to edges (e.g., pairwise comparisons).

The fundamental operator connecting these is the coboundary operator d^0. It transforms a 0-cochain (potential) f into a 1-cochain (flow): $(d^0 f)(h_i, h_j) = f(h_j) - f(h_i)$.

Definition 4.3 (Gradient Flow)**.** A flow C is a *gradient flow* (or coboundary) if it can be derived from a global potential function, i.e., $C = d^0 f$. Gradient flows represent perfectly consistent preferences corresponding to a global ranking.

4.2 Aggregating Preferences: The Importance of Ordinal Data

To analyze stability, we must aggregate the local preferences in the profile P into a single global flow C_P on the graph K. The method of aggregation is critical.

Cardinal Aggregation. If we aggregate the losses (cardinal values) directly, e.g., by averaging the loss differences across datasets: $C_P^{card}(h_i, h_j) = \mathbb{E}_{S \sim \mathcal{D}}[L(h_j, S) - L(h_i, S)]$. This construction *always* results in a gradient flow, as it is the gradient of the aggregate loss function $L_{agg}(h) = \mathbb{E}_S[L(h, S)]$. Therefore, C_P^{card} cannot detect Condorcet cycles.

Ordinal Aggregation (Pairwise Majority Vote). To detect structural inconsistencies analogous to those in social choice theory, we must aggregate the rankings (ordinal preferences). We define the Ordinal Preference Flow C_P^{ord} using Pairwise Majority Vote (PMV): $C_P^{ord}(h_i, h_j) = \mathbb{P}_{S \sim \mathcal{D}}(h_i >_S h_j) - \mathbb{P}_{S \sim \mathcal{D}}(h_j >_S h_i)$. This represents the net preference for h_i over h_j across the dataset distribution. Crucially, C_P^{ord} is generally *not* a gradient flow.

4.3 Hodge Decomposition and Cyclical Obstructions

We use the Hodge decomposition theorem to analyze the structure of the ordinal preference flow C_P^{ord}.

Theorem 4.4 (Hodge Decomposition on Graphs)**.** *Any flow C on the graph K can be uniquely and orthogonally decomposed as: $C = C_{gradient} + C_{cycle}$. Where $C_{gradient}$ is a gradient flow (the consistent part), and C_{cycle} is a divergence-free flow (the cyclical part).*

(See Appendix A.4 for mathematical details and the definition of the divergence operator).

This theorem provides the central tool for our framework. It allows us to quantify the degree to which a preference profile deviates from a global consensus.

Definition 4.5 (Cyclical Obstruction)**.** The component C_{cycle} in the Hodge decomposition of the ordinal preference flow C_P^{ord} is the *cyclical obstruction*. It mathematically characterizes the structural inconsistency (Condorcet cycles) inherent in the preference profile P.

The magnitude $||C_{cycle}||$ quantifies the severity of the obstruction. When $||C_{cycle}|| = 0$, the preferences are perfectly consistent. When $||C_{cycle}|| > 0$, a structural inconsistency exists. In the language of algebraic topology, C_{cycle} corresponds to a non-trivial element in the first cohomology group (H^1).

Example 4.6 (The Condorcet Cocycle)**.** The profile $(A > B, B > C, C > A)$ from Example 1.1, if dominant across the data, results in a C_P^{ord} where the flow is purely cyclical ($C_{gradient} = 0$). This is the canonical example of a structural obstruction.

4.4 Mapping Obstructions to Instability

We now connect this mathematical structure back to the Stable Selection Problem. While the existence of a cyclical obstruction ($||C_{cycle}|| > 0$) indicates inconsistency, instability arises when this inconsistency is sufficiently strong and localized.

Definition 4.7 (Strong Structural Inconsistency)**.** A preference profile P exhibits a strong structural inconsistency with margin $\gamma > 0$ if there exists a cycle of hypotheses $\{h_1, \ldots, h_k\}$ and disjoint regions of the dataset space, $\mathcal{D}_i \subseteq \mathcal{D}$, such that for $S \in \mathcal{D}_i$, h_i is strongly preferred over h_{i+1} (indices modulo k): $L(h_i, S) < L(h_{i+1}, S) - \gamma$.

This condition implies that different regions of the data space have strongly conflicting preferences, forcing an accurate algorithm to choose different hypotheses in different regions.

Proposition 4.8 (Instability Induced by Strong Obstructions)**.** *If the preference profile induced by the data contains a strong structural inconsistency with margin γ, then for sufficiently small accuracy tolerance $\delta < \gamma/2$ and stability parameter ϵ, the Stable Selection Problem is unsolvable.*

Proof. The condition of strong structural inconsistency ensures a separation margin γ. We choose the accuracy tolerance $\delta < \gamma/2$. If $S \in \mathcal{D}_i$, assuming h_i is near-optimal, $L(h_i, S) < L(h_{i+1}, S) - \gamma$. Since $\gamma > 2\delta$, h_{i+1} cannot be in the near-optimal set: $h_{i+1} \notin O_\delta(S)$. This confirms that the optimal sets in different regions are disjoint with respect to the hypotheses in the cycle.

To satisfy the accuracy constraint, an algorithm $\mathcal{A}$ must select $\mathcal{A}(S) \in O_\delta(S)$. The structural inconsistency forces the algorithm to choose between distinct hypotheses in different regions. We demonstrate that stability prevents this.

Intuitive Case: Discrete Metric. Consider the simplified case where $\mathcal{H}$ has the discrete metric ($d(h_a, h_b) = 1$ if $a \neq b$) and $\epsilon < 1$. Stability implies $\mathcal{A}(S) = \mathcal{A}(S')$ for adjacent S, S'. Since $\mathcal{D}$ is connected (Assumption 3.5), this forces a single global choice h^*. However, the strong structural inconsistency (a cycle) implies that for any h^*, there exists a region $\mathcal{D}_i$ where h^* is strongly dispreferred and thus $h^* \notin O_\delta(S)$ for $S \in \mathcal{D}_i$. Thus, the problem is unsolvable.

General Case: General Metric Space. We generalize the argument using the triangle inequality. This requires that the structural inconsistency manifests metrically (standard regularity conditions assumed in Section 3.2).

