Spectral Byzantine Consensus: Anisotropic Aggregation on Continuous Semantic Fields

Cyfrpunk Raelizde
Peak&Tower Research
research@peakandtower.io

November 27, 2025

Abstract

We present a mathematical framework for achieving Byzantine fault-tolerant consensus on continuous functions defined over graph-structured domains. Traditional Byzantine consensus protocols operate on discrete state (transaction ordering, block validation), while distributed AI systems require consensus on *continuous semantic representations*: neural network embeddings, probability distributions, or geometric fields with inherent manifold structure.

Our protocol operates in the *spectral domain* of the graph Laplacian, representing distributed state as coefficients in an eigenbasis. We combine three key components: (1) *coordinate-wise weighted median aggregation* in the spectral basis, breaking rotational invariance to create anisotropic Byzantine resistance (achieving the optimal 1/2 breakdown point), (2) *mode-dependent heat flow dynamics* for structural modes ($k \geq 1$) that couple damping rates to eigenvalues, preferentially preserving low-frequency (global) structure while suppressing high-frequency (local) noise, combined with direct Byzantine-robust consensus for the mean value (mode 0). We also describe optional geometric coherence checks that can serve as auxiliary validation when combined with external energy budget enforcement.

We prove that for Byzantine weight fraction $\rho_w < 1/2$, the protocol achieves consensus with per-mode error bounded by $C_1\sigma_k\sqrt{\rho_w/(1-\rho_w)} + C_2\sigma_k\sqrt{\log(1/\delta)}$ with probability $1-\delta$, where σ_k is the honest noise variance in mode k and C_1, C_2 are explicit constants. The protocol exhibits two convergence regimes: mode 0 (mean value) converges at rate $O(\lambda_\star^{-1})$ governed by gossip mixing, while structural modes converge at rate $O((\eta\lambda_1)^{-1})$ governed by anisotropic heat flow, where λ_1 is the first nonzero eigenvalue of the semantic Laplacian.

Our framework bridges three research communities: (1) Byzantine fault tolerance (extending discrete consensus to continuous fields), (2) graph signal processing (adding adversarial robustness to spectral methods), and (3) distributed machine learning (enabling federated learning on geometric data with provable Byzantine resilience). We provide fast approximation algorithms using Chebyshev polynomials, avoiding explicit eigendecomposition while maintaining theoretical guarantees. For grid-structured domains (common in 3D world modeling, occupancy grids, and multi-dimensional sensor fields), we show the Laplacian admits a Kronecker product decomposition enabling exact spectral transforms in $O(d \cdot n \log n)$ time—achieving 10–100× speedups over Chebyshev for large spatial domains while preserving all robustness guarantees.

1 Introduction

1.1 The Problem: Consensus on Continuous Semantic Fields

Distributed agents increasingly need to agree on *continuous representations*—not just discrete transactions or state machine transitions. Autonomous robots building shared maps of their environment. Federated learning nodes aggregating neural network gradients. Distributed inference systems reaching consensus on probability distributions or embeddings. These applications share a common structure:

- **Multi-agent world modeling**: Agents maintain beliefs about 3D environments as neural radiance fields, occupancy grids, or semantic embeddings over spatial graphs.
- **Federated learning on geometric data**: Training neural networks on graph-structured data (molecules, social networks, knowledge graphs) requires aggregating high-dimensional parameters that respect geometric structure.
- **Decentralized AI inference**: Validators must reach consensus on model outputs (probability distributions, embeddings) without trusting any single node.

These settings demand protocols that (1) handle continuous state spaces, (2) respect geometric structure, (3) provide Byzantine robustness, and (4) scale to high-dimensional representations. Classical Byzantine fault-tolerant (BFT) consensus protocols [4, 3, 14] excel at *state machine replication*—achieving exact agreement on a discrete sequence of transactions—but are not designed for continuous semantic fields with inherent smoothness or manifold structure. Our work instead follows the tradition of *approximate Byzantine agreement* [6, 12], where agents with continuous-valued inputs converge to values within ϵ of each other rather than agreeing on an exact total ordering. We extend this classical framework to high-dimensional spectral representations on graphs.

Scope: This work focuses on Byzantine-robust consensus over continuous graph signals in federated learning and distributed semantic modeling. We assume a **fixed, trusted semantic graph** whose structure is established offline (formalized in Assumption 2.14). Byzantine adversaries can inject arbitrary signal values but cannot corrupt the graph topology or spectral basis itself. This is not a blockchain consensus protocol—we do not address transaction ordering, finality, or network-level attacks. Our focus is robust aggregation of continuous functions when the structural prior is trusted.

1.2 Our Contribution: Coordinate-Wise Spectral Medians

The core idea is simple: represent distributed state in the *spectral domain* of the graph Laplacian, then apply Byzantine-robust aggregation *coordinate-wise* to spectral coefficients.

Consider agents $\mathcal{A}$ maintaining beliefs $f^{(a)} : V \to \mathbb{R}$ over a graph $G = (V, E)$ (e.g., a spatial grid, knowledge graph, or neural network feature space). Instead of exchanging raw function values, agents communicate *spectral coefficients*:

$$c_k^{(a)} = \phi_k^\top f^{(a)} = \sum_{i \in V} \phi_k(i) f^{(a)}(i) \in \mathbb{R}$$

where $\{\phi_k\}$ are the eigenvectors of the normalized graph Laplacian L, ordered by eigenvalue λ_k.

Key insight: The eigenbasis forms a *frequency decomposition*:

- Low modes ($\lambda_k \ll 1$): Global, smooth structure (slowly varying across the graph).
- High modes ($\lambda_k \approx 2$): Local, rapid oscillations (high-frequency detail).

By applying the weighted scalar median *separately to each spectral coefficient* $c_k \in \mathbb{R}$ (reducing to the 1D weighted median since each mode is scalar), we achieve three properties simultaneously:

1. Optimal Byzantine Breakdown Point The coordinate-wise weighted median achieves the optimal $\rho_w < 1/2$ breakdown threshold (vs. 1/3 for many multivariate estimators). Byzantine outliers in specific frequency bands are filtered locally before they can corrupt consensus. Unlike rotationally-invariant vector aggregation, the coordinate-wise approach privileges the graph's eigenbasis—adversarial errors in unstable high-frequency modes cannot corrupt stable low-frequency structure through rotational mixing.

2. Frequency-Dependent Dynamics (Anisotropic Heat Flow) We apply mode-dependent update rates $\alpha_k = \eta\lambda_k$ to create frequency-dependent convergence:

$$c_{i,k}^{(t+1)} = (1 - \alpha_k)c_{i,k}^{(t)} + \alpha_k \tilde{c}_{i,k}^{(t)}$$

High-frequency modes (large λ_k) damp exponentially fast, while low-frequency modes (small λ_k) evolve slowly but robustly. This *anisotropic* (direction-dependent) convergence preferentially preserves global structure.

3. Optional Geometric Sanity Checks In truncated K-mode spectral space, we can perform geometric validation: if claimed events violate triangle inequality beyond what the Dirichlet energy budget allows, this provides a geometric sanity check for inconsistency (Section 6). However, this check is *auxiliary and optional*—it requires external enforcement of energy constraints and does not replace cryptographic fork detection mechanisms. The core protocol achieves Byzantine robustness through coordinate-wise median aggregation and heat flow dynamics alone.

1.3 Contributions

Theoretical contributions:

1. **Mode-wise Byzantine robustness** (Theorems 4.5, 4.8): We prove that the weighted median in scalar spectral space achieves error bounded by $C_1\sigma_k\sqrt{\rho_w/(1-\rho_w)}+C_2\sigma_k\sqrt{\log(1/\delta)}$ for Byzantine weight fraction $\rho_w < 1/2$, achieving the optimal breakdown threshold. This extends standard median robustness to the weighted case with sub-Gaussian noise. Trimmed mean aggregation builds on this with additional guarantees.

2. **Anisotropic convergence analysis** (Theorems 5.4, 5.6, 5.9): We establish two convergence regimes: (a) mode 0 (mean value) achieves consensus at rate $O(\lambda_\star^{-1}\log(1/\epsilon))$ governed by gossip mixing on the honest subgraph, and (b) structural modes ($k \geq 1$) exhibit exponential convergence with mode-dependent rates $e^{-2\eta\lambda_k t}$, with global contraction rate $O((\eta\lambda_1)^{-1}\log(1/\epsilon))$ via a weighted Lyapunov function $V^{(t)} = \sum_{k\geq 1} w_k V_k^{(t)}$.

3. **Intrinsic fork detection** (Theorem 6.3): We prove that the anisotropic heat flow dynamics establish an intrinsic energy ceiling $\mathcal{E}_{\max}$ for honest agent disagreement. Any pair exceeding this bound after mixing time T_{mix} indicates Byzantine majority or protocol deviation—without requiring external energy budget enforcement. This "thermodynamic" bound emerges naturally from the protocol's own Lyapunov contraction.

Algorithmic efficiency and baseline comparison:

Table 1 compares our approach to existing Byzantine-robust aggregation methods for federated learning on high-dimensional state. While traditional methods like Krum [1] and Bulyan [7] scale quadratically in the number of agents N and linearly in state dimension d, our spectral approach achieves $O(K|E| + KN)$ complexity by operating on $K \ll d$ spectral modes over a sparse graph with $|E|$ edges.

Method	Per-Round Cost	Message Size	Byzantine ρ	Geometry
FedAvg	$O(Nd)$	$O(d)$ floats	None	No
Krum [1]	$O(N^2d)$	$O(d)$ floats	$< 1/2$	No
Bulyan [7]	$O(N^2d\log N)$	$O(d)$ floats	$< 1/4$	No
HotStuff	$O(N)$ messages	$O(1)$ hashes	$< 1/3$	No
This work	$\mathbf{O(K\|E\| + KN)}$	$O(K)$ floats	$< \mathbf{1/2}$	**Yes**

Table 1: Complexity comparison for Byzantine-robust aggregation. For high-dimensional state ($d \sim 10^6$) on sparse graphs ($|E| = O(N)$) with $K \ll d$ modes, our method achieves both computational and communication efficiency: $K \approx$ 50–100 floats per message vs. $d \approx 10^6$ for parameter-space methods. HotStuff [14] is included as a discrete BFT baseline for communication complexity comparison; it operates on hashes rather than continuous state and achieves consensus on transaction ordering, not semantic content—it is not directly comparable for our continuous aggregation setting but illustrates the communication advantage of spectral compression.

Our efficient implementation uses:

- **Chebyshev polynomial approximation**: Avoids explicit eigendecomposition by approximating spectral projections with Chebyshev polynomials of the Laplacian, achieving $O(K|E|)$ complexity for K modes and $|E|$ edges.

- **Kronecker-separable exact transform**: For grid-structured domains (regular d-dimensional lattices common in 3D occupancy grids, voxel worlds, and multi-dimensional sensor arrays), the Laplacian admits a Kronecker product decomposition enabling exact spectral projections in $O(d \cdot n \log n)$ time via separable transforms. This eliminates Chebyshev approximation error entirely while achieving 10–100× speedups for large spatial domains ($n \geq 10^5$), making real-time consensus on 128^3–256^3 grids tractable on edge devices.

- **Fast weighted median computation**: We use Weiszfeld's algorithm with warm-starting across modes, achieving $O(m \log(1/\epsilon))$ iterations for m agents and tolerance ϵ.

Remark 1.1. While weighted medians, Chebyshev approximations, and spectral graph theory are classical tools, their combination into a Byzantine-robust consensus protocol operating natively in spectral space with mode-dependent treatment appears novel. Prior work applies spectral methods for post-hoc analysis (Section 8) or treats all dimensions uniformly in robust aggregation. Our key insight is the mode-wise application of scalar robust statistics to achieve anisotropic Byzantine resistance.

2 Preliminaries

2.1 Graph Notation and Agent-Domain Identification

This protocol involves two distinct graph structures:

- **Semantic domain graph** $G_{\text{dom}} = (V_{\text{dom}}, E_{\text{dom}})$ with $n = |V_{\text{dom}}|$ nodes: The fixed structure over which belief signals are defined (e.g., spatial grid, knowledge graph, feature space).

- **Agent communication graph** $G_{\text{com}} = (\mathcal{A}, E_{\text{com}})$ with $N = |\mathcal{A}|$ agents: The network over which consensus participants exchange messages.

Each agent $a \in \mathcal{A}$ maintains a belief signal $f^{(a)} : V_{\text{dom}} \to \mathbb{R}$, represented as a vector $f^{(a)} \in \mathbb{R}^n$. The spectral basis $\{\phi_k\}_{k=0}^{n-1}$ is derived from the *domain* graph Laplacian (Section 2), while gossip dynamics operate over the *communication* graph (Section 3).

Remark 2.1 (Notational Convention: $n = N$ Identification)**.** For expository simplicity, we identify $n = N$ throughout this paper: each agent corresponds to a domain node, and agents maintain beliefs over the full N-node domain. This is natural when the semantic domain *is* the agent network (e.g., distributed sensors where each agent's location is a domain node). The analysis extends to $n \neq N$ by treating $f^{(a)} \in \mathbb{R}^n$ as each agent's belief over a shared n-dimensional domain, with N agents participating in consensus. When $n = N$, we use N uniformly to denote both quantities.

2.2 Graph Laplacian and Spectral Decomposition

Definition 2.2 (Normalized Graph Laplacian)**.** Let $G = (V, E, w)$ be a weighted undirected graph with $N = |V|$ vertices and edge weights $w_{ij} \geq 0$. The degree matrix is $D = \mathrm{diag}(d_1, \ldots, d_N)$ where $d_i = \sum_j w_{ij}$. The *normalized graph Laplacian* is:

$$L = I - D^{-1/2} W D^{-1/2}$$

where $W = (w_{ij})$ is the adjacency matrix.

The Laplacian is symmetric positive semidefinite with spectral decomposition:

$$L = \sum_{k=0}^{N-1} \lambda_k \phi_k \phi_k^\top$$

where:

- Eigenvalues: $0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_{N-1} \leq 2$
- Eigenvectors: $\{\phi_k\}_{k=0}^{N-1} \subset \mathbb{R}^N$ (orthonormal basis)

Definition 2.3 (Spectral Coefficients)**.** For a function $f : V \to \mathbb{R}$ represented as a vector $f \in \mathbb{R}^N$, the *spectral coefficient* for mode k is:

$$c_k := \phi_k^\top f = \sum_{i=1}^{N} \phi_k(i) f(i) \in \mathbb{R}$$

The function can be reconstructed as $f = \sum_{k=0}^{N-1} c_k \phi_k$.

Definition 2.4 (Dirichlet Energy)**.** The *Dirichlet energy* (or graph Laplacian quadratic form) of f is:

$$\mathcal{E}(f) := f^\top L f = \frac{1}{2} \sum_{(i,j) \in E} w_{ij} \left(\frac{f(i)}{\sqrt{d_i}} - \frac{f(j)}{\sqrt{d_j}} \right)^2.$$

In spectral space: $\mathcal{E}(f) = \sum_{k=0}^{N-1} \lambda_k c_k^2$.

Intuition: The Dirichlet energy measures how much f varies across edges. Smooth (low-energy) functions have most energy in low-λ_k modes.