Let h_i and h_j be two distinct hypotheses in the cycle, separated by distance $D = d_{\mathcal{H}}(h_i, h_j) > 0$. Consider the corresponding regions $\mathcal{D}_i$ and $\mathcal{D}_j$. We assume that for sufficiently small δ, the accuracy constraint implies metric localization. Formally, we assume there exists a δ' (which decreases as δ decreases) such that for $S \in \mathcal{D}_i$, $d_{\mathcal{H}}(\mathcal{A}(S), h_i) < \delta'$, and for $S' \in \mathcal{D}_j$, $d_{\mathcal{H}}(\mathcal{A}(S'), h_j) < \delta'$. We choose δ small enough such that $2\delta' < D/2$.

Since $\mathcal{D}$ is connected, there exists a finite path of adjacent datasets $S_0, S_1, \ldots, S_M$, where $S_0 \in \mathcal{D}_i$ and $S_M \in \mathcal{D}_j$.

By the stability constraint, $d_{\mathcal{H}}(\mathcal{A}(S_k), \mathcal{A}(S_{k+1})) \leq \epsilon$. By the triangle inequality along the path:

$$d_{\mathcal{H}}(\mathcal{A}(S_0), \mathcal{A}(S_M)) \leq \sum_{k=0}^{M-1} d_{\mathcal{H}}(\mathcal{A}(S_k), \mathcal{A}(S_{k+1})) \leq M \cdot \epsilon.$$

Now we relate this to the separation D using the triangle inequality again:

$$\begin{aligned} D = d_{\mathcal{H}}(h_i, h_j) &\leq d_{\mathcal{H}}(h_i, \mathcal{A}(S_0)) + d_{\mathcal{H}}(\mathcal{A}(S_0), \mathcal{A}(S_M)) + d_{\mathcal{H}}(\mathcal{A}(S_M), h_j) \\ &< \delta' + (M \cdot \epsilon) + \delta'. \end{aligned}$$

This gives the requirement $D < 2\delta' + M \cdot \epsilon$.

The distance D and path length M are fixed by the problem structure. We have chosen δ such that $2\delta' < D/2$. We can now choose ϵ sufficiently small such that $M \cdot \epsilon \leq D/2$. Then $2\delta' + M \cdot \epsilon < D$. This violates the derived inequality, creating a contradiction. Thus, the Stable Selection Problem is unsolvable. □

When a strong cyclical obstruction exists, no single-valued algorithm can simultaneously satisfy the inconsistent preferences and the metric constraints of stability.

5 A Mathematical Explanation for Stability Methods

Our framework, centered on the Hodge decomposition, rigorously distinguishes between two sources of instability and unifies the methods used to address them based on the presence or absence of structural obstructions (C_{cycle}).

5.1 Addressing Structural Obstructions ($C_{cycle} > 0$)

When the data preferences contain fundamental inconsistencies, the instability is structural. This requires methods that either resolve the ambiguity or enforce consistency.

5.1.1 Obstruction Resolution: Target Space Enlargement

When obstructions are inherent ($C_{cycle} \neq 0$), a stable single-valued solution may not exist (Proposition 4.13). The mathematical solution is to change the target space.

In social choice theory, when a single consensus winner does not exist due to cycles, the solution is to identify a consensus *set*, such as the Top Trading Cycle (TTC) set—the minimal set of hypotheses H^* that dominate those outside H^*. Inflated operators, rigorously analyzed by Adrian et al. (2024) and Liang et al. (2025), implement this principle by enlarging the target space from $\mathcal{H}$ to $\mathcal{P}(\mathcal{H})$.

Definition 5.1 (Inflated Argmax (Adrian et al., 2024))**.** $argmax_\epsilon(w) := \{j | w_j \geq max_k w_k - \epsilon\}$.

Proposition 5.2 (Cohomological Significance of Inflation)**.** *Inflated operators resolve cyclical obstructions by systematic target space enlargement. The set-valued output of the inflated operator approximates the Top Trading Cycle set associated with the underlying preference cycle.*

Proof. Consider a profile P with a significant cyclical obstruction, represented by a cycle $h_1 > \cdots > h_k > h_1$. As shown in Proposition 4.8, if this obstruction is strong, it leads to the failure of the single-valued Stable Selection Problem.

The Top Trading Cycle (TTC) set associated with this cycle is $H^* = \{h_1, ..., h_k\}$. Now consider the application of the inflated operator (e.g., returning the near-optimal set $O_\delta(S)$). If the cycle is balanced, the losses of the hypotheses in the cycle will be close across the relevant region of the dataset space: $|L(h_i, S) - L(h_j, S)| \leq \epsilon'$ for $h_i, h_j \in H^*$.

If we choose the tolerance parameter $\delta \geq \epsilon'$, the near-optimal set $O_\delta(S)$ will contain H^*. If hypotheses outside the cycle have significantly higher loss, then $O_\delta(S) \approx H^*$.

The inflated operator thus identifies the set of hypotheses involved in the structural ambiguity characterized by the obstruction. The stability of this approach relies on the established result (Adrian et al., 2024) that for nearby datasets S, S' (under bounded loss variations), their near-optimal sets have a non-empty intersection: $O_\delta(S) \cap O_\delta(S') \neq \emptyset$. This ensures the set-valued output remains consistent under perturbation, resolving the obstruction by changing the requirements on the output. □

5.1.2 Obstruction Prevention: Enforcing Consistency

Alternatively, we can attempt to prevent obstructions by enforcing consistency during the learning process. Regularization-based ensemble stability adds penalty terms that discourage disagreement (e.g., $\mathcal{L}_{glue} = \sum_{i,j} ||h_i(x) - h_j(x)||^2$). This explicitly enforces consensus among the ensemble members. By minimizing $\mathcal{L}_{glue}$, the optimization procedure forces the local preferences to align. When models are forced to produce similar outputs, their rankings across the data become more consistent, directly minimizing the inconsistencies that contribute to the cyclical component C_{cycle}. Our proposed method in Section 6.1 is a higher-order generalization of this approach.

5.2 Addressing Statistical Variance ($C_{cycle} \approx 0$)

When the cyclical component is near zero, the aggregated preferences are globally consistent. However, instability can still occur due to statistical variance—the sensitivity of the learning algorithm or the selection procedure (like argmax) to noise, near-ties, or specific data samples.

5.2.1 Bagging: Variance Reduction

Bagging (Breiman, 1996) utilizes bootstrap samples $\{S_i\}$ from S and aggregates the results. Its effectiveness is statistically grounded in variance reduction (Soloff et al., 2024).

Interpretation 5.3. Role of Bagging in the Hodge Framework *Bagging is a mechanism for reducing statistical instability when the underlying preference structure is largely consistent ($C_{cycle} \approx 0$). It is not designed to resolve strong structural obstructions ($C_{cycle} \neq 0$).*

Argument. Let $V_{ij}(S) = L(h_j, S) - L(h_i, S)$ be the cardinal preference score. When $C_{cycle} \approx 0$, the ordinal preferences are largely aligned with a global ranking. Instability occurs when $V_{ij}(S)$ has high variance, causing the sign of $V_{ij}(S)$ (the ordinal preference) to flip frequently across the dataset distribution, especially in near-ties.

The bagged learner $\mathcal{A}_{bag}$ smooths the cardinal preferences, resulting in $V_{ij}^{bag}(S)$ with significantly reduced variance. This stabilization makes the ordinal preferences more robust and less likely to fluctuate, thereby improving stability.

However, if a strong structural obstruction exists ($C_{cycle} \neq 0$), the data inherently contains conflicting preferences. Variance reduction alone cannot create a stable consensus, as the underlying problem is fundamentally ambiguous. Bagging averages over the conflicting preferences but does not resolve the cycle itself. As we demonstrate in Experiment 7.1, bagging fails to stabilize the selection process in the presence of a strong Condorcet cycle.

6 Obstruction-Aware Ensembling: A Framework Inspired by Cohomological Insights

Our cohomological perspective suggests principled approaches to designing stable algorithms by systematically targeting these mathematical structures.

6.1 Cohomologically-Inspired Regularization

Motivated by the mathematical structure of disagreement cycles, we propose regularization terms that target higher-order inconsistencies, moving beyond pairwise agreement. Consider an ensemble with models $\{h_i\}$ and learnable alignment maps $\{\phi_{ji}\}$ between their representation spaces. Global consistency requires the alignment maps to satisfy the cocycle condition: $\phi_{ki} = \phi_{kj} \circ \phi_{ji}$.

Definition 6.1 (Obstruction-Aware Ensemble Loss)**.** A cohomologically-inspired loss includes a term enforcing higher-order consistency: $\mathcal{L}_{cocycle} = \sum_{i,j,k} ||\phi_{ki} - \phi_{kj} \circ \phi_{ji}||_F^2$ The total loss is $\mathcal{L}_{total} = \mathcal{L}_{task} + \lambda_1 \mathcal{L}_{glue} + \lambda_2 \mathcal{L}_{cocycle}$.

Minimizing $\mathcal{L}_{cocycle}$ encourages global consistency in the geometric relationships between models, directly targeting the structures identified by the framework in the representation space.

Computational Complexity and Stochastic Approximation. The proposed $\mathcal{L}_{cocycle}$ involves a summation over all triplets (i, j, k), resulting in a computational complexity of $O(N^3)$ for an ensemble of size N. This can be computationally prohibitive for very large ensembles. However, this complexity can be readily mitigated using a stochastic approximation. Instead of calculating the loss over all triplets at every step, we can randomly sample a fixed number of triplets K and calculate the average loss:

$$\mathcal{L}_{\text{stochastic-cocycle}} = \frac{1}{K} \sum_{(i,j,k) \in \text{Samples}} \|\phi_{ki} - \phi_{kj} \circ \phi_{ji}\|_F^2$$

This reduces the complexity to $O(K)$ per training step and provides an unbiased estimator of the average cocycle defect. As we demonstrate in Section 7.4, this stochastic approach effectively enforces the cocycle constraint while maintaining computational tractability.

Connection to Cycle Consistency. The mathematical form of $\mathcal{L}_{cocycle}$ is analogous to the 'Cycle-Consistency Loss' widely used in unsupervised learning domains like image translation and unsupervised machine translation Grover et al. (2020). In those domains, cycle consistency serves as an engineering heuristic to constrain ill-posed, unsupervised problems. Our work re-contextualizes this mechanism: we derive the cocycle constraint from a theoretical framework, arguing that violations of this consistency are manifestations of a mathematical obstruction with direct consequences for stability. This elevates the mechanism from a heuristic to a principled, theoretically-grounded tool for improving robustness.

6.2 Adaptive Target Space Enlargement

Inspired by the success of inflated operators, we propose adaptive mechanisms that automatically transition between single-valued outputs when consensus exists (suggesting $C_{cycle} \approx 0$) and set-valued outputs when obstructions may be present ($C_{cycle} \neq 0$), based on measures of ensemble disagreement.

7 Computational Experiments

We conduct five experiments to validate our theoretical framework. To rigorously validate the core hypothesis—that structural inconsistencies (Condorcet cycles) cause instability and that cohomologically-inspired methods resolve them—it is necessary to operate in controlled settings. Real-world benchmark datasets often undergo a selection bias; they are typically chosen because standard methods generalize well on them, implying they may inherently lack the strong structural inconsistencies that cause severe instability. Therefore, we utilize synthetic data (Experiments 1, 3, 4) and conflict-engineered standard data (Experiment 2) to precisely engineer or isolate these obstructions. Finally, Experiment 5 demonstrates that these structural obstructions emerge naturally in fairness-constrained model selection on the UCI Adult dataset, confirming the framework's applicability to real-world scenarios involving conflicting objectives. These experiments are essential for demonstrating the framework in action and confirming the underlying mechanisms. (Full details and code are provided in Appendix B).

7.1 Experiment 1: Structural Instability (Engineered Condorcet Cycle)

This experiment validates the core distinction of our framework: structural instability requires structural solutions (inflation), while statistical methods (bagging) fail. We test this in a scenario with a pure structural obstruction.

Setup: We engineered a scenario with three hypotheses (A, B, C) where the underlying data distribution is a mixture of three regions (Z_1, Z_2, Z_3) inducing a balanced Condorcet cycle ($A > B > C > A$). We simulate the process of drawing a dataset S and an adjacent dataset S' (by perturbing the sampling weights by a small amount). We ran 5000 trials.

Methods: We compare the stability of:

1. Standard 'argmax'.
2. Inflated 'argmax' ($\epsilon = 0.01$) (Structural solution).
3. Bagged 'argmax' (50 bootstraps) (Statistical solution).