Remark 2.5 (Notation: Two Distinct Spectral Gaps)**.** This paper involves two graph Laplacians with distinct spectral gaps:

1. **Semantic Laplacian** L_{semantic} (domain graph): First nonzero eigenvalue $\lambda_1 > 0$ controls how quickly structural modes ($k \geq 1$) converge under anisotropic heat flow.
2. **Gossip Laplacian** L_H (honest communication graph): Spectral gap $\lambda_\star := \lambda_1(L_H) > 0$ governs gossip averaging (mode 0 convergence) and mixing time.

These are *independent* parameters: λ_1 is a property of the semantic domain structure (precomputed offline), while $\lambda_\star$ is a property of the agent communication topology. To avoid confusion, we reserve $\lambda_\star$ exclusively for the gossip gap throughout.

2.3 Chebyshev Polynomial Approximation

Computing eigenvectors $\{\phi_k\}$ directly via eigendecomposition costs $O(N^3)$, prohibitive for large graphs. We use Chebyshev polynomials to approximate spectral projections.

Definition 2.6 (Chebyshev Polynomials)**.** The Chebyshev polynomials of the first kind $\{T_k\}_{k=0}^{\infty}$ on $[-1,1]$ satisfy:

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x).$$

Lemma 2.7 (Chebyshev Approximation)**.** *Let L have eigenvalues in $[0, \lambda_{\max}]$. Define the scaled operator:*

$$\widetilde{L} = \frac{2}{\lambda_{\max}} L - I$$

with eigenvalues in $[-1,1]$. For any $f \in \mathbb{R}^N$, the Chebyshev features:

$$u_k = T_k(\widetilde{L})f$$

can be computed via the recurrence:

$$u_0 = f, \quad u_1 = \widetilde{L}f, \quad u_{k+1} = 2\widetilde{L}u_k - u_{k-1}$$

in $O(K|E|)$ time for K polynomials.

Moreover, $\left\| T_k(\widetilde{L}) \right\|_{\mathrm{op}} \leq 1$, so noise in f does not amplify.

Remark 2.8. For the theoretical analysis in this paper we assume direct access to the spectral coefficients $c_k = \phi_k^\top f$, which can be obtained by standard Krylov methods in $O(K|E|)$ time. In practice, they can be approximated using the Chebyshev recurrence in Lemma 2.7; the approximation error does not affect the Byzantine robustness bounds, only the effective noise levels σ_k.

2.3.1 Kronecker-Separable Acceleration (Optional Fast Path)

When the semantic domain V is a *Cartesian product* of smaller domains—most importantly, a regular d-dimensional grid $V = V_1 \times V_2 \times \cdots \times V_d$ with $|V_\delta| = m$ for each dimension δ—the graph Laplacian admits a *Kronecker product decomposition* that enables dramatic computational speedups while preserving all theoretical guarantees.

Kronecker Structure for Product Graphs For a d-dimensional grid graph $G = G_1 \square G_2 \square \cdots \square G_d$ (Cartesian product), the (unnormalized) Laplacian decomposes as a Kronecker sum:

$$L = \bigoplus_{\delta=1}^{d} L_\delta := \sum_{\delta=1}^{d} (I \otimes \cdots \otimes L_\delta \otimes \cdots \otimes I)$$

where L_δ is the 1D Laplacian on G_δ and $\otimes$ denotes the Kronecker product.

Key properties:

1. **Eigenvector factorization:** Eigenvectors are exact tensor products:
$$\phi_{k_1,k_2,\ldots,k_d}(i_1,i_2,\ldots,i_d) = \phi^{(1)}_{k_1}(i_1) \otimes \phi^{(2)}_{k_2}(i_2) \otimes \cdots \otimes \phi^{(d)}_{k_d}(i_d)$$
2. **Eigenvalue additivity:** Eigenvalues sum across dimensions:
$$\lambda_{k_1,\ldots,k_d} = \sum_{\delta=1}^{d} \lambda^{(\delta)}_{k_\delta}$$
3. **Separable transforms:** The graph Fourier transform factors into d independent 1D transforms (DCT-II/DCT-III for path graphs, FFT for cycle graphs).

Remark 2.9 (Normalized vs Unnormalized Laplacians)**.** The Kronecker structure is *exact* for unnormalized Laplacians $L = D - W$. For normalized Laplacians $L_{\text{norm}} = I - D^{-1/2}WD^{-1/2}$, the Kronecker decomposition is approximate but excellent for regular grids (degrees nearly constant). Since eigenvectors are identical (up to scaling) and only eigenvalue magnitudes differ slightly, all spectral methods remain valid with negligible approximation error.

Protocol Modifications When the Kronecker fast path is available, agents make the following substitutions in Algorithm 1:

State representation: Instead of vector $c_i^{(t)} \in \mathbb{R}^K$, agent i maintains a d-dimensional spectral tensor:
$$C_i^{(t)} \in \mathbb{R}^{K_1 \times K_2 \times \cdots \times K_d}$$
where K_δ is the number of retained modes in dimension δ and $\prod_\delta K_\delta \approx K$ (total modes).

Projection (before line 1): Replace Chebyshev approximation with d-dimensional separable transform:

```
$F_i^{(t)} \leftarrow \text{reshape}(f^{(i)}(t), (m_1, m_2, \ldots, m_d))$ {Vectorize as tensor}
for $\delta = 1, \ldots, d$ do
    Apply 1D DCT-II along dimension $\delta$
end for
$C_i^{(t)} \leftarrow$ truncate to low-frequency $K_1 \times K_2 \times \cdots \times K_d$
```

Complexity: $O(d \cdot N \log m)$ where $m \approx N^{1/d}$ is the side length per dimension.

Mode-wise aggregation (lines 7–9): Iterate over multi-index $(k_1, \ldots, k_d) \in [K_1] \times \cdots \times [K_d]$:
$$\tilde{C}_i^{(t)}[k_1,\ldots,k_d] = \text{med}_w\big(\{C_j^{(t)}[k_1,\ldots,k_d] : j \in \mathcal{N}_i^{(t)}\}, \{w_{ij}^{(t)}\}\big)$$
This remains a scalar weighted median per frequency bin—*completely unchanged* from Algorithm 1.

Anisotropic heat flow (lines 12–14): Compute damping coefficient from additive eigenvalue:
$$\alpha_{k_1,\ldots,k_d} = \eta \cdot \sum_{\delta=1}^{d} \lambda^{(\delta)}_{k_\delta}$$
Then apply elementwise:
$$C_i^{(t+1)}[k_1,\ldots,k_d] \leftarrow (1 - \alpha_{k_1,\ldots,k_d}) \cdot C_i^{(t)}[k_1,\ldots,k_d] + \alpha_{k_1,\ldots,k_d} \cdot \tilde{C}_i^{(t)}[k_1,\ldots,k_d]$$

Reconstruction (when needed): Inverse d-dimensional DCT-III, again $O(d \cdot N \log m)$.

Theoretical Guarantees (Unchanged) **All theorems in this paper remain valid under the Kronecker fast path:**

- **Byzantine robustness** (Theorems 4.5, 4.8): Mode-wise median aggregation is identical—only the indexing changes from linear to multi-dimensional.
- **Dirichlet energy** (Definition in §2):

$$\mathcal{E}(f) = \sum_{k_1,\ldots,k_d} \lambda_{k_1,\ldots,k_d} |C[k_1,\ldots,k_d]|^2$$

  with $\lambda_{k_1,\ldots,k_d} = \sum_\delta \lambda^{(\delta)}_{k_\delta}$ preserves the same quadratic form.
- **Convergence rates** (Theorems 5.6, 5.9): The weighted Lyapunov function sums over all modes with weights $w_{k_1,\ldots,k_d} = (1+\lambda_{k_1,\ldots,k_d})^{-\beta}$. Convergence analysis is identical.
- **Fork detection** (Theorem 6.3): Spectral distance checks and the intrinsic energy ceiling $\mathcal{E}_{\max}$ operate on the coefficient space, which has the same dimension $K = \prod_\delta K_\delta$ regardless of indexing.

The only change is *how* coefficients are computed and indexed—the aggregation, energy, and convergence properties are invariant to the choice of basis ordering.

Complexity Comparison For a d-dimensional grid with $N = m^d$ nodes and $K = K_0^d$ retained modes (e.g., $64^3 = 262,144$ nodes, $K_0 = 32$ per dimension, $K = 32,768$ total modes):

Operation	**Dense**	**Chebyshev**	**Kronecker-Separable**
Memory (coefficients)	$O(N)$	$O(K)$	$O(K)$
Projection	$O(KN)$ exact	$O(K\|E\|)$ approx	$O(d \cdot N \log m)$ exact
Matvec (per iteration)	$O(N^2)$	$O(\|E\|)$	$O(d \cdot N)$
Total (K-mode projection)	$O(KN)$	$O(K\|E\|) \approx O(KdN)$	$O(d \cdot N \log m)$
Example: 64^3 **grid** $(N = 262k, K = 32k, d = 3)$	2*—	2*$\approx 2 \times 10^9$ ops	2*$\approx 1.5 \times 10^7$ ops
Speedup	—	**1×** (baseline)	**133× faster**

Table 2: Computational complexity for spectral projection. For 3D grids with $K \gtrsim d \log m$, the Kronecker path is asymptotically superior: $O(d \cdot N \log N^{1/d}) = O(d \cdot N \log N/d) \ll O(KdN)$ when $K \gg \log m$.

Concrete speedups:

- **2D grid** ($100 \times 100 = 10k$ nodes): Chebyshev $\approx 10^8$ ops, Kronecker $\approx 2 \times 10^6$ ops → **50× speedup**
- **3D grid** ($64^3 = 262k$ nodes): Chebyshev $\approx 2 \times 10^9$ ops, Kronecker $\approx 1.5 \times 10^7$ ops → **133× speedup**

- **3D grid** ($128^3 = 2M$ nodes): Chebyshev $\approx 1.6 \times 10^{10}$ ops, Kronecker $\approx 1.4 \times 10^8$ ops $\rightarrow$ **114× speedup**

For applications requiring on-device/real-time operation (drone swarms, autonomous vehicles, edge federated learning), the Kronecker path makes 3D consensus at 128^3–256^3 resolution *tractable* where Chebyshev would be prohibitively slow.

Applicability The Kronecker fast path applies when:

- **Semantic domain is a product graph:** Regular grids (most common), hypercubes, torus graphs, or Cartesian products of arbitrary factor graphs.
- **Pre-computed 1D eigenbases:** The 1D Laplacians L_δ for each dimension must be factorizable (trivial for path/cycle graphs; basis vectors are known analytically).
- **Dimensions are not too small:** For $d = 2$ with $m \geq 50$, speedups are $> 10\times$. For $d = 3$ with $m \geq 32$, speedups are $> 50\times$.

Non-grid domains (irregular meshes, social networks, molecule graphs) cannot use Kronecker structure and must use Chebyshev approximation. The protocol gracefully degrades to Chebyshev (already efficient at $O(K|E|)$) when Kronecker structure is unavailable.

Remark 2.10 (Practical Recommendation)**.** For typical federated learning or multi-agent scenarios on 2D/3D spatial domains with $N \geq 10^4$ nodes:

1. **If grid-structured**: Use Kronecker path (10–100× faster, exact coefficients).
2. **Otherwise**: Use Chebyshev approximation (still $O(K|E|)$, general-purpose).

Both paths are *mathematically equivalent* in terms of Byzantine robustness and convergence—only computational efficiency differs.

2.4 Byzantine-Robust Statistics

Definition 2.11 (Weighted Median (Scalar))**.** For scalar values $\{x_a\}_{a \in \mathcal{A}} \subset \mathbb{R}$ with weights $w_a > 0$ satisfying $\sum_a w_a = 1$, the *weighted median* $\mathrm{med}_w(x)$ is:

$$m = \inf \left\{ y \in \mathbb{R} : \sum_{a : x_a \leq y} w_a \ \geq \ \tfrac{1}{2} \right\}.$$

When $\sum_{a : x_a = m} w_a > 1/2$ (non-unique case), any value in the interval $[\max\{x_a : x_a < m\}, \min\{x_a : x_a > m\}]$ is a valid median. For determinacy, we take:

$$m = \min \left\{ x_a : \sum_{b : x_b \leq x_a} w_b \ \geq \ \tfrac{1}{2} \right\}.$$

Remark 2.12. In one dimension ($d = 1$), the geometric median reduces to the weighted median. The weighted median has a breakdown point of $1/2$: it remains bounded as long as the Byzantine weight fraction satisfies $\rho_w < 1/2$. This is superior to the mean (breakdown point 0) and matches the optimal breakdown point for any affine-equivariant estimator.

In this work, we apply the weighted median *mode-wise* in spectral space: each spectral coefficient $c_k \in \mathbb{R}$ is a scalar, and we aggregate coefficients independently across modes. The robustness properties are established in Theorem 4.5.

2.5 System Model and Assumptions

We now formalize the assumptions underlying our Byzantine-robust spectral consensus protocol.

2.5.1 Semantic Constitution

Definition 2.13 (Semantic Domain Graph)**.** The semantic domain graph $G_{\text{dom}} = (V_{\text{dom}}, E_{\text{dom}}, w)$ is a weighted undirected graph where:

- V_{dom}: Domain nodes representing semantic concepts (e.g., spatial locations, word embeddings, ontology nodes);
- E_{dom}: Domain edges encoding semantic similarity with weights $w_{ij} \geq 0$;
- Laplacian: $L_{\text{dom}} = D - W$ where $D[i,i] = \sum_j w_{ij}$ and $W = (w_{ij})$.

Under the $n = N$ identification (Remark 2.1), we write L for L_{dom} throughout.

Assumption 2.14 (Fixed Eigenbasis)**.** The eigenpairs $\{(\lambda_k, \phi_k)\}_{k=0}^{N-1}$ of L are computed offline via:

$$L\phi_k = \lambda_k \phi_k, \quad \text{where} \quad 0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_{N-1}.$$

We retain K modes: $\Phi_K = [\phi_0, \phi_1, \ldots, \phi_{K-1}] \in \mathbb{R}^{N \times K}$.

Computational Cost: $O(N^3)$ dense, or $O(NK^2)$ via Lanczos for sparse L.

Immutability: Once computed, $\{\phi_k, \lambda_k\}$ are fixed for all agents and all time $t \geq 0$. Byzantine agents cannot corrupt the eigenbasis.

2.5.2 Agent Model

Definition 2.15 (Agent State)**.** Each agent $i \in \{1, \ldots, N\}$ maintains:

- State vector: $f^{(i)}(t) \in \mathbb{R}^n$;
- Spectral coefficients: $c_k^{(i)}(t) = \langle \phi_k, f^{(i)}(t) \rangle$ for $k = 0, \ldots, K-1$.

Dynamics: Agents update via spectral gossip (Algorithm 1).

2.5.3 Byzantine Adversary Model

Assumption 2.16 (Byzantine Scope)**.** The communication graph G is constructed from honest agent reports of pairwise similarities or stakes. Byzantine agents can:

- Report arbitrary belief signals $f^{(a)}$ (data poisoning);
- Submit malformed spectral features $\{c_k^{(a)}(t)\}$ (feature attacks).

Byzantine agents **cannot**:

- Forge edge weights in G for other agents' views;
- Modify the Laplacian structure L seen by honest nodes;
- Corrupt the eigenbasis $\{(\lambda_k, \phi_k)\}$ (Assumption 2.14);
- Perform Sybil attacks (total Byzantine weight ρ_w is fixed).

Rationale: This scoping reflects federated learning scenarios where graph structure derives from trusted metadata (geographical proximity, organizational hierarchy, or cryptographic stake) while belief data comes from potentially compromised sources.