Metrics: Agreement Rate (probability the winner is the same for S and S') for single-valued methods; Intersection Consistency (probability the output sets intersect) for the inflated method.

Table 1: Experiment 1: Stability under a pure structural obstruction (Condorcet cycle).

Method	Type	Stability (Agreement/Intersection)	Avg. Output Size
Standard Argmax	Baseline	0.4094	1.00
Bagged Argmax	Statistical	0.3432	1.00
Inflated Argmax	Structural	1.0000	3.00

Results: As shown in Table 1, the standard 'argmax' exhibits low stability (0.4094). Crucially, the Bagged 'argmax' also fails to stabilize the selection, performing slightly worse (0.3432). This confirms Interpretation 5.3: statistical variance reduction is ineffective against structural instability. In contrast, the Inflated 'argmax' achieves perfect stability (1.0000) by identifying the ambiguity inherent in the cycle and returning the consensus set {A, B, C}.

7.2 Experiment 2: Diagnosing Instability via Hodge Decomposition

This experiment utilizes the Hodge decomposition as a diagnostic tool to determine the source of instability in standard machine learning scenarios and verifies the effectiveness of bagging when instability is statistical.

Setup: We analyze instability induced by:

- **Scenario A (Synthetic Noise):** A complex synthetic dataset (high noise/low separation). Baseline: unstable Decision Tree (DT) trained on small subsamples.
- **Scenario B (Real Data Heterogeneity - DT):** Forest Covertype dataset. Baseline: "Specialist Ensemble" of DTs trained only on specific data partitions (Wilderness Areas) to induce conflict.
- **Scenario C (Real Data Heterogeneity - 1-NN):** Forest Covertype dataset. Baseline: Specialist Ensemble using 1-Nearest Neighbor (1-NN), a canonically unstable algorithm.

Methodology: In all scenarios, we trained $K = 20$ hypotheses over multiple trials. We applied the Hodge decomposition to the Ordinal Preference Cochain. We measure the cyclical norm $||C_{cycle}||$ and the empirical instability (average disagreement rate between models; lower is better).

Results and Interpretation: The results are summarized in Table 2. In all scenarios, the measured cyclical norms $||C_{cycle}||$ are extremely small (10^{-15} to 10^{-17}). These values are at the level of machine precision and are indistinguishable from zero. Crucially, the baseline models exhibit high empirical instability across all scenarios. The coexistence of high instability and zero cyclical

Table 2: Experiment 2: Diagnosis of instability. High instability coexists with near-zero cyclical norm, indicating the source is statistical variance, which bagging successfully reduces.

			Baseline (Unstable)		Bagged (Stable)	
ID	Data	Algorithm (Source)	Instability ↓	$\|\|C_{cycle}\|\|$	Instability ↓	$\|\|C_{cycle}\|\|$
A	Synth.	DT (Noise)	0.6210	2.40×10^{-16}	0.5122	1.19×10^{-16}
B	Forest	DT (Specialist)	0.4610	1.98×10^{-15}	0.1963	3.26×10^{-17}
C	Forest	1-NN (Specialist)	0.5025	1.99×10^{-15}	0.0892	1.54×10^{-17}

norm rigorously diagnoses the source of instability as statistical variance rather than structural obstruction.

As predicted by our framework (Interpretation 5.3), bagging successfully reduces this statistical instability across all scenarios (e.g., from 0.4610 to 0.1963 in Scenario B; from 0.5025 to 0.0892 in Scenario C). This contrasts sharply with Experiment 1, where bagging failed because the instability was structural. These findings highlight the diagnostic power of the framework: it allows us to differentiate the source of instability and correctly predict the efficacy of stabilization methods.

7.3 Experiment 3: Validation of Obstruction-Aware Regularization

We validate the cohomologically-inspired regularization $\mathcal{L}_{cocycle}$ proposed in Section 6 by demonstrating a direct correspondence between enforced mathematical consistency and improved algorithmic stability.

Setup: We constructed a system of three interconnected 2D representation spaces with learnable linear transformations ϕ_{ji}. We engineered conflicting ground truth transformations (using rotations of 60°, 60°, and −30° for ϕ_{10}, ϕ_{21}, ϕ_{20} respectively). This results in a 150° error on the loop ($\phi_{21} \circ \phi_{10} \neq \phi_{20}$), simulating an engineered structural inconsistency (a non-trivial obstruction) in the representation space. We compared standard training (using only pairwise alignment loss) with framework-guided training (adding $\mathcal{L}_{cocycle}$ with weight λ_2).

Methodology: To robustly assess the correspondence between mathematical consistency and stability, we performed a comprehensive parameter sweep across various levels of data noise (0.01 to 0.5) and regularization strengths (λ_2 from 1.0 to 20.0), running 8 independent trials per configuration. We measured the Cocycle Defect (the mathematical inconsistency) and several stability metrics based on the variance across trials: Parameter Variance (consistency of learned weights), Prediction Variance (consistency of outputs), Functional Variance (stability of the behavior of composed maps), and Transformation Consistency (average distance between learned transformations across runs).

Results: The parameter sweep revealed a strong correspondence between obstruction reduction and stability improvement under moderate conditions. The configuration demonstrating the best overall correspondence (Noise 0.1, $\lambda_2 = 5.0$) is detailed in Table 3. Enforcing the mathematical constraints significantly reduced the cocycle defect (24.9x). Crucially, this correlated strongly with improvements across all four key stability metrics, ranging from 1.5x (Functional Variance) to 2.4x (Parameter Variance). This strongly supports the central claim of this work: targeting the math-

Table 3: Experiment 3: Effect of obstruction-aware regularization (Best correspondence configuration: Medium noise, $\lambda_2 = 5.0$).

Metric	Standard training	Framework-guided	Improvement
Mathematical Consistency			
Cocycle defect	2.738	0.110	24.9x reduction
Algorithmic Stability			
Parameter variance	0.000418	0.000177	2.4x improvement
Prediction variance	0.000159	0.000088	1.8x improvement
Transformation consistency	0.028042	0.017872	1.6x improvement
Functional variance	0.000085	0.000056	1.5x improvement

ematical structures identified by obstruction theory provides a principled and effective mechanism for improving algorithmic stability.

7.4 Experiment 4: Scalability of Stochastic Cocycle Regularization

This experiment addresses the $O(N^3)$ computational complexity of the proposed $\mathcal{L}_{cocycle}$ (Section 6.1) by validating the effectiveness and efficiency of its stochastic approximation.

Setup: To highlight the scalability challenge, we increased the ensemble size to $N = 15$ (455 total triplets) and initialized the transformation maps ϕ_{ji} randomly, creating a high initial cocycle defect. We trained the system using only the regularization loss to isolate its effect. We compared three scenarios: (1) Full Batch $O(N^3)$; (2) Stochastic Approximation, sampling $K = 50$ triplets per step $O(K)$; (3) Baseline (no regularization).

Metrics: We measured the reduction of the True Cocycle Defect over epochs (Convergence Rate) and wall-clock time (Computational Efficiency).

Results: The results (detailed in Appendix B.4 confirm the validity of the stochastic approach. The Baseline maintains a high defect ($\approx 10^1$). Both the Full Batch and Stochastic approaches significantly reduce the defect. Crucially, the Stochastic approach (4.68s) achieved a 3.34x speedup over the Full Batch approach (15.66s). While the Full Batch method reaches near machine precision (10^{-10}), the Stochastic method effectively reduces the defect by five orders of magnitude (10^{-5}), demonstrating its efficacy as a regularizer. This illustrates the expected trade-off: the stochastic approximation provides a scalable and practical solution for enforcing mathematical consistency in large ensembles.

7.5 Experiment 5: Structural Obstructions in Fairness-Constrained Model Selection

This experiment demonstrates that structural obstructions—absent in standard classification tasks—emerge naturally in real-world fairness-constrained model selection, where incompatible objectives create genuine Condorcet cycles.

Motivation: While Experiments 1-4 showed that $||C_{cycle}|| \approx 0$ in standard tasks, the framework predicts that structural obstructions should arise when stakeholders impose conflicting constraints. Fairness-aware machine learning provides a natural testbed: it is well-established that different fair-

ness metrics are mathematically incompatible and cannot be simultaneously optimized (Kleinberg et al., 2016). Our framework characterizes this incompatibility as a topological obstruction.

Setup: We used the UCI Adult dataset (Dua & Graff, 2019) to simulate fairness-constrained selection. We trained five Logistic Regression models, employing different heuristics (e.g., reweighting, regularization adjustments) to generate a diverse set of models exhibiting varied fairness-accuracy trade-offs. We emphasize that these heuristics are not intended to optimally maximize specific fairness metrics, but rather to simulate the pool of candidate models available for selection. We then defined five simulated stakeholder perspectives (Business, Civil Rights, Equal Opportunity, Calibration, Regulator), each prioritizing different combinations of Accuracy, Demographic Parity (DP), Equalized Odds (EO), and Predictive Parity (PP). Crucially, the structural obstruction arises from these conflicting stakeholder preferences over the candidate models, not the specific training procedures. (See Appendix B.5 for details).

Methodology: For each of 20 trials with different train-test splits, we:

1. Trained the five models with different fairness-accuracy tradeoffs.
2. Computed stakeholder preferences based on model performance.
3. Constructed the ordinal preference flow C_P^{ord} via Pairwise Majority Vote (PMV).
4. Applied Hodge decomposition to measure $||C_{cycle}||$ and the Cycle Ratio ($||C_{cycle}||/||C_{total}||$).

Table 4: Experiment 5: Structural obstructions in fairness model selection (UCI Adult). Fairness problems exhibit cyclical norms 15 orders of magnitude larger than standard classification tasks.

Task Type	$\|\|C_{cycle}\|\|$ (mean)	Cycle Ratio (mean)	vs. Standard Tasks
Standard Classification (Exp. 2)	$\sim 10^{-16}$	~ 0	Baseline
Fairness Model Selection (Adult)	0.857 ± 0.173	0.363 ± 0.075	$\sim 10^{15}\times$ larger

Results and Interpretation: The results, summarized in Table 4, reveal a striking contrast with standard classification tasks. The mean cyclical norm is $||C_{cycle}|| = 0.857$, approximately 10^{15} times larger than the machine-precision values observed in Experiment 2. The mean cycle ratio is 0.363, indicating that over 36% of the preference disagreement is fundamentally structural rather than statistical.

Theoretical Significance: This validates the core distinction of our framework on real data. Standard learning problems produce negligible structural obstructions ($C_{cycle} \approx 0$), explaining the effectiveness of bagging. In contrast, fairness-constrained selection involves genuinely conflicting objectives that create strong cyclical obstructions ($C_{cycle} \neq 0$), making single-model selection fundamentally unstable.

Topological Characterization of Fairness Impossibility: This provides a topological characterization of known fairness impossibility theorems (Kleinberg et al., 2016). The incompatibility between fairness definitions manifests as non-trivial cohomology ($H^1 \neq 0$) in the preference space. Our contribution is not discovering that fairness metrics conflict—this is well-established—but providing the first rigorous quantification of this conflict via topological invariants ($||C_{cycle}||$), unifying

these impossibilities with the broader theory of algorithmic stability and social choice theory under a common mathematical framework.

Practical Implications: When $C_{cycle} \neq 0$, Proposition 4.8 establishes that stable, single-valued selection is impossible. The appropriate structural solution (Proposition 5.2) is target space enlargement via inflated operators, returning a *portfolio* of models (the Pareto frontier). Crucially, this does not "solve" the fairness conflict; it merely acknowledges the impossibility of a single solution characterized by the cycle norm, and externalizes the decision to policymakers. The framework clarifies that the challenge is structural and cannot be resolved by purely statistical means.

8 Limitations and Future Directions

While our framework provides a unifying perspective on algorithmic stability, it has several limitations that suggest promising directions for future research.

8.1 Limitations

Continuous Hypothesis Spaces: The framework of topological social choice theory and Hodge decomposition is primarily developed for a finite set of alternatives $\mathcal{H}_k$. Extension to continuous hypothesis spaces requires additional technical machinery, such as analysis on function spaces or discretization schemes that preserve the relevant topological structure.