Assumption 2.17 (Byzantine Corruption)**.** The adversary controls a subset $V_B \subseteq \{1, \ldots, N\}$ of Byzantine agents. Byzantine agents can broadcast arbitrary values $\{c_k^{(i)}(t)\}$. Honest agents $V_H = \{1, \ldots, N\} \setminus V_B$ follow Algorithm 1.

Assumption 2.18 (Local Byzantine Bound)**.** For each honest agent $i \in V_H$, the fraction of Byzantine neighbors satisfies:

$$\rho_i = \frac{|N_i \cap V_B|}{|N_i|} \leq \rho_{\text{loc}} < \frac{1}{2}$$

where N_i is agent i's neighborhood in the communication graph.

Assumption 2.19 (Honest Connectivity)**.** The communication subgraph induced by honest agents $G_H = (V_H, E_H)$ is connected with spectral gap (first nonzero eigenvalue):

$$\lambda_\star := \lambda_1(L_H) > 0.$$

This parameter governs the convergence rate of gossip averaging (mode 0) and appears in the global contraction rate for structural modes.

Remark 2.20. Assumptions 2.18–2.19 are standard in Byzantine consensus literature [12, 11]. Our contribution is combining these with spectral decomposition to enable geometry-aware robust aggregation. Note that the local bound $\rho_{\text{loc}} < 1/2$ aligns with the optimal breakdown point of the weighted median.

3 Protocol Description

3.1 System Model

Agents: A set $\mathcal{A} = H \cup B$ of N agents, where:

- Honest agents H: follow the protocol, represent ground truth with bounded noise;
- Byzantine agents B: controlled by an adversary, send arbitrary values.

Stake weights: Each agent a has weight $w_a > 0$ with $\sum_{a \in \mathcal{A}} w_a = 1$. We assume:

$$\rho_w := \sum_{a \in B} w_a < \frac{1}{2}$$

(Byzantine agents control less than 1/2 of total weight, the breakdown threshold for weighted median.)

Communication model: Let $G_{\text{full}} = (\mathcal{A}, E_{\text{full}})$ denote the full communication graph including both honest and Byzantine agents. Honest agents communicate over the induced *honest subgraph* $G_H = (H, E_H)$ with Laplacian L_H. We assume G_H is connected (spectral gap $\lambda_\star := \lambda_1(L_H) > 0$).

Byzantine injection mechanism: Neighbor sampling for gossip weights restricts to G_H—honest agents sample neighbors only from among other honest agents. However, the aggregation step (Algorithm 1, line 9) accepts coefficient proposals from any agent $j \in \mathcal{N}_i^{(t)} \cup B_i$, where $B_i \subseteq B$ are Byzantine agents adjacent to honest agent i in G_{full}. Byzantine messages are included with their stake weights, subject to the global bound $\sum_{b \in B} w_b \leq \rho_w < 1/2$. This models scenarios where Byzantine agents can broadcast messages but cannot influence which honest agents are sampled as neighbors.

Semantic representation: Each agent a maintains a belief $f^{(a)} : V \to \mathbb{R}$ over a fixed domain graph $G = (V, E)$ (the "semantic space"). For honest agents:

$$\left\| f^{(a)} - f^\star \right\|_2 \leq \sigma_0$$

where $f^\star$ is the ground truth function.

Assumption 3.1 (Honest Spectral Basis)**.** The eigenpairs $\{(\lambda_k, \phi_k)\}_{k=0}^{K-1}$ of the semantic Laplacian L_{sem} are computed during a *trusted setup phase* and remain immutable throughout protocol execution. Byzantine agents may inject arbitrary signal values (coefficient proposals) but **cannot** manipulate the spectral basis itself.

This assumption is appropriate when the semantic graph structure is derived from:

- **Physical constraints**: Spatial adjacency in SLAM, sensor networks, or multi-robot systems;
- **Pre-trained embeddings**: Feature spaces from models trained on trusted corpora;
- **Collaborative initialization**: Graph constructed during a Byzantine-free bootstrapping phase.

Byzantine robustness to adversarial manifold learning (where adversaries poison the basis construction) remains an open problem; see Section 7 for discussion and potential mitigation strategies.

3.2 Protocol State

At round t, each agent $i \in \mathcal{A}$ maintains:

$$c_i^{(t)} = (c_{i,0}^{(t)}, c_{i,1}^{(t)}, \dots, c_{i,K-1}^{(t)}) \in \mathbb{R}^K$$

where $c_{i,k}^{(t)} \in \mathbb{R}$ is the coefficient for mode k.

Initialization: Each agent computes initial coefficients:

$$c_{i,k}^{(0)} = \phi_k^\top f^{(i)}(0)$$

using Chebyshev approximation (Lemma 2.7).

3.3 Protocol Round

At each round t, agent i executes:

Algorithm 1 Spectral Byzantine Consensus with Static Basis (S+B)

```
1:  Input: Trusted eigenpairs $\{(\lambda_k, \phi_k)\}_{k=0}^{K-1}$ of $L_{\text{sem}}$, stepsize $\eta > 0$.
2:  State at agent $i$: Spectral coefficients $c_i^{(t)} \in \mathbb{R}^K$.
3:  Optional fast path: If domain is a $d$-dimensional grid, replace Chebyshev projection
    with separable $d$-dimensional DCT/FFT (Section 2.3.1) for 10–100× speedup.
4:  for each round $t = 0, 1, 2, \dots$ do
5:    for each honest agent $i \in H$ in parallel do
6:      Sample neighbor multiset $\mathcal{N}_i^{(t)}$ from $\mathcal{D}_i$ (lazy random walk on $G_H$).
7:      Receive messages $\{c_j^{(t)} : j \in \mathcal{N}_i^{(t)}\}$.
8:      for each mode $k = 0, 1, \dots, K-1$ do
9:        Form scalar reports $x_{j,k} = c_{j,k}^{(t)}$ for $j \in \mathcal{N}_i^{(t)} \cup \{i\}$, with weights $w_{ij}^{(t)}$ (stake weights
          of agent $j$, normalized over sampled multiset).
10:       Compute weighted scalar median (robust center): $m_k = \text{med}_w(\{x_{j,k}\}_j, \{w_{ij}^{(t)}\}_j)$.
11:       Compute trimming parameter: $\alpha_k^{\text{trim}} = \text{clip}(\eta\lambda_k, \rho_w, 1 - 2\rho_w)$.
12:       Trim agents: sort by distance $|x_{j,k} - m_k|$, remove until trimmed weight $\geq \alpha_k^{\text{trim}}$.
13:       Compute trimmed weighted mean: $\tilde{c}_{i,k}^{(t)} = \sum_{j \in S_k} \tilde{w}_{ij} x_{j,k}$, where $S_k$ is kept set and
          $\tilde{w}_{ij}$ renormalized.
14:       if $k = 0$ then
15:         $c_{i,0}^{(t+1)} \leftarrow \tilde{c}_{i,0}^{(t)}$ {direct consensus on mean}
16:       else
17:         $\alpha_k \leftarrow \eta\lambda_k$
18:         $c_{i,k}^{(t+1)} \leftarrow (1 - \alpha_k) c_{i,k}^{(t)} + \alpha_k \tilde{c}_{i,k}^{(t)}$
19:       end if
20:     end for
21:   end for
22: end for
```

Remark 3.2 (Two-Step Aggregation: Median then Trimmed Mean)**.** The aggregation procedure (lines 8–12) uses a two-step process combining the strengths of both estimators:

1. **Geometric median** (line 8): Provides a robust center estimate with 50% breakdown point (Theorem 4.5).

2. **Distance-based trimming** (lines 9–11): Sorts agents by distance from the median, trims outliers, then computes a weighted mean (Theorem 4.8).

This combines the robustness of the geometric median with the computational efficiency of trimmed means. The mode-dependent trimming parameter α_k^{trim} (line 10) enables *anisotropic* Byzantine resistance: high-frequency modes are trimmed more aggressively than low-frequency modes.

Remark 3.3 (Why Coordinate-Wise Aggregation?)**.** We apply the two-step aggregation *separately per mode* rather than computing a single median in the full K-dimensional spectral coefficient space. This design choice breaks rotational invariance, effectively enforcing a "Manhattan" (L^1-like) geometry in the spectral domain that aligns with the graph's eigenbasis.

Alternative (not used): Computing $\tilde{c}_i^{(t)} = \arg\min_{z\in\mathbb{R}^K} \sum_j w_{ij}\|z - c_j^{(t)}\|_2$ would aggregate the full coefficient vector as a single geometric object, allowing rotational freedom in spectral space.

Why coordinate-wise is stronger: By aggregating mode-by-mode, we prevent adversarial agents from "rotating" their attack to mix low- and high-frequency components. An adversary trying to inject high-frequency noise cannot hide it by aligning with low-frequency modes—each mode is filtered independently based on its own statistics. This ensures that Byzantine outliers in specific frequency bands (particularly high frequencies) are rejected locally before they can corrupt the global consensus structure.

This architectural decision is critical for the mode-dependent Byzantine robustness guarantees in Theorems 4.5 and 4.8.

Parameters:

- K: number of spectral modes (typically $K \ll N$, e.g., $K = O(\log N)$);
- $\eta \in (0, 1/\lambda_{\max})$: heat flow stepsize;
- $\alpha_k = \eta\lambda_k$ for $k \geq 1$: mode-dependent damping coefficient;
- mode weights for Lyapunov analysis: $w_k = (1 + \lambda_k)^{-\beta}$ with $\beta > 0$.

Remark 3.4 (Spectral Anisotropy)**.** Note that we apply the two-step aggregation (median then trimmed mean) *per mode* (lines 8–12 of Algorithm 1) rather than on the full vector $c_i^{(t)} \in \mathbb{R}^K$. This coordinate-wise non-linearity is what creates the *anisotropic resistance profile*, preventing Byzantine noise in high modes from rotating into or corrupting the structural low modes.

In contrast, applying a multivariate geometric median to the full coefficient vector would be *isotropic*—treating all spectral directions equally and allowing adversarial perturbations to "leak" between modes. The coordinate-wise application privileges the eigenbasis $\{\phi_k\}$ as the fundamental axes of semantic truth, enforcing a Manhattan-like ℓ^1 geometry in spectral space rather than Euclidean ℓ^2 geometry. Combined with mode-dependent damping ($\alpha_k =$

$\eta\lambda_k$) and mode-dependent trimming (α_k^{trim}), this creates a *triply* anisotropic mechanism: aggregation center, trimming aggressiveness, and dynamics are all aligned with the graph's spectral structure.

Remark 3.5 (Mode 0 vs Structural Modes: Critical Architectural Distinction)**.** The protocol treats mode 0 and structural modes *fundamentally differently*, which is essential for both convergence and Byzantine robustness:

Mode 0 (Value Consensus):

- **Semantics**: $\phi_0 = [1/\sqrt{N}, \dots, 1/\sqrt{N}]^\top$ represents the *mean value* (DC component): $c_{i,0} = \sqrt{N} \cdot \text{mean}(f^{(i)})$
- **Update rule**: Direct consensus with $\alpha_0 = 1$ (full step toward aggregated value): $c_{i,0}^{(t+1)} = \tilde{c}_{i,0}^{(t)}$
- **Convergence**: Governed by gossip graph spectral gap $\lambda_\star = \lambda_1(L_H)$, with rate $O(\lambda_\star^{-1} \log(1/\epsilon))$ (Theorem 5.4)
- **Why necessary**: Using heat flow $\alpha_0 = \eta\lambda_0 = 0$ would *freeze* the coefficient (no updates), preventing agents from agreeing on the mean—violating the consensus objective

Structural Modes $k \geq 1$ (Pattern Consensus):

- **Semantics**: Represent *how the function varies* (patterns, gradients, structures)
- **Update rule**: Anisotropic heat flow with $\alpha_k = \eta\lambda_k$: $c_{i,k}^{(t+1)} = (1 - \alpha_k)c_{i,k}^{(t)} + \alpha_k \tilde{c}_{i,k}^{(t)}$
- **Convergence**: Mode-dependent exponential decay at rate $e^{-2\eta\lambda_k t}$ (Theorem 5.6)
- **Anisotropic property**: Low-frequency modes (small λ_k) converge slowly, preserving global structure; high-frequency modes (large λ_k) damp rapidly, suppressing Byzantine noise

Mathematical justification: In continuous heat diffusion $\partial f/\partial t = -Lf$, the constant mode (eigenvalue 0) is conserved—the total "heat" (integral of f) does not change. Byzantine consensus, however, requires agents to *agree* on the mean, necessitating active updates to mode 0. The dual-path architecture resolves this fundamental tension between physical heat flow and consensus objectives.

This architectural distinction is critical for the convergence analysis in Section 5: Theorem 5.4 analyzes mode 0 separately from Theorems 5.6–5.9 for structural modes. Implementations must respect this separation to ensure both correctness and convergence guarantees.

3.4 Computational Complexity

The protocol has two distinct cost regimes: offline (one-time setup) and online (per-round).

Remark 3.6. Weiszfeld convergence is linear with rate depending on the condition number of the point configuration. For bounded Byzantine displacement, $I_{\text{med}} = O(\log(1/\epsilon))$ suffices. For scalar median when $d = 1$, standard sorting achieves $O(N \log N)$ complexity. (Note: Vardi-Zhang [13] addresses the multivariate geometric median, not scalar median.)

Phase	**Operation**	**Complexity**
2*Offline	Dense eigendecomposition	$O(N^3)$ (if done)
	Sparse Chebyshev (K modes)	$O(K\lvert E\rvert)$
4*Per-Round	Chebyshev features	$O(K\lvert E\rvert)$
	Communication (K coefficients)	$O(K\lvert\mathcal{N}_i\rvert)$
	Aggregation (per mode)	$O(KN + K\lvert\mathcal{N}_i\rvert \cdot I_{\text{med}})$
	Reconstruction	$O(KN)$
Total Per-Round		$\mathbf{O(K\lvert E\rvert + KN + K\lvert\mathcal{N}_i\rvert \cdot I_{\text{med}})}$

Table 3: Computational complexity breakdown. I_{med} is the number of Weiszfeld iterations (typically 5–50 for $\epsilon = 10^{-6}$). For sparse graphs with $|\mathcal{N}_i| = O(1)$ and $K \ll N$, per-round cost is dominated by $O(K|E|)$ for Chebyshev filtering.

Key observations:

- **Offline cost** is amortized over all consensus rounds and can use sparse methods for structured graphs;
- **Per-round cost** scales linearly with K (number of modes) and N (semantic domain dimension);
- For typical parameters ($K = O(\log N)$, sparse neighborhoods), per-round cost is $O(N \log N)$;
- Communication scales with K, not N, enabling efficient distributed operation.

4 Main Results I: Robust Aggregation

We establish Byzantine robustness guarantees for the weighted median aggregator applied to scalar spectral coefficients. This is the regime relevant for per-mode consensus in spectral space.

4.1 Weighted Scalar Median

We specialize to the scalar case $d = 1$, which applies to each spectral mode coefficient. In one dimension, the geometric median reduces to the (weighted) median, for which stronger robustness guarantees are available.