Theoretical Scope and Empirical Findings: Our framework characterizes two distinct mathematical sources of instability: structural obstructions (cyclical inconsistencies) and statistical variance. The empirical findings of Experiment 7.2—that $||C_{cycle}|| \approx 0$ in standard classification tasks—constitute a significant theoretical result rather than a limitation. They explain mathematically why the field has empirically converged on variance reduction methods: the structural obstructions that would defeat such methods simply do not arise in well-posed supervised learning problems with convex or near-convex loss surfaces and i.i.d. data. The cases where $C_{cycle} \neq 0$ represent important boundary conditions where fundamentally different mathematical structures emerge. We provide concrete evidence for this in Experiment 7.5, demonstrating that significant structural obstructions arise in fairness scenarios due to incompatible metrics. Investigating the precise conditions under which structural obstructions manifest in other domains (e.g., adversarial settings, multi-agent systems), and their prevalence, remains an important direction for future theoretical and empirical work under which structural obstructions manifest, and their prevalence in these specialized domains, remains an important direction for future theoretical and empirical work.

Computational Scalability of Analysis: Calculating the Hodge decomposition for theoretical analysis (as in Experiment 7.2) requires operations such as computing the pseudoinverse of the graph Laplacian, which scale as $O(K^3)$ where K is the number of hypotheses. This limits the direct application of the decomposition as an online diagnostic tool for very large hypothesis spaces. However, the proposed obstruction-aware regularization (Section 6.1), which is the primary practical contribution, admits efficient stochastic approximation with $O(N)$ complexity per step, as demonstrated in Section 7.4.

Connection Between Ordinal and Cardinal Structures: While we rigorously distinguish between ordinal aggregation (which detects cycles) and cardinal aggregation (which does not), the

precise relationship between high variance in cardinal preferences and the emergence of near-cycles in ordinal preferences remains to be fully characterized. A deeper understanding of this connection would strengthen the theoretical link between variance reduction (bagging) and cycle suppression.

8.2 Future Directions

Algorithmic Applications: Further development of obstruction-aware algorithms. This includes designing efficient methods to estimate when $C_{cycle} \neq 0$ during training and adaptively deploying regularization (structural) or bagging (statistical) strategies based on the diagnosed source of instability.

Theoretical Extensions: Investigating the precise mathematical relationship between the smoothing effect of bagging (variance reduction) and the reduction of the cyclical component in the Hodge decomposition. Furthermore, exploring the use of higher-order cohomology groups (H^q for $q \geq 2$) to identify more complex structural inconsistencies beyond pairwise cycles.

Connections to Fairness and Explainability: Preference cycles often underlie paradoxes in fairness and explainability, where different metrics or perspectives lead to conflicting conclusions. Extending this cohomological framework to analyze the stability and consistency of fairness metrics is a promising direction.

9 Conclusion

We have presented a unified framework for algorithmic stability based on Combinatorial Hodge Theory and the geometry of preferences. This framework rigorously distinguishes between structural instability, caused by cyclical obstructions (Condorcet cycles, $C_{cycle} > 0$), and statistical instability, caused by variance ($C_{cycle} \approx 0$). We demonstrated that structural instability requires targeted solutions like inflated operators or novel obstruction-aware regularization, while statistical instability is addressed by methods like bagging. Our experiments validate this distinction, showing that bagging fails in the presence of strong structural obstructions but succeeds when instability is statistical. This framework provides a mathematical foundation for understanding and improving the robustness of machine learning systems.

References

Adrian, M., Soloff, J. A., and Willett, R. Stabilizing black-box model selection with the inflated argmax. *arXiv preprint arXiv:2410.18268*, 2024.

Barannikov, S., Voronkova, D., Trofimov, I., Korotin, A., Sotnikov, G., and Burnaev, E. Topological obstructions in neural networks' learning. *arXiv preprint arXiv: 2012.15834v2*, 2022.

Baryshnikov, Y. M. Unifying impossibility theorems: A topological approach. *Advances in Applied Mathematics*, 14(4):404-415, 1993.

Bousquet, O. and Elisseeff, A. Stability and generalization. *Journal of Machine Learning Research*, 2(Mar):499-526, 2002.

Breiman, L. Bagging predictors. *Machine Learning*, 24(2):123-140, 1996.

Chichilnisky, G. Social choice and the topology of spaces of preferences. *Advances in Mathematics*, 37(2):165-176, 1980.

Dua, D. and Graff, C. UCI machine learning repository, 2019. URL `http://archive.ics.uci.edu/ml`.

Grover, A., Chute, C., Shu, R., Cao, Z., and Ermon, S. Alignflow: Cycle consistent learning from multiple domains via normalizing flows. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(4):4028-4035, 2020.

Hensel, F., Moor, M., and Rieck, B. A Survey of Topological Machine Learning Methods. *Frontiers in Artificial Intelligence*, 4:681108, 2021.

Jiang, X., Lim, L.-H., Yao, Y., and Ye, Y. Learning to rank with combinatorial Hodge theory. *arXiv preprint arXiv:0811.1067*, 2008.

Kleinberg, J., Mullainathan, S., and Raghavan, M. Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*, 2016.

Lim, L.-H. Hodge Laplacians on graphs. *SIAM Review*, 62(3):685–715, 2020

Liang, R., Soloff, J. A., Barber, R. F., and Willett, R. Assumption-free stability for ranking problems. *arXiv preprint arXiv: 2506.02257*, 2025.

Soloff, J. A., Barber, R. F., and Willett, R. Bagging provides assumption-free stability. *Journal of Machine Learning Research*, 25(131):1-35, 2024.

Wasserman, L. Topological data analysis. *Annual Review of Statistics and Its Application*, 5:501–532, 2018.

Zia, A., Khamis, A., Nichols, J., et al. Topological deep learning: a review of an emerging paradigm. *Artificial Intelligence Review*, 57, 77, 2024.

A Appendix A: Mathematical Foundations: Cohomology of Preferences

This appendix provides the rigorous mathematical details underlying the connection between social choice theory, algebraic topology, and the cohomological obstructions discussed in the main text. We adapt the framework established by Chichilnisky (1980) and Baryshnikov (1993), and incorporate tools from Hodge theory for ranking.

A.1 The Geometry of Rankings and Preferences

Let $\mathcal{H}_k = \{h_1, ..., h_k\}$ be a finite set of hypotheses.

Definition A.1 (Space of Preferences)**.** A utility function $U : \mathcal{H}_k \to \mathbb{R}$ induces preferences. The space of utility functions is $\mathcal{U} = \mathbb{R}^k$. A preference profile is a collection of utility functions $\{U_S\}_{S \in \mathcal{D}}$. Alternatively, we can consider the space of strict linear orders $\mathcal{P}$.