Let $\{x_a\}_{a\in\mathcal{A}} \subset \mathbb{R}$ be scalar reports with nonnegative weights $\{w_a\}_{a\in\mathcal{A}}$ satisfying $\sum_a w_a = 1$. Let $H \subset \mathcal{A}$ denote the honest agents and $B = \mathcal{A} \setminus H$ the Byzantine agents. We assume the local Byzantine weight fraction is bounded:

Assumption 4.1 (Local Byzantine Weight)**.** There exists $\rho_w \in [0, 1/2)$ such that

$$\sum_{a\in B} w_a \leq \rho_w < \frac{1}{2}.$$

Remark 4.2 (Adversary Model: Static Byzantine)**.** We consider a *static* Byzantine adversary: Byzantine values $\{c_{b,k}^{(t)}\}_{b\in B}$ may depend on all public information (graph structure, eigenvalues, protocol parameters) and past honest states $\{c_{i,k}^{(s)}\}_{s<t,i\in H}$, but are chosen *before* observing the round-t gossip sampling randomness. Honest noises $\{\xi_{a,k}\}_{a\in H}$ are i.i.d. sub-Gaussian across agents and rounds, independent of Byzantine choices.

Extension to *fully adaptive* adversaries (who observe sampling outcomes before committing Byzantine values) would require additional assumptions on neighborhood size concentration or randomized commitment schemes. The static model captures realistic scenarios where message latency prevents real-time adaptation to sampling outcomes.

Definition 4.3 (Weighted Scalar Median)**.** A *weighted scalar median* $\mathrm{med}_w(x)$ is any value $m \in \mathbb{R}$ such that the total weight on either side does not exceed 1/2:

$$\sum_{a:x_a\le m} w_a \;\ge\; \tfrac{1}{2}, \qquad \sum_{a:x_a\ge m} w_a \;\ge\; \tfrac{1}{2}.$$

When the set $\{a : x_a = m\}$ has positive total weight, any m in the interval of medians is acceptable; we take the midpoint for uniqueness.

Remark 4.4 (Terminology: Scalar vs. Spatial Median)**.** We use the term "weighted scalar median" to emphasize that aggregation occurs in $\mathbb{R}^1$ (one-dimensional space). In higher dimensions ($d > 1$), the analogous robust estimator is the *weighted spatial median* (also called weighted geometric median), which minimizes $\sum_a w_a\|x - x_a\|_2$ with unsquared Euclidean norm. In our protocol, each spectral coefficient $c_k \in \mathbb{R}$ is scalar, so we apply the weighted scalar median mode-by-mode. This achieves the optimal 1/2 breakdown threshold, superior to the $O(1/\sqrt{d})$ degradation in high-dimensional spatial median applications.

We use the weighted median as our per-mode robust aggregator.

Theorem 4.5 (Weighted Scalar Median Robustness)**.** *Let $\theta \in \mathbb{R}$ be the (unknown) honest target value and assume honest reports satisfy*

$$x_a = \theta + \xi_a, \qquad a \in H,$$

with $\{\xi_a\}_{a\in H}$ independent, mean-zero, and sub-Gaussian with parameter σ^2, i.e., $\mathbb{E}[\exp(t\xi_a)] \le \exp(\frac{1}{2}\sigma^2t^2)$ for all $t \in \mathbb{R}$. Suppose Assumption 4.1 holds with $\rho_w < 1/2$.

Then with explicit constants $C_1 = 4$ and $C_2 = 2\sqrt{2}$, for any $\delta \in (0, 1)$ and $N \ge 2$, with probability at least $1 - \delta$ the weighted median $m = \mathrm{med}_w(x)$ satisfies

$$|m - \theta| \;\le\; 4\sigma\sqrt{\frac{\rho_w}{1-\rho_w}} \;+\; 2\sqrt{2}\sigma\sqrt{\frac{\log(2/\delta)}{N}},$$

where $N = |H|$ is the number of honest agents.

In particular, the breakdown point of the weighted median, in terms of the adversarial weight fraction ρ_w, is 1/2: as long as $\rho_w < 1/2$ the estimator remains bounded, and its error grows smoothly with ρ_w.

Proof sketch. The first term bounds the bias induced by Byzantine weight via a Hoeffding-style inequality on the honest empirical CDF, using the fact that the median cannot be pushed beyond a $(1 - 2\rho_w)$-quantile of the honest distribution. The second term bounds concentration of the honest sub-Gaussian order statistics around the median. □

Corollary 4.6 (Mode-wise Application)**.** *For mode k, let honest agents satisfy $x_{a,k} = \theta_k + \xi_{a,k}$ where $\theta_k = \phi_k^\top f^\star$ is the ground truth coefficient and $\{\xi_{a,k}\}_{a\in H}$ are independent sub-Gaussian noise with parameter σ_k^2. The weighted median m_k of $\{x_{a,k}\}_{a\in\mathcal{A}}$ satisfies, with probability at least $1 - \delta$:*

$$|m_k - \theta_k| \le 4\sigma_k\sqrt{\frac{\rho_w}{1-\rho_w}} + 2\sqrt{2}\sigma_k\sqrt{\frac{\log(2/\delta)}{N}} =: C_{\mathrm{med}}\sigma_k,$$

where $C_{\mathrm{med}} = 4\sqrt{\rho_w/(1-\rho_w)} + 2\sqrt{2}\sqrt{\log(2/\delta)/N}$ and $N = |H|$.

Remark 4.7 (Dimensionality and Breakdown Points)**.** Since each mode aggregation is in $\mathbb{R}^1$ (scalar coefficients), we achieve the optimal **50% breakdown point** with tight constants. This contrasts with ambient-space geometric median in $\mathbb{R}^n$, which would suffer an $O(\sqrt{n})$ penalty in the error bound.

Moreover, in the protocol (Algorithm 1), each agent i aggregates over its local neighborhood $\mathcal{N}_i^{(t)}$. Under Assumption 2.18, the local Byzantine fraction $\rho_i = |\mathcal{N}_i^{(t)} \cap V_B|/|\mathcal{N}_i^{(t)}| < 1/2$ ensures that the weighted median in each neighborhood is robust. The global bound $\rho_w < 1/2$ on total Byzantine weight is used for the overall protocol analysis, but the per-agent robustness depends on the *local* fraction in each neighborhood.

4.2 Trimmed Weighted Mean

The weighted median provides robustness but requires iterative optimization (Weiszfeld's algorithm). We follow it with distance-based trimming to enable fast linear aggregation.

Theorem 4.8 (Trimmed Mean Bound)**.** *Under the setup of Theorem 4.5, let ρ_w denote the total Byzantine weight, and assume $\rho_w < 1/2$. Fix a trimming parameter α satisfying*

$$\rho_w \le \alpha \le 1 - 2\rho_w.$$

Given scalar values $\{c_a\}_{a\in\mathcal{A}}$ with weights $\{w_a\}_{a\in\mathcal{A}}$ such that $\sum_a w_a = 1$, compute:

1. *weighted median m_k;*
2. *distances $d_a = |c_a - m_k|$ for all $a \in \mathcal{A}$;*
3. *trim agents with largest d_a (breaking ties arbitrarily) until the total trimmed weight W_T satisfies $W_T \le \alpha$ and adding any further agent would make $W_T > \alpha$;*
4. *let $S \subset \mathcal{A}$ be the remaining (kept) agents;*
5. *compute the trimmed mean with renormalized weights $\widetilde{w}_a = w_a / \sum_{j\in S} w_j$:*

$$\hat{c} = \sum_{a\in S} \widetilde{w}_a c_a.$$

Let $C_{\text{med}} = C_1\sqrt{\rho_w/(1-\rho_w)} + C_2\sqrt{\log(2/\delta)/N}$ be the weighted median constant from Theorem 4.5 (with $C_1 = 4$, $C_2 = 2\sqrt{2}$), and define

$$R_H := \sigma(1 + C_{\text{med}}).$$

Then

$$|\hat{c} - \theta| \leq R_H.$$

Proof. Let H and B denote the sets of honest and Byzantine agents, respectively, with total Byzantine weight $\rho_w = \sum_{b\in B} w_b$.

Step 1: Honest agents are close to the median. By assumption, for each honest agent $h \in H$,

$$|c_h - \theta| \leq \sigma.$$

By Theorem 4.5, the weighted median m_k satisfies

$$|m_k - \theta| \leq C_{\text{med}}\sigma.$$

Hence for any honest $h \in H$,

$$d_h = |c_h - m_k| \leq |c_h - \theta| + |\theta - m_k| \leq \sigma + C_{\text{med}}\sigma = R_H.$$

Thus every honest agent lies in the interval $\{a : d_a \leq R_H\}$ around m_k.

Step 2: Trimming preserves an honest majority in weight. Let

$$W_H = \sum_{h\in H} w_h = 1 - \rho_w, \qquad W_T = \sum_{a\notin S} w_a \leq \alpha,$$

be the total honest and trimmed weights, respectively. In the worst case, all trimmed weight is honest, so the honest weight remaining in S satisfies

$$W_{S_H} := \sum_{a\in S\cap H} w_a \geq W_H - W_T \geq (1-\rho_w) - \alpha.$$

Byzantine weight in S is at most

$$W_{S_B} := \sum_{a\in S\cap B} w_a \leq \rho_w.$$

Total kept weight is

$$W_S := \sum_{a\in S} w_a = 1 - W_T \geq 1 - \alpha.$$

Therefore the *honest weight fraction* after trimming satisfies

$$\widetilde{W}_{S_H} := \frac{W_{S_H}}{W_S} \geq \frac{1-\rho_w-\alpha}{1-\alpha}.$$

Using $\alpha \leq 1 - 2\rho_w$ and $\rho_w < 1/2$,

$$\frac{1-\rho_w-\alpha}{1-\alpha} \geq \frac{1-\rho_w-(1-2\rho_w)}{1-(1-2\rho_w)} = \frac{\rho_w}{2\rho_w} = \frac{1}{2}.$$

Hence

$$\widetilde{W}_{S_H} \geq \frac{1}{2}, \qquad \widetilde{W}_{S_B} := \frac{W_{S_B}}{W_S} = 1 - \widetilde{W}_{S_H} \leq \frac{1}{2}.$$

Step 3: Kept Byzantine agents cannot be too far from m_k. Define the "far" set

$$F := \{a \in \mathcal{A} : d_a > R_H\}.$$

By Step 1, all honest agents have $d_h \leq R_H$, so $F \subseteq B$ and its total weight satisfies

$$W_F := \sum_{a \in F} w_a \leq \rho_w.$$

The trimming procedure removes agents in order of decreasing d_a until the total trimmed weight reaches some $W_T \leq \alpha$, and by assumption $\alpha \geq \rho_w \geq W_F$. Thus the algorithm can remove all of F without exceeding the trimming budget α; since we always remove the farthest available agents first, every agent in F is trimmed. Hence

$$F \cap S = \emptyset,$$

and every kept agent $a \in S$ satisfies $d_a \leq R_H$.

For any kept Byzantine agent $b \in S \cap B$, we therefore have

$$|c_b - \theta| \leq |c_b - m_k| + |m_k - \theta| \leq R_H + C_{\text{med}}\sigma = \sigma(1 + 2C_{\text{med}}).$$

Step 4: Bound the trimmed mean. We can now bound the deviation of the trimmed mean:

$$\begin{aligned} |\hat{c} - \theta| &= \left| \sum_{a \in S} \widetilde{w}_a (c_a - \theta) \right| \\ &\leq \sum_{a \in S_H} \widetilde{w}_a \, |c_a - \theta| + \sum_{a \in S_B} \widetilde{w}_a \, |c_a - \theta| \\ &\leq \widetilde{W}_{S_H} \cdot \sigma + \widetilde{W}_{S_B} \cdot \sigma(1 + 2C_{\text{med}}), \end{aligned}$$

where $\widetilde{W}_{S_H} = \sum_{a \in S_H} \widetilde{w}_a$ and similarly for $\widetilde{W}_{S_B}$. Using $\widetilde{W}_{S_H} + \widetilde{W}_{S_B} = 1$ and $\widetilde{W}_{S_B} \leq 1/2$, we get

$$\begin{aligned} |\hat{c} - \theta| &\leq \sigma + \widetilde{W}_{S_B} \, \sigma(2C_{\text{med}}) \\ &\leq \sigma + \frac{1}{2}\sigma(2C_{\text{med}}) \\ &= \sigma(1 + C_{\text{med}}) \\ &= R_H. \end{aligned}$$

This completes the proof. □

Remark 4.9. In the spectral protocol we use mode-dependent trimming levels α_k satisfying $\rho_w \leq \alpha_k \leq 1 - 2\rho_w$ (e.g. $\alpha_k = \text{clip}(\eta\lambda_k, \rho_w, 1 - 2\rho_w)$ for a small stepsize η). This only changes constants, not the asymptotic dependence on ρ_w or σ_k.

Constants summary:

- $C_{\text{med}} = 4\sqrt{\rho_w/(1-\rho_w)} + 2\sqrt{2}\sqrt{\log(2/\delta)/N}$ (weighted median, Theorem 4.5);
- $C_{\text{trim}} = 1 + C_{\text{med}}$ (trimmed mean).

For $\rho_w = 0.4$, $\delta = 0.01$, and $N = 100$, we have $C_{\text{med}} \approx 5.6$ and $C_{\text{trim}} \approx 6.6$. Note that the breakdown threshold is $\rho_w < 1/2$ (optimal for affine-equivariant estimators).

4.3 Local-to-Global Byzantine Robustness

We now establish how neighborhood-level Byzantine bounds guarantee global convergence. First, we clarify the relationship between global and local Byzantine fractions.

Lemma 4.10 (Local-Global Byzantine Fraction Concentration)**.** *Under uniform stake weights $w_a = 1/N$ and minimum neighborhood size $|\mathcal{N}_i| \geq d_{\min}$ for all honest agents $i \in H$, if the global Byzantine fraction is $\rho_w = |B|/N$, then by Chernoff bounds, with probability at least $1 - N\exp(-\Omega(d_{\min}))$, every local Byzantine fraction satisfies:*

$$\rho_i := \frac{|\mathcal{N}_i \cap B|}{|\mathcal{N}_i|} \leq \rho_w + O\left(\sqrt{\frac{\log N}{d_{\min}}}\right).$$

For dense neighborhoods with $d_{\min} = \Omega(\log N)$, we have $\max_{i \in H} \rho_i \approx \rho_w$ with high probability.

Proof sketch. For each honest agent i, the local Byzantine fraction ρ_i is the empirical mean of $|\mathcal{N}_i|$ indicator variables. By the multiplicative Chernoff bound, $\Pr[\rho_i > \rho_w(1+\delta)] \leq \exp(-\delta^2 \rho_w d_{\min}/3)$. Setting $\delta = \sqrt{3\log N/(\rho_w d_{\min})}$ and taking a union bound over N agents gives the result. □

Remark 4.11. For simplicity, the analysis assumes a network design guaranteeing $\rho_i \leq \rho_{\text{loc}} < 1/2$ for all $i \in H$ (Assumption 2.18 in Section 2), which holds under the above concentration when $\rho_w < 1/2 - \epsilon$ for some margin $\epsilon > 0$ and $d_{\min} = \Omega(\log N/\epsilon^2)$.

Theorem 4.12 (Local-to-Global Byzantine Robustness)**.** *Assume:*

- *Each honest agent i has neighborhood $\mathcal{N}_i$ with local Byzantine fraction $\rho_i < 1/2$;*
- *Honest subgraph G_H is connected with spectral gap $\lambda_\star > 0$;*
- *Agents use weighted scalar median locally (each mode $c_k \in \mathbb{R}$ is aggregated independently).*

Then the global consensus error satisfies:

$$\mathbb{E}\left[\max_{i \in H}\left|c_i^{(t)} - \theta\right|^2\right] \leq (1-\lambda_\star)^t D_0^2 + \frac{C^2\sigma^2}{\lambda_\star},$$

where $C = 4\sqrt{\max_i \rho_i/(1-\rho_i)} + 2\sqrt{2\log(1/\delta)/\min_i |\mathcal{N}_i|}$ and D_0^2 is the initial disagreement.