A.2 The Source of Obstructions: Topological Social Choice

The motivation for analyzing preference cycles comes from Topological Social Choice Theory (Chichilnisky, 1980; Baryshnikov, 1993). This theory analyzes the topology of the *space of all possible preferences* ($\mathcal{P}$) to determine if fair aggregation rules $F : \mathcal{P}^N \to \mathcal{P}$ exist.

A.2.1 Topological Interpretation (Chichilnisky's Approach)

Chichilnisky (1980) showed that the impossibility of fair aggregation (satisfying certain axioms like Arrow's conditions) stems from the topological properties of $\mathcal{P}$. To define a continuous aggregation rule, $\mathcal{P}$ must be equipped with a topology. The standard approach identifies the space of preferences $\mathcal{P}$ with a subset of the space of utility functions $\mathcal{U} = \mathbb{R}^k$ (e.g., normalized utility vectors). The topology on $\mathcal{P}$ is then induced by the standard Euclidean metric on $\mathcal{U}$. This means preferences are considered "close" if their utility values are close.

Theorem A.2 (Chichilnisky, 1980)**.** *If a continuous aggregation rule $F : \mathcal{P}^N \to \mathcal{P}$ exists satisfying certain desirable axioms (e.g., anonymity, unanimity), then $\mathcal{P}$ must be contractible.*

A space is contractible if it has no topological "holes" (i.e., its homology and cohomology groups are trivial).

A.2.2 Cohomology of Preference Spaces

We analyze the topology of the space of strict linear orders $\mathcal{S}_k$. This space is known to be non-contractible for $k \geq 3$.

The Condorcet cycle ($A > B, B > C, C > A$) represents a loop in the space of preferences. This loop cannot be contracted to a point while maintaining the structure of the preferences.

Lemma A.3. *The existence of a non-contractible loop in the space of preferences corresponds to a non-trivial element in the first cohomology group $H^1(\mathcal{S}_k)$.*

This abstract theory shows that the possibility of Condorcet cycles arises because the space of preferences itself has a non-trivial topology (it is not contractible). This motivates the search for these structures in specific preference profiles.

A.3 Connecting Topology to Stability

While the abstract theory analyzes the space of all preferences, in machine learning we analyze a specific profile P induced by the data. Combinatorial Hodge Theory provides the tools to detect the manifestation of these topological obstructions within a specific profile by analyzing flows on the graph of hypotheses K.

A.4 Algebraic Tools: Cochains and Hodge Decomposition

This section provides the rigorous definitions for the tools used in Section 4.

A.4.1 Cochains and the Coboundary Operator

Let K be the 1-skeleton (the complete graph) on the vertices $\mathcal{H}_k$.

- $C^0(K;\mathbb{R})$: 0-cochains (functions assigning values to vertices, i.e., utility functions).
- $C^1(K;\mathbb{R})$: 1-cochains (functions assigning values to edges, i.e., pairwise comparisons or flows).

The coboundary operator $d^0 : C^0 \to C^1$ maps a potential function $f \in C^0$ to a flow: $(d^0 f)(h_i, h_j) = f(h_j) - f(h_i)$. This represents a perfectly consistent ranking.

A.4.2 Preference Cochain Construction: Cardinal vs. Ordinal

We define the preference cochain $C_P \in C^1(K, \mathbb{R})$. The construction method is critical, as detailed in Section 4.2.

Cardinal Aggregation: If we aggregate the losses directly (e.g., average loss difference): $C_P^{card}(h_i, h_j) = \int_{\mathcal{D}} (L(h_j, S) - L(h_i, S)) d\mu(S)$. This construction always results in a coboundary, as it is the gradient of the aggregate loss function $L_{agg}(h) = \int L(h, S) d\mu(S)$. Therefore, C_P^{card} cannot detect Condorcet cycles.

Ordinal Aggregation (Pairwise Majority Vote): To detect cohomological obstructions relevant to social choice theory, we must aggregate ordinal preferences. Let R_S be the ranking induced by $L(\cdot, S)$. $C_P^{ord}(h_i, h_j) = \int_{\mathcal{D}} sign(R_S(h_j) - R_S(h_i)) d\mu(S)$. In practice (Experiment 7.2), this is calculated as: (Proportion of voters preferring h_i over h_j) - (Proportion preferring h_j over h_i). This construction allows for C_P to be non-coboundary.

A.4.3 Hodge Decomposition

We use the Hodge decomposition theorem to analyze flows (cochains) on the graph K.

Theorem A.4 (Hodge Decomposition on Graphs)**.** *The space of 1-cochains C^1 can be orthogonally decomposed as: $C^1 = im(d^0) \oplus ker(\delta^1)$ where δ^1 is the adjoint of d^0 (the divergence operator).*

Here, $im(d^0)$ is the space of coboundaries (gradient flows, $C_{gradient}$). $ker(\delta^1)$ is the space of harmonic cochains (cyclical flows, C_{cycle}). This harmonic component corresponds to the cohomological obstruction identified in the topological framework.

This theorem allows us to rigorously decompose a preference profile C_P into its consistent part ($C_{gradient}$) and its cyclical obstruction (C_{cycle}), providing the diagnostic tool used in Section 7.

B Appendix B: Experimental Details and Reproducibility

This appendix provides the detailed configurations and methodologies for the experiments presented in Section 7, ensuring reproducibility. We utilized Python with the following key libraries: NumPy, SciPy, scikit-learn (including `fetch_covtype`), and PyTorch.

B.1 Experiment 1: Structural Instability

Objective: Validate that structural instability (Condorcet cycles) is resolved by inflation but not by bagging.

Setup Details:

- Hypotheses: $K = 3$ (A, B, C).
- Data Regions (defining the cycle):
  - Z_1 (prefers $A > B > C$): Utility vector $U_1 = [3.0, 2.0, 1.0]$.
  - Z_2 (prefers $B > C > A$): Utility vector $U_2 = [1.0, 3.0, 2.0]$.
  - Z_3 (prefers $C > A > B$): Utility vector $U_3 = [2.0, 1.0, 3.0]$.
- Simulation: 5000 trials.
- Dataset Generation: A dataset S is simulated by sampling weights W_S for the three regions around a base weight of 100 (with uniform noise $[-0.5, 0.5]$).
- Perturbation: An adjacent dataset S' is generated by perturbing one weight in W_S by ± 1.
- Argmax: Standard argmax with random tie-breaking.
- Inflated Argmax: $\epsilon = 0.01$.
- Bagged Argmax: Simulated by drawing $B = 50$ bootstrap samples (using a multinomial distribution based on W_S) and aggregating the winners by majority vote (with random tie-breaking).