Proof sketch. The key mechanism: Local robustness in each neighborhood propagates through gossip dynamics on the honest subgraph. The global convergence rate is governed by the spectral gap $\lambda_\star$, while the steady-state error depends on the worst-case local Byzantine density $\max_i \rho_i$.

Step 1: Apply Theorem 4.5 in each neighborhood to bound local aggregation error.

Step 2: Model gossip on G_H as a Markov chain with mixing time $O(\lambda_\star^{-1})$.

Step 3: Show that local errors compose via triangle inequality, with accumulation bounded by $C\sigma$.

Step 4: Use Lyapunov analysis with variance functional $V^{(t)} = \sum_{i\in H} \left| c_i^{(t)} - \bar{c}^{(t)} \right|^2$.

Step 5: Prove one-step contraction: $\mathbb{E}[V^{(t+1)}] \leq (1-\lambda_\star)V^{(t)} + C^2\sigma^2$.

Step 6: Unroll recursion to obtain stated bound. □

Remark 4.13. This theorem formalizes how neighborhood-level Byzantine bounds (Assumption 2.18) guarantee global convergence despite adversarial agents. The result applies to any mode k by setting $\theta = \theta_k$ (ground truth coefficient for mode k) and $\sigma = \sigma_k$ (noise level for that mode).

4.4 Aggregate Reconstruction Error

The per-mode robustness bounds (Corollary 4.6) combine via Parseval's identity to yield an L^2 reconstruction guarantee.

Theorem 4.14 (Aggregate L^2 Reconstruction Bound)**.** *Let $\hat{f}_K = \sum_{k=0}^{K-1} \hat{c}_k\phi_k$ be the K-mode reconstruction from consensus coefficients $\{\hat{c}_k\}$, and let $f^\star = \sum_{k=0}^{N-1} c_k^\star\phi_k$ be the ground truth signal. Under Assumptions 2.14–2.18 with $\rho_w < 1/2$, the reconstruction error satisfies:*

$$\|\hat{f}_K - f^\star\|_{L^2}^2 \;=\; \underbrace{\sum_{k=0}^{K-1} |\hat{c}_k - c_k^\star|^2}_{\textit{consensus error}} \;+\; \underbrace{\sum_{k\geq K} |c_k^\star|^2}_{\textit{truncation error}} \;.$$

Applying Corollary 4.6 to each mode:

$$\|\hat{f}_K - f^\star\|_{L^2}^2 \;\leq\; \sum_{k=0}^{K-1} C_{\mathrm{med}}^2\sigma_k^2 \;+\; \mathcal{E}_{>K}(f^\star),$$

where $\mathcal{E}_{>K}(f^\star) := \sum_{k\geq K} |c_k^\star|^2$ is the tail energy.

For homogeneous noise $\sigma_k = \sigma$ across modes:

$$\|\hat{f}_K - f^\star\|_{L^2} \;\leq\; C_{\mathrm{med}}\sigma\sqrt{K} \;+\; \sqrt{\mathcal{E}_{>K}(f^\star)}.$$

Proof. By Parseval's identity for the orthonormal eigenbasis $\{\phi_k\}$:

$$\|\hat{f}_K - f^\star\|_{L^2}^2 = \left\| \sum_{k=0}^{K-1} (\hat{c}_k - c_k^\star)\phi_k - \sum_{k\geq K} c_k^\star\phi_k \right\|^2 = \sum_{k=0}^{K-1} |\hat{c}_k - c_k^\star|^2 + \sum_{k\geq K} |c_k^\star|^2.$$

The first sum is bounded by applying Corollary 4.6 independently to each mode and using the union bound (or Bonferroni correction for K modes with failure probability δ/K each). □

Remark 4.15 (Cross-Mode Coordination)**.** A Byzantine adversary with knowledge of $\{\phi_k\}$ could coordinate attacks across modes, injecting bias at the $\sim (1/2 - \rho_w)$ quantile in each mode simultaneously. However, this produces **coherent bias**, not unbounded error: each mode's median shift is still bounded by $C_{\text{med}}\sigma_k$. The aggregate bias $\|\text{bias}\|_{L^2} \leq C_{\text{med}}\sigma\sqrt{K}$ may be semantically meaningful (e.g., shifting all concepts in a consistent direction), but the **magnitude** remains controlled.

This distinguishes spectral aggregation from parameter-space methods: the bias is *structured* (expressible in K modes) rather than arbitrary, enabling downstream detection via Dirichlet energy bounds (Theorem 6.3).

5 Main Results II: Convergence Analysis

5.1 Gossip Contraction with Byzantine Noise

We model the gossip process on the honest subgraph G_H.

Assumption 5.1 (Honest Gossip Graph)**.** The honest agents H communicate over a connected graph $G_H = (H, E_H)$ with normalized Laplacian L_H. The gossip process is a lazy random walk on G_H with transition matrix:

$$P = I - \frac{1}{2}L_H$$

(lazy random walk with spectral gap $\lambda_\star := \lambda_1(L_H) > 0$).

Byzantine agents send values but do not appear in neighbor sampling (i.e., agents only sample neighbors from H).

Remark 5.2. We consider a standard Byzantine adversary that controls the values sent by agents in B, but cannot alter the topology or randomness of the honest gossip process on G_H. This aligns the model with the contraction properties of the lazy random walk on the honest subgraph.

Lemma 5.3 (Weighted Gossip Contraction)**.** *Under Assumption 5.1, for mode k at round t, let:*

- $c_{i,k}^{(t)}$*: agent i's current coefficient;*
- $\bar{c}_k^{(t)} = |H|^{-1}\sum_{i\in H} c_{i,k}^{(t)}$*: honest average;*
- $\tilde{c}_{i,k}^{(t)}$*: aggregated value from protocol (includes Byzantine contributions).*

Define the Byzantine noise:

$$\xi_{i,k}^{(t)} := \tilde{c}_{i,k}^{(t)} - \sum_{j\in H} P_{ij} c_{j,k}^{(t)}.$$

Then:

$$\mathbb{E}\big[\big|c_{i,k}^{(t)} - \bar{c}_k^{(t)}\big|^2\big] \leq (1-\lambda_\star)\big|c_{i,k}^{(t)} - \bar{c}_k^{(t)}\big|^2 + \mathbb{E}\Big[\Big|\xi_{i,k}^{(t)}\Big|^2\Big],$$

where expectation is over the gossip randomness.

Moreover, by Theorem 4.8, the Byzantine perturbation satisfies

$$\mathbb{E}\big[\big|\xi_{i,k}^{(t)}\big|^2\big] \leq R_H^2 = \sigma_k^2(1 + C_{\mathrm{med}})^2 =: \sigma_\xi^2,$$

where $C_{\mathrm{med}} = C_1\sqrt{\rho_w/(1-\rho_w)} + C_2\sqrt{\log(1/\delta)}$.

Proof. Split the aggregation into honest gossip and Byzantine perturbation:

$$\tilde{c}_{i,k}^{(t)} = \sum_{j \in H} P_{ij} c_{j,k}^{(t)} + \xi_{i,k}^{(t)}.$$

For the honest gossip term, by the standard Boyd et al. (2006) analysis of lazy random walk:

$$\mathbb{E}_{\mathrm{gossip}}\left[\left(\sum_{j \in H} P_{ij} c_{j,k}^{(t)} - \bar{c}_k^{(t)}\right)^2\right] \leq (1-\lambda_\star)\big(c_{i,k}^{(t)} - \bar{c}_k^{(t)}\big)^2.$$

For the Byzantine term, $\xi_{i,k}^{(t)}$ represents the deviation introduced by trimmed aggregation. By Theorem 4.8, if the true honest average in the sampled neighborhood is $\bar{c}_{\mathrm{local}}$, then

$$\left|\xi_{i,k}^{(t)}\right| \leq C_{\mathrm{trim}}\sigma_k\sqrt{\frac{\rho_w}{1-\rho_w}}.$$

Taking expectations and using independence of gossip randomness and Byzantine perturbations (Byzantine agents do not know which neighbors were sampled),

$$\mathbb{E}\big[\big|\tilde{c}_{i,k}^{(t)} - \bar{c}_k^{(t)}\big|^2\big] \leq (1-\lambda_\star)\big|c_{i,k}^{(t)} - \bar{c}_k^{(t)}\big|^2 + \sigma_\xi^2.$$

□

5.2 Mode 0: Value Consensus

Mode 0 represents the mean value and requires separate treatment due to $\lambda_0 = 0$. Since we cannot use $\alpha_0 = \eta\lambda_0 = 0$ (which would freeze the coefficient), the protocol uses direct assignment $\alpha_0 = 1$ (Algorithm 1, line 10).

Theorem 5.4 (Mode 0 Convergence)**.** *Under Assumption 5.1, for mode $k = 0$ using direct weighted median assignment with $\alpha_0 = 1$:*

$$c_{i,0}^{(t+1)} = \tilde{c}_{i,0}^{(t)} = \mathrm{med}_w(\{c_{j,0}^{(t)}\}_{j \in \mathcal{N}_i^{(t)} \cup \{i\}}),$$

the expected squared error satisfies, with probability at least $1-\delta$:

$$\mathbb{E}\left[\left|c_{i,0}^{(t)} - \theta_0\right|^2\right] \leq (1-\lambda_\star)^t D_0^2 + C_{\mathrm{med}}^2\sigma_0^2,$$

where:

- *$D_0 = \max_{i,j \in H} |c_{i,0}^{(0)} - c_{j,0}^{(0)}|$ is initial disagreement among honest agents;*
- *$\theta_0 = \phi_0^\top f^\star = \sqrt{N} \cdot \mathrm{mean}(f^\star)$ is the ground truth mode-0 coefficient;*
- *$\lambda_\star = \lambda_1(L_H) > 0$ is the spectral gap of the honest gossip graph G_H;*
- *$C_{\mathrm{med}} = C_1\sqrt{\rho_w/(1-\rho_w)} + C_2\sqrt{\log(1/\delta)}$ from Theorem 4.5.*

Convergence time to ϵ-consensus is:

$$T_\epsilon = O\left(\frac{1}{\lambda_\star} \log \frac{D_0}{\epsilon}\right),$$

*which depends on the **gossip graph spectral gap** $\lambda_\star$, **not** the heat flow parameter η.*

Proof. The update rule for mode 0 is direct assignment:

$$c_{i,0}^{(t+1)} = \tilde{c}_{i,0}^{(t)} = \mathrm{med}_w(\{c_{j,0}^{(t)}\}_{j \in \mathcal{N}_i^{(t)} \cup \{i\}}).$$

By Theorem 4.5, the weighted median satisfies, with probability at least $1-\delta$:

$$\left|\tilde{c}_{i,0}^{(t)} - \theta_0\right| \leq C_{\mathrm{med}}\sigma_0,$$

where $C_{\mathrm{med}} = C_1\sqrt{\rho_w/(1-\rho_w)} + C_2\sqrt{\log(1/\delta)}$.

Define the deviation from honest average: $e_{i,0}^{(t)} := c_{i,0}^{(t)} - \bar{c}_0^{(t)}$ where $\bar{c}_0^{(t)} = |H|^{-1}\sum_{j \in H} c_{j,0}^{(t)}$.

From the standard gossip averaging analysis (Boyd et al., 2006), the lazy random walk on G_H with spectral gap $\lambda_\star$ satisfies:

$$\mathbb{E}\left[\left(e_{i,0}^{(t+1)}\right)^2\right] \leq (1-\lambda_\star)\left(e_{i,0}^{(t)}\right)^2 + C_{\mathrm{med}}^2\sigma_0^2.$$

Unrolling the recursion:

$$\mathbb{E}\left[\left(e_{i,0}^{(t)}\right)^2\right] \leq (1-\lambda_\star)^t\left(e_{i,0}^{(0)}\right)^2 + C_{\mathrm{med}}^2\sigma_0^2\sum_{s=0}^{t-1}(1-\lambda_\star)^s.$$

The geometric series converges:

$$\sum_{s=0}^{t-1}(1-\lambda_\star)^s = \frac{1-(1-\lambda_\star)^t}{\lambda_\star} < \frac{1}{\lambda_\star}.$$

Therefore, with high probability:

$$\mathbb{E}\left[\left(c_{i,0}^{(t)} - \theta_0\right)^2\right] \leq (1-\lambda_\star)^t D_0^2 + C_{\mathrm{med}}^2\sigma_0^2.$$

For ϵ-consensus, we require $(1-\lambda_\star)^t D_0^2 \leq \epsilon^2$, giving:

$$t \geq \frac{\log(D_0/\epsilon)}{\log(1/(1-\lambda_\star))} \approx \frac{1}{\lambda_\star}\log\frac{D_0}{\epsilon}$$

using $\log(1/(1-x)) \approx x$ for small $x = \lambda_\star$. □

Remark 5.5. The key distinction: mode 0 convergence depends on the **gossip graph spectral gap** $\lambda_\star = \lambda_1(L_H)$, not the heat flow parameter η. This is why mode 0 cannot use $\alpha_0 = \eta\lambda_0 = 0$ (which would freeze it) and instead requires $\alpha_0 = 1$ (direct assignment). The convergence is entirely governed by the mixing time of the gossip process on the honest subgraph.

5.3 Heat Flow Convergence

Theorem 5.6 (Mode-wise Exponential Decay for Structural Modes)**.** *For structural modes $k \in \{1, \ldots, K-1\}$, under Assumption 5.1 and with update rule:*

$$c_{i,k}^{(t+1)} = (1-\alpha_k)c_{i,k}^{(t)} + \alpha_k \tilde{c}_{i,k}^{(t)},$$

where $\alpha_k = \eta\lambda_k$ with $\eta \in (0, 1/\lambda_{\max})$, the expected squared error satisfies:

$$\mathbb{E}\big[\big|c_{i,k}^{(t)} - c_k^\infty\big|^2\big] \le e^{-2\eta\lambda_k t} D_0^2 + C_{\text{trim}}^2 \sigma_k^2 \frac{\rho_w}{1-\rho_w}\big(1 - e^{-\eta\lambda_k t}\big)^2,$$

where:

- $D_0 = \max_{i,j \in H} \left|c_{i,k}^{(0)} - c_{j,k}^{(0)}\right|$ *is the initial disagreement;*
- c_k^∞ *is the limiting consensus value (Byzantine-perturbed equilibrium).*

In particular, the steady-state variance is bounded by:

$$\lim_{t\to\infty} \mathbb{E}\big[\big|c_{i,k}^{(t)} - c_k^\star\big|^2\big] \le C_{\text{trim}}^2 \sigma_k^2 \frac{\rho_w}{1-\rho_w},$$

where $c_k^\star$ is the ground truth coefficient.

Proof. Define the deviation from the honest average:

$$e_{i,k}^{(t)} := c_{i,k}^{(t)} - \bar{c}_k^{(t)}.$$

From the update rule,

$$\begin{aligned} e_{i,k}^{(t+1)} &= (1-\alpha_k)e_{i,k}^{(t)} + \alpha_k\big(\tilde{c}_{i,k}^{(t)} - \bar{c}_k^{(t)}\big) \\ &= (1-\alpha_k)e_{i,k}^{(t)} + \alpha_k\left(\sum_{j\in H} P_{ij}(c_{j,k}^{(t)} - \bar{c}_k^{(t)}) + \xi_{i,k}^{(t)}\right). \end{aligned}$$

By Lemma 5.3,

$$\mathbb{E}\Big[\Big|e_{i,k}^{(t+1)}\Big|^2\Big] \le (1-\alpha_k)^2\left|e_{i,k}^{(t)}\right|^2 + \alpha_k^2\big[(1-\lambda_\star)\left|e_{i,k}^{(t)}\right|^2 + \sigma_\xi^2\big].$$

Simplifying,

$$\begin{aligned} \mathbb{E}\Big[\Big|e_{i,k}^{(t+1)}\Big|^2\Big] &\le \big[(1-\alpha_k)^2 + \alpha_k^2(1-\lambda_\star)\big]\left|e_{i,k}^{(t)}\right|^2 + \alpha_k^2\sigma_\xi^2 \\ &= \big[1 - 2\alpha_k + \alpha_k^2 + \alpha_k^2(1-\lambda_\star)\big]\left|e_{i,k}^{(t)}\right|^2 + \alpha_k^2\sigma_\xi^2 \\ &= \big[1 - 2\alpha_k + \alpha_k^2(2-\lambda_\star)\big]\left|e_{i,k}^{(t)}\right|^2 + \alpha_k^2\sigma_\xi^2. \end{aligned}$$

For small $\alpha_k = \eta\lambda_k \ll 1$, the contraction coefficient is approximately

$$1 - 2\alpha_k \approx e^{-2\alpha_k} = e^{-2\eta\lambda_k}.$$

Unrolling the recursion,

$$\begin{aligned}\mathbb{E}\Big[\Big|e_{i,k}^{(t)}\Big|^2\Big] &\le e^{-2\eta\lambda_k t}\Big|e_{i,k}^{(0)}\Big|^2 + \sigma_\xi^2 \sum_{s=0}^{t-1} \alpha_k^2 e^{-2\eta\lambda_k(t-s)} \\ &\le e^{-2\eta\lambda_k t} D_0^2 + \sigma_\xi^2 \alpha_k^2 \frac{1 - e^{-2\eta\lambda_k t}}{1 - e^{-2\eta\lambda_k}}.\end{aligned}$$

For small α_k, we have

$$\frac{\alpha_k^2}{1 - e^{-2\eta\lambda_k}} = O(\alpha_k),$$

and we obtain the stated bound (up to a mild constant relabeling) with the noise term behaving like $\sigma_\xi^2(1 - e^{-\eta\lambda_k t})^2$.

Substituting $\sigma_\xi^2 = C_{\mathrm{trim}}^2 \sigma_k^2 \rho_w/(1 - \rho_w)$ completes the proof. □

Remark 5.7 (Finite Step-Size Effects)**.** The rate $e^{-2\eta\lambda_k t}$ is asymptotically exact as $\eta \to 0$. For practical step sizes, second-order corrections of $O(\eta^2\lambda_k^2(2 - \lambda_\star))$ reduce the effective rate. Empirical validation confirms the qualitative scaling with correlation > 0.97 between λ_k and decay rate (see validation notebook).

5.4 Anisotropic Lyapunov Contraction for Structural Modes

We establish global convergence for structural modes $k \ge 1$ by defining a weighted variance functional. Mode 0 is analyzed separately (Theorem 5.4) since it uses a different update rule.

Definition 5.8 (Weighted Variance Lyapunov Function for Structural Modes)**.** Define the per-mode variance among honest agents:

$$V_k^{(t)} := \frac{1}{|H|} \sum_{i \in H} \big(c_{i,k}^{(t)} - \bar{c}_k^{(t)}\big)^2,$$

and the weighted total variance over **structural modes only**:

$$V^{(t)} := \sum_{k=1}^{K-1} w_k V_k^{(t)},$$

where $w_k = (1 + \lambda_k)^{-\beta}$ for some $\beta > 0$ (mode weighting exponent).

Theorem 5.9 (Global Lyapunov Contraction for Structural Modes)**.** *Under the heat flow protocol (Algorithm 1) with $\alpha_k = \eta\lambda_k$ for $k \ge 1$, the weighted variance over structural modes satisfies:*

$$\mathbb{E}\big[V^{(t+1)}\big] \le (1 - \gamma)V^{(t)} + \Delta_B,$$

where:

- *contraction coefficient $\gamma \approx 2\eta\lambda_1$ (depends on lowest structural mode λ_1);*
- *Byzantine noise floor $\Delta_B = C_{\text{med}}^2 \sum_{k=1}^{K-1} w_k \sigma_k^2$.*

Unrolling,

$$V^{(t)} \leq e^{-\gamma t} V^{(0)} + \frac{\Delta_B}{\gamma}\big(1 - e^{-\gamma t}\big).$$

Convergence time to ϵ-consensus for structural modes is

$$T_\epsilon = O\left(\frac{1}{\gamma} \log \frac{V^{(0)}}{\epsilon}\right) = O\left(\frac{1}{2\eta\lambda_1} \log \frac{V^{(0)}}{\epsilon}\right),$$

where $V^{(0)}$ depends on initial disagreement across honest agents.

Proof. From Theorem 5.6, for each structural mode $k \geq 1$, the per-mode variance satisfies (averaging over agents):

$$\mathbb{E}\big[V_k^{(t+1)}\big] \leq (1 - 2\alpha_k) V_k^{(t)} + \alpha_k^2 \sigma_{\xi,k}^2,$$

where we use the small-α_k approximation $(1 - 2\alpha_k + \alpha_k^2(2 - \lambda_\star)) \approx (1 - 2\alpha_k)$.

Multiplying by w_k and summing over structural modes $k = 1, \ldots, K-1$,

$$\begin{aligned}\mathbb{E}\big[V^{(t+1)}\big] &= \sum_{k=1}^{K-1} w_k \mathbb{E}\big[V_k^{(t+1)}\big] \\ &\leq \sum_{k=1}^{K-1} w_k (1 - 2\alpha_k) V_k^{(t)} + \sum_{k=1}^{K-1} w_k \alpha_k^2 \sigma_{\xi,k}^2 \\ &= \sum_{k=1}^{K-1} w_k V_k^{(t)} - 2 \sum_{k=1}^{K-1} w_k \alpha_k V_k^{(t)} + \Delta_B.\end{aligned}$$

Lower-bound the contraction term using $\alpha_k = \eta\lambda_k$:

$$\sum_{k=1}^{K-1} w_k \alpha_k V_k^{(t)} \geq \min_{k \geq 1}(\alpha_k) \sum_{k=1}^{K-1} w_k V_k^{(t)} = \eta\lambda_1 V^{(t)},$$

where $\lambda_1 > 0$ is the first structural mode (spectral gap of the semantic graph).

Therefore,

$$\mathbb{E}\big[V^{(t+1)}\big] \leq \big(1 - 2\eta\lambda_1\big) V^{(t)} + \Delta_B.$$

Setting $\gamma = 2\eta\lambda_1$, the recursion becomes:

$$\mathbb{E}\big[V^{(t+1)}\big] \leq (1 - \gamma) V^{(t)} + \Delta_B.$$

At steady-state: $V^{(\infty)} \leq \Delta_B/\gamma = \Delta_B/(2\eta\lambda_1)$. The unrolling and convergence time follow from standard geometric series analysis. □

Corollary 5.10 (Mode-Dependent Convergence Rates)**.** *For $\rho_w < 1/2$ and $\sigma_k^2 = \sigma_0^2/\lambda_k$ (natural noise scaling),*

$$\Delta_B = C_{\mathrm{med}}^2 \sigma_0^2 \sum_{k=1}^{K-1} \frac{w_k}{\lambda_k}.$$

For $w_k = (1+\lambda_k)^{-\beta}$ and uniformly spaced eigenvalues $\lambda_k \approx k/K$,

$$\Delta_B \approx C_{\mathrm{med}}^2 \sigma_0^2 \cdot K \cdot \log K,$$

i.e., logarithmic dependence on K.

Remark 5.11 (Summary of Convergence Rates)**.** The protocol exhibits two distinct convergence regimes:

- **Mode 0 (value consensus):** $T_0 = O(\lambda_\star^{-1} \log(D_0/\epsilon))$, governed by gossip mixing;
- **Structural modes $k \geq 1$:** $T_{\mathrm{struct}} = O(\gamma^{-1} \log(V^{(0)}/\epsilon)) = O((2\eta\lambda_1)^{-1} \log(V^{(0)}/\epsilon))$, governed by anisotropic heat flow.

Overall convergence is the maximum: $T_{\mathrm{total}} = \max(T_0, T_{\mathrm{struct}})$.

5.5 Global Consensus Convergence

We now combine the mode 0 and structural mode analyses to state a unified global convergence result.

Theorem 5.12 (Global Consensus Convergence)**.** *Under Assumptions 2.14–5.1 with Byzantine weight $\rho_w < 1/2$, there exists a random limiting coefficient vector $c^\infty \in \mathbb{R}^K$ such that for all honest agents $i \in H$:*

$$\max_{i \in H} \|c_i^{(t)} - c^\infty\|_2 \leq \epsilon$$

for $t \geq T_\epsilon$, where the convergence time is:

$$T_\epsilon := \max\left(\frac{1}{\lambda_\star} \log \frac{D_0}{\epsilon},\ \frac{1}{\eta\lambda_1} \log \frac{V^{(0)}}{\epsilon} \right).$$

The limiting consensus error relative to ground truth satisfies, with probability at least $1-\delta$:

$$\|c^\infty - c^\star\|_2 \leq C_{\mathrm{med}}(\delta) \sqrt{\sum_{k=0}^{K-1} \sigma_k^2},$$

where $c^\star = (\theta_0, \theta_1, \ldots, \theta_{K-1})$ is the ground truth coefficient vector and $C_{\mathrm{med}}(\delta) = C_1\sqrt{\rho_w/(1-\rho_w)} + C_2\sqrt{\log(2K/\delta)/N}$ accounts for union bound over K modes.

Proof. **Step 1 (Mode 0 convergence):** By Theorem 5.4, mode 0 achieves $|c_{i,0}^{(t)} - \theta_0| \leq \epsilon_0$ for $t \geq T_0 = O(\lambda_\star^{-1} \log(D_0/\epsilon_0))$, with steady-state error bounded by $C_{\mathrm{med}}\sigma_0$.

Step 2 (Structural mode convergence): By Theorem 5.9, the weighted variance over structural modes satisfies $V^{(t)} \leq \epsilon^2$ for $t \geq T_{\text{struct}} = O((\eta\lambda_1)^{-1} \log(V^{(0)}/\epsilon^2))$, with per-mode steady-state error bounded by $C_{\text{trim}}\sigma_k\sqrt{\rho_w/(1-\rho_w)}$.

Step 3 (Combine via Parseval): The total ℓ^2 error decomposes as:

$$\|c_i^{(t)} - c^\star\|_2^2 = \sum_{k=0}^{K-1} |c_{i,k}^{(t)} - \theta_k|^2.$$

Applying Corollary 4.6 to each mode with failure probability δ/K (union bound):

$$\sum_{k=0}^{K-1} |c_{i,k}^{(\infty)} - \theta_k|^2 \leq C_{\text{med}}^2(\delta) \sum_{k=0}^{K-1} \sigma_k^2$$

with probability at least $1 - \delta$.

Step 4 (Convergence time): Taking $T_\epsilon = \max(T_0, T_{\text{struct}})$ ensures both mode 0 and structural modes have converged to within ϵ of their limiting values. □

Remark 5.13 (Interpretation)**.** Theorem 5.12 unifies the convergence guarantees: honest agents reach ϵ-consensus on all K spectral coefficients in time $O(\max(\lambda_\star^{-1}, (\eta\lambda_1)^{-1}) \log(1/\epsilon))$, with Byzantine-induced bias bounded by $O(\sqrt{K}\sigma\sqrt{\rho_w/(1-\rho_w)})$. The bias is *structured* (lies in the K-mode spectral subspace) rather than arbitrary, enabling geometric sanity checks via Dirichlet energy bounds (Theorem 6.3).

6 Fork Detection and Dynamic Topology

6.1 Time-Varying Honest Graphs

Assumption 6.1 (Uniform Spectral Gap)**.** The honest communication graph $G_H^{(t)}$ may change over time, but satisfies:

$$\lambda_1(L_H^{(t)}) \geq \lambda_\star > 0 \quad \forall t,$$

a uniform lower bound on spectral gap.

Theorem 6.2 (Dynamic Topology Resilience)**.** *Under Assumption 6.1, Theorems 5.6 and 5.9 hold with spectral gap replaced by $\lambda_\star$ (the worst-case gap). In particular, the convergence time becomes*

$$T_\epsilon = O\left(\frac{1}{\alpha_0\lambda_\star} \log\frac{1}{\epsilon}\right),$$

where the dependence on $\lambda_\star$ reflects the "bottleneck" epoch with poorest connectivity.

Proof. At each round t, apply Lemma 5.3 with the current graph's spectral gap $\lambda_1(L_H^{(t)})$. The variance contraction is:

$$\mathbb{E}\big[V_k^{(t+1)} \mid G_H^{(t)}\big] \leq (1 - 2\alpha_k\lambda_1(L_H^{(t)}))V_k^{(t)} + \sigma_\xi^2.$$

Taking expectations over the sequence of graphs and using $\lambda_1(L_H^{(t)}) \geq \lambda_\star$,

$$\mathbb{E}\big[V_k^{(t+1)}\big] \leq (1 - 2\alpha_k\lambda_\star)V_k^{(t)} + \sigma_\xi^2.$$

The rest of the proof proceeds identically to Theorem 5.9. □

6.2 Intrinsic Fork Detection via Lyapunov Bounds

The anisotropic heat flow dynamics of Theorem 5.9 establish an intrinsic energy ceiling for honest agent disagreement. This enables fork detection *without external energy budget enforcement*: any pair of honest agents whose spectral disagreement exceeds the steady-state bound indicates Byzantine activity.

Theorem 6.3 (Intrinsic Fork Detection (High Probability))**.** *Under Assumptions 2.14–5.1 with Byzantine weight $\rho_w < 1/2$, semantic spectral gap $\lambda_1 > 0$, and gossip spectral gap $\lambda_\star > 0$, define:*

1. ***Contraction rate:*** $\gamma = 2\eta\lambda_1$
2. ***Mixing time:*** $T_{\text{mix}} = \lceil \gamma^{-1} \log(W^{(0)}\gamma/\Delta_B) \rceil$
3. ***Steady-state energy ceiling (confidence** $1-\delta$**):***

$$\mathcal{E}_{\max}(\delta) = \frac{\Delta_B(\delta)}{\gamma} = \frac{C_{\text{trim}}^2(\delta)}{2\eta\lambda_1} \sum_{k=1}^{K-1} w_k \lambda_k \sigma_k^2 \frac{\rho_w}{1-\rho_w}$$

where $w_k = (1+\lambda_k)^{-\beta}$, $\Delta_B(\delta)$ is the Byzantine noise floor from Theorem 5.9, and $C_{\text{trim}}(\delta)$ includes the $\sqrt{\log(1/\delta)}$ term from Theorem 4.5.