B.2 Experiment 2: Diagnosing Instability via Hodge Decomposition

Objective: Diagnose the source of instability (structural vs. statistical) using Hodge decomposition and verify stability improvements.

General Methodology (All Scenarios):

- Hypotheses (Candidates): $K = 20$.
- Voters: Test data points (50% split).
- Analysis (Structural): Ordinal Preference Cochain (PMV) based on 0-1 loss. Hodge decomposition implemented using the numerically stable least-squares solution (`numpy.linalg.lstsq`) to find the potential function f that minimizes $||C - d^0 f||$.
- Stability Metric (Instability): Average disagreement rate. Calculated as the average fraction of models whose prediction disagrees with the majority vote (mode) across the test set.

Setup Scenario A (Synthetic Noise):

- Data Generation: `sklearn.datasets.make_classification`. Samples: 5000. High noise (flip_y=0.45), low separation (class_sep=0.5).
- Base Learner: Deep, unpruned Decision Tree (DT).
- Comparison: Base Learner trained on small subsamples ($N = 80$) vs. Bagged Learner (20 estimators) trained globally.

- Trials: 50.

Setup Scenario B (Real Data Heterogeneity - DT):

- Data Generation: Forest Covertype dataset (`fetch_covtype`), subsampled to 50,000 points. Features standardized.
- Conflict Engineering: Specialists trained on distinct Wilderness Areas (4 areas, corresponding to feature indices 10-13).
- Base Learner: Deep, unpruned Decision Tree (DT).
- Comparison: Specialist Ensemble (5 models per area, subsample size 250) vs. Bagged Learner trained globally.
- Trials: 10.

Setup Scenario C (Real Data Heterogeneity - 1-NN):

- Data Generation: Forest Covertype (same as Scenario B).
- Base Learner: 1-Nearest Neighbor (1-NN).
- Comparison: Specialist Ensemble (5 models per area, subsample size 250, using bootstrapping within the area to induce diversity among the deterministic 1-NN learners) vs. Bagged 1-NN trained globally.
- Trials: 10.

B.3 Experiment 3: Obstruction-Aware Regularization

Objective: Validate the proposed $\mathcal{L}_{cocycle}$ regularization.

Setup Details:

- System: $N = 3$ models/representation spaces. Dimension: 2D.
- Transformations: Learnable linear maps (ϕ_{ji}) implemented as PyTorch `nn.Linear` layers (bias=False).
- Engineered Inconsistency (Ground Truth Rotations):
  - ϕ_{10}: 60°. ϕ_{21}: 60°.
  - ϕ_{20}: −30° (Inconsistent, as composition yields 120°).
  - Total loop error: 150°.
- Training: Adam optimizer, Learning Rate 0.01, 1000 epochs.
- Loss: Pairwise alignment loss (MSE) $\pm\lambda_2\mathcal{L}_{cocycle}$.
- Parameter Sweep: Noise levels (0.01 to 0.5), λ_2 (1.0 to 20.0).
- Trials: 8 independent runs per configuration.
- Best Configuration (Reported in Table 3): Noise=0.1, $\lambda_2 = 5.0$.

B.4 Experiment 4: Scalability of Stochastic Approximation

Objective: Validate the efficiency and effectiveness of stochastic $\mathcal{L}_{cocycle}$.

Setup Details:

- System: $N = 15$ models. Dimension: 5D.
- Total Triplets: $\binom{15}{3} = 455$.
- Initialization: Xavier uniform initialization for ϕ_{ji} (high initial defect).
- Training: Adam optimizer, Learning Rate 0.05, 250 epochs.
- Loss: Pure regularization loss (no task loss).
- Scenarios: Full Batch (455 triplets), Stochastic ($K = 50$ sampled triplets), Baseline (loss=0).

Results

Table 5: Experiment 4 Summary (N=15, K=50).

Method	Complexity	Final Cocycle Defect	Wall-clock Time (s)
Baseline	-	1.0235×10^{1}	1.25
Full Batch	$O(N^3)$	9.1000×10^{-10}	15.66
Stochastic	$O(K)$	4.3446×10^{-5}	4.68

As shown in Table 5, the stochastic approximation provides a significant speedup (3.34x). The Full Batch approach (blue line) converges to a lower final defect due to the precise gradient calculation. The Stochastic approach (orange dashed line) exhibits the expected behavior of stochastic optimization: the variance introduced by sampling leads to a higher convergence floor (10^{-5}). However, this reduction is substantial, confirming that the stochastic approximation provides a practical and scalable implementation for obstruction-aware regularization.

B.5 Experiment 5: Structural Obstructions in Fairness

Objective: Demonstrate the emergence of structural obstructions in fairness-constrained model selection on the UCI Adult dataset.

Setup Details:

- Dataset: UCI Adult Income (Dua & Graff, 2019) (fetched via OpenML). N=48842.
- Features: 12 features (5 numeric, 7 categorical) processed via standardization and one-hot encoding. Missing values handled by mode imputation.
- Sensitive Attribute: Sex (Male/Female).
- Target: Income (>50K / <=50K).

- Models (K=5): Logistic Regression models trained with different heuristics. The objective is to induce diversity in the fairness-accuracy trade-offs, not necessarily to optimally achieve the target metrics:
  - Accuracy: Standard optimization (C=1.0).
  - Demographic Parity: Reweighting minority group (weight=2.5).
  - Equalized Odds: Reduced regularization (C=0.1).
  - Predictive Parity: Using `class_weight='balanced'`.
  - Balanced: Moderate reweighting (weight=1.5) and regularization (C=0.5).
- Stakeholders (Voters=5): Simulated utility functions prioritizing different metrics: Business (Accuracy only), Civil Rights (DP/EO), Equal Opportunity (EO/Acc), Calibration (PP/Acc), Regulator (Balanced).
- Aggregation: Pairwise Majority Vote (PMV) across the 5 stakeholders.
- Trials: 20 trials with different randomized train-test splits (70/30).