Then for any $\delta \in (0,1)$, with probability at least $1-\delta$ over gossip randomness and honest noise, for all $T \geq T_{\text{mix}}$ and any honest pair $i, j \in H$:

$$d_\Lambda^2(i,j;T) := \sum_{k=1}^{K-1} w_k \lambda_k (c_{i,k}^{(T)} - c_{j,k}^{(T)})^2 \leq 4\mathcal{E}_{\max}(\delta)$$

Fork detection criterion (high probability): *Observing any pair (i,j) with weighted spectral disagreement energy $d_\Lambda^2(i,j;T) > 4\mathcal{E}_{\max}(\delta)$ implies, with probability at least $1-\delta$, at least one of:*

(a) Byzantine majority: $\rho_w \geq 1/2$

(b) Insufficient mixing: $T < T_{\text{mix}}$

(c) Agent i or j is Byzantine

Proof. The proof follows from Theorem 5.9 (Lyapunov contraction for structural modes).

Step 1: Variance-to-disagreement bound. For any honest pair $i, j \in H$, the squared spectral distance satisfies:

$$d_\Lambda^2(i,j;T) = \sum_{k=1}^{K-1} w_k \lambda_k (c_{i,k}^{(T)} - c_{j,k}^{(T)})^2 \leq 2 \sum_{k=1}^{K-1} w_k \lambda_k \left[(c_{i,k}^{(T)} - \bar{c}_k^{(T)})^2 + (c_{j,k}^{(T)} - \bar{c}_k^{(T)})^2 \right]$$

by the parallelogram law, where $\bar{c}_k^{(T)} = |H|^{-1} \sum_{i \in H} c_{i,k}^{(T)}$.

Step 2: Apply Lyapunov bound. From Theorem 5.9, the weighted variance functional satisfies:

$$W^{(T)} := \sum_{k=1}^{K-1} w_k \lambda_k V_k^{(T)} \leq (1-\gamma)^T W^{(0)} + \frac{\Delta_B}{\gamma}$$

where $V_k^{(T)} = |H|^{-1} \sum_{i \in H} (c_{i,k}^{(T)} - \bar{c}_k^{(T)})^2$.

Step 3: Steady-state after mixing. For $T \geq T_{\text{mix}} = \lceil \gamma^{-1} \log(W^{(0)}\gamma/\Delta_B) \rceil$:

$$(1-\gamma)^T W^{(0)} \leq W^{(0)} e^{-\gamma T} \leq W^{(0)} \cdot \frac{\Delta_B}{W^{(0)}\gamma} = \frac{\Delta_B}{\gamma}$$

Hence $W^{(T)} \leq 2\Delta_B/\gamma = 2\mathcal{E}_{\max}$.

Step 4: Combine. For honest i, j:

$$\mathbb{E}[d_\Lambda^2(i,j;T)] \leq 2(V_i^{(T)} + V_j^{(T)}) \leq 4W^{(T)} \leq 4\mathcal{E}_{\max}$$

The contrapositive gives the fork detection criterion. □

Remark 6.4 (Thermodynamic Interpretation)**.** The steady-state energy ceiling $\mathcal{E}_{\max}(\delta)$ represents a *statistical equilibrium* of the consensus protocol: Byzantine agents continuously inject adversarial perturbations (heat source), while anisotropic heat flow continuously dissipates them (cooling). The equilibrium temperature is:

$$\mathcal{E}_{\max}(\delta) \propto \frac{\text{injection rate}}{\text{dissipation rate}} = \frac{\rho_w/(1-\rho_w)}{\eta\lambda_1}$$

Important caveat: This is a *statistical* equilibrium, not an absolute guarantee. Honest pairs remain below $\mathcal{E}_{\max}(\delta)$ with probability $1-\delta$, not with certainty. Rare honest noise exceedances (occurring with probability δ) may trigger false positives in fork detection. For high-confidence detection, set δ small (e.g., $\delta = 10^{-6}$), which increases $\mathcal{E}_{\max}(\delta)$ via the $\sqrt{\log(1/\delta)}$ term in C_{trim}.

This provides fork detection without external energy enforcement: the protocol's own dynamics establish the energy ceiling intrinsically.

Remark 6.5 (Computational Complexity of Fork Detection)**.** Computing $d_\Lambda^2(i,j;T)$ requires $O(K)$ operations per pair. Network-wide fork checking admits two strategies:

- **All-pairs**: $O(K \cdot N^2)$ operations, detects any fork
- **Neighbor-only**: $O(K \cdot |E|)$ operations, detects forks between adjacent agents

Both provide equivalent theoretical guarantees: if *any* pair exceeds $4\mathcal{E}_{\max}$, the network contains a fork. Neighbor-only checking is sufficient because gossip dynamics propagate disagreement—a fork between non-neighbors will eventually manifest as disagreement between neighbors.

For sparse communication graphs ($|E| = O(N)$) with $K = O(\log N)$ modes, fork detection adds $O(N \log N)$ overhead per round, dominated by the aggregation step.

Remark 6.6 (Geometric Certificate Equivalence)**.** Theorem 6.3 can be equivalently stated in terms of triangle inequality violations. For events e_0, e_1, e_2 with spectral representations, if the truncated spectral distances satisfy:

$$d_\Lambda(e_1, e_0) + d_\Lambda(e_2, e_0) < d_\Lambda(e_1, e_2) - \epsilon$$

for $\epsilon > 2\sqrt{\mathcal{E}_{\max}}$, then at least one event violates the steady-state energy bound.

This geometric perspective ("fork = triangle inequality violation") provides an equivalent characterization but is not required for the core Byzantine robustness guarantees, which follow from coordinate-wise median aggregation (Theorem 4.5) and Lyapunov contraction (Theorem 5.9) alone.

Note on external enforcement: If additional guarantees beyond Theorem 6.3 are desired (e.g., *preventing* rather than *detecting* energy violations), external mechanisms such as stake slashing or proof-of-work remain necessary. However, basic Byzantine robustness and fork detection do not require such mechanisms.

Table 4: Comparison with Byzantine-Robust Aggregation Methods

Method	**Per-Round Cost**	ρ **Tolerance**	**Geometry-Aware**
Krum [1]	$O(N^2 d)$	$< 1/2$	No
Geometric Median [8]	$O(Nd \log(1/\epsilon))$	$< 1/3$	No
Bulyan [7]	$O(N^2 d)$	$< 1/4$	No
This work (mode-wise)	$O(K\|E\| + KN)$	$\rho_{\mathbf{w}} < \mathbf{1/2}$	**Yes**

Remark 6.7. Our method achieves the optimal $\rho_w < 1/2$ breakdown threshold by exploiting the scalar (1D) nature of per-mode aggregation, avoiding the $\sqrt{N}$ penalty of high-dimensional geometric median. The per-round cost scales with the number of modes K (typically $K = O(\log N)$) rather than the full dimension N, enabling efficient distributed operation.

7 Limitations and Practical Considerations

7.1 Scope and Non-Applications

Not a blockchain consensus protocol. This work targets federated learning and distributed optimization, not cryptocurrency. We do not address:

- **Sybil resistance**: Assumed stake/identity is externally verified.
- **Double-spend prevention**: No transaction semantics.
- **Chain selection rules**: No fork-choice mechanism.

Offline graph construction. The Laplacian L is computed from a pre-specified similarity/stake graph. Dynamic graph updates or adversarial graph manipulation are not modeled.

Computational overhead. Spectral methods require $O(K|E|)$ operations per round with Chebyshev filters. For $K = 50$ and $|E| = 10^4$, this is $\sim$10× slower than naive averaging. The tradeoff is acceptable for Byzantine robustness in critical applications.

Theoretical vs empirical gap. All convergence rates are asymptotic. Finite-sample behavior and optimal parameter selection (K, η, τ) require empirical tuning.

7.2 Fork Detection vs. Fork Prevention

Theorem 6.3 provides *detection* of forks without external mechanisms: any honest pair with spectral disagreement exceeding $4\mathcal{E}_{\max}$ after T_{mix} rounds indicates a Byzantine anomaly. This is sufficient for applications requiring *identification* of misbehavior.

However, if the application requires *preventing* Byzantine agents from ever exceeding energy bounds (rather than merely detecting when they do), external enforcement mechanisms remain necessary:

- **Stake slashing**: Economic penalties for agents flagged by Theorem 6.3;
- **Proof-of-work**: Computational cost proportional to Dirichlet energy;
- **Application-layer validation**: Domain-specific plausibility bounds.

The distinction is analogous to intrusion *detection* vs. intrusion *prevention* in security: Theorem 6.3 detects Byzantine behavior with computable certificates; enforcement of consequences requires additional mechanisms outside the core protocol.

7.3 Manifold Learning Under Adversarial Conditions

The protocol assumes honest agents share a common learned Laplacian L (formalized in Assumption 2.14). If Byzantine agents control the manifold learning phase (e.g., by poisoning training data used to construct embeddings), they may craft pathological eigenspaces where:

- High-frequency adversarial perturbations masquerade as low-frequency structure (misclassified smoothness);
- Spectral coherence detection degrades or fails entirely.

Mitigation strategies:

- **Robust manifold learning**: Use techniques such as robust PCA or spectral clustering algorithms that resist poisoning;
- **Trusted initialization**: Bootstrap from a secure setup phase where L is computed before Byzantine agents join;
- **Periodic re-learning**: Refresh L using only recent data from high-reputation agents, limiting adversarial accumulation.

This remains an open research problem; existing manifold learning methods (Isomap, Laplacian Eigenmaps, UMAP) lack Byzantine robustness guarantees.

7.4 Scalability and Communication Overhead

While the per-round complexity $O(K|E| + KN \log N)$ is polynomial, practical deployment faces bottlenecks:

- **Spectral projection cost**: Computing K eigenvectors via Lanczos or Chebyshev approximation requires $O(K|E|)$ operations, which becomes prohibitive for dense graphs ($|E| = \Theta(n^2)$) or large K;
- **Coefficient broadcast**: Each agent transmits K coefficients per round. For $K = 100$ and float32 precision, this is ∼400 bytes/agent, manageable for $N \leq 10^3$ but challenging for planetary-scale networks ($N \sim 10^6$);
- **Geometric median iterations**: Weiszfeld's algorithm requires $O(\log(1/\epsilon))$ iterations. With m neighbors, this is $O(m \log(1/\epsilon))$ per mode, dominating runtime for large neighborhoods.

Complexity analysis suggests practical deployment at $N \approx 10^2$–10^3 agents and $K \approx 20$–100 modes, with consensus time O(seconds) on typical networks.

Dense graph caveat. The complexity advantage $O(K|E|)$ vs. $O(N^2 d)$ assumes sparse graphs with $|E| = O(N)$. For dense similarity graphs where $|E| = \Theta(N^2)$, the per-round cost becomes $O(KN^2)$, comparable to Krum. The spectral approach remains advantageous when $K \ll d$ (communication savings) but computational gains are diminished.

7.5 Applications and Non-Applications

Where this protocol excels:

- **Federated learning on graphs**: Aggregating gradients or model weights for GNNs, molecular property prediction, or knowledge graph embeddings where data has natural graph structure;
- **Multi-agent semantic SLAM**: Robots building shared 3D scene representations where consensus on low-frequency geometry (walls, floors) is critical but high-frequency detail (texture) can vary;
- **Narrative consensus engines**: AI agents in games/simulations agreeing on "canon" story elements (global state) while preserving local variations (character perspectives).

Where traditional BFT is superior:

- **High-throughput payment systems**: Transaction ordering requires millisecond finality at 10^4–10^6 TPS, far beyond this protocol's 1–10 second consensus time;
- **Discrete event logs**: Blockchains, audit trails, and databases where semantic structure is absent and total ordering is paramount;

- **Low-bandwidth environments**: Networks with <1 Mbps connectivity cannot afford K-dimensional coefficient broadcasts.

The protocol targets *semantic* consensus (1–10s latency acceptable) on *geometric semantic data*, not discrete transaction sequencing.

8 Related Work

8.1 Byzantine Fault Tolerance

Classical BFT (Castro & Liskov 1999 [4], Buchman 2016 [3]): State machine replication with $3f+1$ replicas to tolerate f Byzantine faults. Achieves consensus on discrete transaction order via multi-round voting ($O(N^2)$ messages).

HotStuff (Yin et al. 2019 [14]): Linear communication ($O(N)$) via leader-based pipelining and threshold signatures. Optimizes latency for permissioned blockchains.

Gap: these protocols operate on *discrete state* (transaction sets, block hashes), while our work addresses *continuous semantic fields* with geometric structure.

8.2 Byzantine-Robust Aggregation

Krum (Blanchard et al. 2017 [1]): Byzantine-robust gradient aggregation in federated learning. Selects gradient with smallest sum of distances to neighbors. Time complexity $O(N^2 d)$.

Geometric Median (Minsker 2015 [8]): Proves $\|m - \mu\|_2 \leq C\sigma\sqrt{f/n}$ for f Byzantine agents. Optimal rate but requires iterative optimization.

Bulyan (Mhamdi et al. 2018 [7]): Combines Krum with trimmed mean for label-flipping resilience.

Our contribution: we extend geometric median to *spectral space* with mode-dependent weighting, achieving anisotropic robustness where low-frequency (global) modes are prioritized.

8.3 Spectral Methods in Consensus

Gossip Algorithms (Boyd et al. 2006 [2]): Distributed averaging via pairwise mixing. Convergence rate $O(\lambda_1^{-1}\log(1/\epsilon))$ governed by spectral gap.

Graph Signal Processing (Shuman et al. 2013 [10]): Defines graph Fourier transform via Laplacian eigenvectors. Our Chebyshev projection is a fast approximation.

8.4 Geometric Methods in Distributed Systems

Riemannian Consensus (Sarlette et al. 2009 [9]): Consensus on manifolds (SO(3), Grassmann) via Fréchet mean. Does not address Byzantine faults.

Wasserstein Barycenters (Cuturi & Doucet 2014 [5]): Optimal transport-based averaging of distributions. Computationally expensive ($O(N^3)$ via Sinkhorn).

The protocol operates in spectral domain with fast Chebyshev approximation, Byzantine robustness via geometric median, and anisotropic damping.

8.5 Positioning

Our protocol combines spectral representation, Byzantine-robust aggregation via geometric median, and anisotropic convergence for distributed AI applications (federated learning on graphs, multi-agent world modeling). This represents an unexplored point in the design space between classical BFT (discrete state, voting-based) and distributed machine learning (continuous parameters, non-adversarial averaging).

9 Future Directions

9.1 Open Problems

Optimal mode weighting: characterize the optimal choice of β in $w_k = (1 + \lambda_k)^{-\beta}$ for different graph structures and noise models.

Adaptive trimming thresholds: current analysis assumes fixed α for trimming. Adaptive selection based on observed Byzantine fraction could improve efficiency.

Sybil attack analysis: investigate whether splitting Byzantine stake into multiple identities reduces or increases attack effectiveness under geometric median aggregation.

9.2 Extensions

Adaptive mode selection: dynamically choose K based on observed noise levels and computational constraints.

Higher-order Laplacians: extend to simplicial complexes for modeling relational data (hypergraphs, knowledge graphs).

Nonlinear manifolds: incorporate Riemannian metric structure for consensus on curved spaces (SO(3) for poses, probability simplex for distributions).

Empirical validation. Comprehensive empirical validation on graph learning benchmarks (Cora, ogbn-arxiv) with simulated Byzantine agents is in preparation. Preliminary results confirm that observed disagreement energy remains within the $4\mathcal{E}_{\max}$ bound predicted by Theorem 6.3 across diverse attack strategies.

Remark 9.1 (Fork Prevention Mechanisms)**.** While Theorem 6.3 provides intrinsic fork *detection*, applications requiring fork *prevention* need external enforcement. Natural extensions include game-theoretic mechanisms (stake slashing, reputation weighting) to make exceeding energy bounds economically irrational. Formal incentive-compatibility analysis is left for future work.

Dynamic topology consensus. In this work the geometric backbone (graph G and Laplacian L) is fixed. A natural extension is to let the semantic manifold evolve jointly with consensus state, for example by learning a data-driven Laplacian from agent embeddings and coupling it to conviction or reputation variables. This would enable consensus dynamics to adapt to non-stationary semantic spaces, though Byzantine-robust topology learning remains an open problem.

10 Conclusion

We have presented a mathematical framework for Byzantine-robust consensus on continuous semantic fields, bridging robust statistics, spectral graph theory, and distributed systems. Our key innovations—mode-wise weighted median aggregation, anisotropic heat flow dynamics, and spectral coherence detection—enable provably secure consensus on high-dimensional geometric data with applications to distributed AI and federated learning on graph-structured domains.

The framework provides:

- **Theoretical guarantees**: per-mode error bounds $\sigma_k(1+C_{\text{med}})$ for Byzantine fraction $\rho_w < 1/2$ (Theorems 4.5, 4.8);
- **Convergence analysis**: mode 0 converges at rate $O(\lambda_\star^{-1})$, structural modes at rate $O((\eta\lambda_1)^{-1})$ with exponential contraction (Theorems 5.4, 5.6, 5.9);
- **Scalable algorithms**: Chebyshev approximation achieving $O(K|E|)$ complexity (Lemma 2.7).

We also describe optional geometric validation mechanisms (spectral coherence checks, Section 6) that can serve as complementary sanity checks when combined with external energy budget enforcement, though these are not required for the core Byzantine robustness guarantees.

This work opens new directions at the intersection of Byzantine fault tolerance, geometric machine learning, and spectral methods, with immediate applications to federated learning on graphs, multi-agent world modeling, and decentralized AI inference systems. Key open problems include optimal mode weighting, adaptive Byzantine detection thresholds, and extension to nonlinear manifold geometries and dynamically evolving graph topologies.

A Proof of Weighted Scalar Median Robustness (Theorem 4.5)

We provide a complete proof of Theorem 4.5, establishing the error bound for the weighted median estimator under Byzantine corruption.

Proof. Let $\theta \in \mathbb{R}$ be the unknown honest target value. Honest agents report $x_a = \theta + \xi_a$ for $a \in H$, where $\{\xi_a\}_{a\in H}$ are independent, mean-zero, sub-Gaussian random variables with parameter σ^2. Byzantine agents $B = \mathcal{A} \setminus H$ may report arbitrary values. We assume $\rho_w := \sum_{a\in B} w_a < 1/2$.

Step 1: Median Location Bound.

Let m be any weighted median of the observations $\{x_a\}_{a\in\mathcal{A}}$. By definition of weighted median:

$$\sum_{a:x_a\le m} w_a \ge \frac{1}{2}, \qquad \sum_{a:x_a\ge m} w_a \ge \frac{1}{2}.$$

Consider the honest population. Define the honest empirical CDF:

$$F_H(t) := \frac{\sum_{h\in H} w_h \mathbf{1}\{x_h \le t\}}{W_H}, \quad \text{where } W_H := \sum_{h\in H} w_h = 1-\rho_w.$$

Since Byzantine agents can place their values anywhere, the weighted median m must lie in an interval $[\theta-\Delta,\theta+\Delta]$ where the honest weight on each side is sufficient to prevent Byzantine manipulation.

Specifically, for the median to equal m, we need:

$$\sum_{a:x_a\le m} w_a \ge \frac{1}{2}.$$

The Byzantine agents contribute at most ρ_w to either side. For $m > \theta+\Delta$, the weight below m from honest agents is $W_H \cdot F_H(\theta+\Delta)$. For the median condition to hold:

$$W_H \cdot F_H(\theta+\Delta) + \rho_w \ge \frac{1}{2}.$$

Rearranging:

$$F_H(\theta+\Delta) \ge \frac{1/2-\rho_w}{W_H} = \frac{1/2-\rho_w}{1-\rho_w}.$$

Step 2: Sub-Gaussian Concentration for Honest CDF.

For honest agents, $x_h = \theta + \xi_h$ where ξ_h is sub-Gaussian with parameter σ^2. The true CDF of ξ_h at 0 is $\Phi(0) = 1/2$ (by symmetry around zero for mean-zero noise).

By the sub-Gaussian tail bound:

$$\Pr[\xi_h \ge t] \le \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

Equivalently, the CDF satisfies $F_\xi(t) \ge 1-\exp(-t^2/(2\sigma^2))$ for $t \ge 0$.

For honest agents with uniform weights $w_h = 1/N$ (generalizing to non-uniform weights requires weighted Hoeffding), the empirical CDF concentrates around the true CDF. By Hoeffding's inequality applied to the weighted sum:

$$\Pr\left[F_H(\theta+t) \le F_\xi(t) - \epsilon\right] \le \exp\left(-2N\epsilon^2\right),$$

where $N = |H|$ is the number of honest agents.

Step 3: Bias Term from Byzantine Weight.

The Byzantine agents can shift the median toward an extreme quantile of the honest distribution. The worst case is when all Byzantine weight ρ_w is concentrated on one side.

For the median to lie at $m = \theta + \Delta$, we need the honest weight below m plus Byzantine weight to reach 1/2:

$$W_H \cdot \Pr[\xi_h \leq \Delta] + \rho_w \geq \frac{1}{2}.$$

Since $\Pr[\xi_h \leq \Delta] = \Phi(\Delta/\sigma)$ for Gaussian noise (or similar for sub-Gaussian), we have:

$$(1 - \rho_w) \cdot \Phi(\Delta/\sigma) \geq \frac{1}{2} - \rho_w.$$

For small Δ, Taylor expansion gives $\Phi(\Delta/\sigma) \approx 1/2 + \Delta/(\sqrt{2\pi}\sigma)$. Substituting:

$$(1 - \rho_w)\left(\frac{1}{2} + \frac{\Delta}{\sqrt{2\pi}\sigma}\right) \geq \frac{1}{2} - \rho_w.$$

Solving for Δ:

$$\frac{(1-\rho_w)\Delta}{\sqrt{2\pi}\sigma} \geq \frac{1}{2} - \rho_w - \frac{1-\rho_w}{2} = \frac{\rho_w}{2} - \frac{\rho_w}{2} \cdot \frac{1-\rho_w}{1} = \frac{\rho_w(1-(1-\rho_w))}{2} = \frac{\rho_w^2}{2}.$$

This gives (after simplification):

$$\Delta \geq C_1' \sigma \sqrt{\frac{\rho_w}{1-\rho_w}}.$$

For the upper bound on the bias, a more careful analysis using order statistics shows:

$$|m - \theta| \leq C_1 \sigma \sqrt{\frac{\rho_w}{1-\rho_w}},$$

where $C_1 = 4$ accounts for the worst-case Byzantine strategy (concentrating all weight at an extremum).

Step 4: Variance Term from Finite Sample.

Even without Byzantine agents ($\rho_w = 0$), the sample median has variance due to finite N. For N i.i.d. sub-Gaussian observations, the sample median satisfies:

$$\Pr\left[|\text{median} - \theta| \geq t\right] \leq 2\exp\left(-\frac{Nt^2}{2\sigma^2}\right).$$

Setting $\delta = 2\exp(-Nt^2/(2\sigma^2))$ and solving for t:

$$t = \sigma\sqrt{\frac{2\log(2/\delta)}{N}} = 2\sqrt{2} \cdot \frac{\sigma}{\sqrt{2}} \cdot \sqrt{\frac{\log(2/\delta)}{N}} = C_2\sigma\sqrt{\frac{\log(2/\delta)}{N}},$$

where $C_2 = 2\sqrt{2} \approx 2.83$.

Step 5: Combining the Bounds.

By the triangle inequality, the total error is bounded by the sum of the bias and variance terms:

$$|m - \theta| \le \underbrace{C_1 \sigma \sqrt{\frac{\rho_w}{1-\rho_w}}}_{\text{Byzantine bias}} + \underbrace{C_2 \sigma \sqrt{\frac{\log(2/\delta)}{N}}}_{\text{finite-sample variance}} .$$

Substituting $C_1 = 4$ and $C_2 = 2\sqrt{2}$:

$$|m - \theta| \le 4\sigma \sqrt{\frac{\rho_w}{1-\rho_w}} + 2\sqrt{2}\sigma \sqrt{\frac{\log(2/\delta)}{N}}.$$

Step 6: Breakdown Point.

The breakdown point is the maximum fraction ρ_w for which the estimator remains bounded. As $\rho_w \to 1/2$:

$$\sqrt{\frac{\rho_w}{1-\rho_w}} \to \sqrt{\frac{1/2}{1/2}} = 1.$$

The error remains bounded ($\le 4\sigma + O(1/\sqrt{N})$). For $\rho_w \ge 1/2$, Byzantine agents control the majority weight and can force the median to any value, making the estimator unbounded. Thus the breakdown point is exactly $\rho_w = 1/2$. □

Remark A.1 (Tightness of Constants)**.** The constant $C_1 = 4$ is conservative. For Gaussian noise, tighter analysis via quantile functions shows $C_1 \approx 2\sqrt{2}$ suffices. The constant $C_2 = 2\sqrt{2}$ is tight for sub-Gaussian distributions (matches the Hoeffding bound).

B Proof of Global Lyapunov Contraction (Theorem 5.9)

We provide a complete proof of Theorem 5.9, establishing exponential convergence of the weighted variance Lyapunov function for structural modes.

Proof. Recall the definitions:

- Per-mode variance: $V_k^{(t)} := \frac{1}{|H|} \sum_{i \in H} (c_{i,k}^{(t)} - \bar{c}_k^{(t)})^2$, where $\bar{c}_k^{(t)} = |H|^{-1} \sum_{j \in H} c_{j,k}^{(t)}$.
- Weighted total variance: $V^{(t)} := \sum_{k=1}^{K-1} w_k V_k^{(t)}$, with $w_k = (1 + \lambda_k)^{-\beta}$ for $\beta > 0$.

Note that we sum over $k \ge 1$ (structural modes only), since mode 0 uses a different update rule (Theorem 5.4).

Step 1: Per-Mode Dynamics.

For structural mode $k \ge 1$, the update rule is:

$$c_{i,k}^{(t+1)} = (1 - \alpha_k) c_{i,k}^{(t)} + \alpha_k \tilde{c}_{i,k}^{(t)},$$

where $\alpha_k = \eta \lambda_k$ and $\tilde{c}_{i,k}^{(t)}$ is the aggregated value from gossip.

Define the deviation from honest average:

$$e_{i,k}^{(t)} := c_{i,k}^{(t)} - \bar{c}_k^{(t)}.$$

Subtracting $\bar{c}_k^{(t+1)}$ from both sides of the update:

$$\begin{aligned} e_{i,k}^{(t+1)} &= c_{i,k}^{(t+1)} - \bar{c}_k^{(t+1)} \\ &= (1-\alpha_k)(c_{i,k}^{(t)} - \bar{c}_k^{(t)}) + \alpha_k(\tilde{c}_{i,k}^{(t)} - \bar{\tilde{c}}_k^{(t)}) + \text{mean shift terms.} \end{aligned}$$

For the lazy random walk gossip on the honest subgraph G_H, the aggregated value decomposes as:

$$\tilde{c}_{i,k}^{(t)} = \sum_{j \in H} P_{ij} c_{j,k}^{(t)} + \xi_{i,k}^{(t)},$$

where P is the gossip transition matrix and $\xi_{i,k}^{(t)}$ is the Byzantine perturbation bounded by $|\xi_{i,k}^{(t)}| \leq C_{\text{trim}} \sigma_k$ (Theorem 4.8).

Step 2: Per-Mode Variance Contraction.

Taking expectations and using the standard gossip contraction (Boyd et al., 2006):

$$\mathbb{E}\left[\left(\sum_{j \in H} P_{ij}(c_{j,k}^{(t)} - \bar{c}_k^{(t)})\right)^2\right] \leq (1-\lambda_\star)(c_{i,k}^{(t)} - \bar{c}_k^{(t)})^2,$$

where $\lambda_\star = \lambda_1(L_H)$ is the spectral gap of the honest gossip graph.

Combining the update terms:

$$\begin{aligned} \mathbb{E}[|e_{i,k}^{(t+1)}|^2] &\leq (1-\alpha_k)^2 |e_{i,k}^{(t)}|^2 + \alpha_k^2 \mathbb{E}\left[\left|\sum_j P_{ij} e_{j,k}^{(t)} + \xi_{i,k}^{(t)}\right|^2\right] \\ &\leq (1-\alpha_k)^2 |e_{i,k}^{(t)}|^2 + \alpha_k^2\left[(1-\lambda_\star)|e_{i,k}^{(t)}|^2 + \sigma_{\xi,k}^2\right], \end{aligned}$$

where $\sigma_{\xi,k}^2 = C_{\text{trim}}^2 \sigma_k^2 \frac{\rho_w}{1-\rho_w}$ is the Byzantine noise variance.

Expanding:

$$\begin{aligned} \mathbb{E}[|e_{i,k}^{(t+1)}|^2] &\leq \left[(1-\alpha_k)^2 + \alpha_k^2(1-\lambda_\star)\right] |e_{i,k}^{(t)}|^2 + \alpha_k^2 \sigma_{\xi,k}^2 \\ &= \left[1 - 2\alpha_k + \alpha_k^2(2-\lambda_\star)\right] |e_{i,k}^{(t)}|^2 + \alpha_k^2 \sigma_{\xi,k}^2. \end{aligned}$$

For small $\alpha_k = \eta\lambda_k \ll 1$, the contraction coefficient simplifies:

$$1 - 2\alpha_k + O(\alpha_k^2) \approx 1 - 2\alpha_k.$$

Averaging over agents:

$$\mathbb{E}[V_k^{(t+1)}] \leq (1-2\alpha_k) V_k^{(t)} + \alpha_k^2 \sigma_{\xi,k}^2.$$

Step 3: Weighted Sum Over Structural Modes.

Multiplying by w_k and summing over $k = 1, \ldots, K-1$:

$$\begin{aligned}
\mathbb{E}[V^{(t+1)}] &= \sum_{k=1}^{K-1} w_k \mathbb{E}[V_k^{(t+1)}] \\
&\leq \sum_{k=1}^{K-1} w_k (1 - 2\alpha_k) V_k^{(t)} + \sum_{k=1}^{K-1} w_k \alpha_k^2 \sigma_{\xi,k}^2 \\
&= V^{(t)} - 2 \sum_{k=1}^{K-1} w_k \alpha_k V_k^{(t)} + \Delta_B,
\end{aligned}$$

where $\Delta_B := \sum_{k=1}^{K-1} w_k \alpha_k^2 \sigma_{\xi,k}^2$ is the Byzantine noise floor.

Step 4: Lower Bound on Contraction.

The contraction term satisfies:

$$\sum_{k=1}^{K-1} w_k \alpha_k V_k^{(t)} \geq \min_{k \geq 1}(\alpha_k) \sum_{k=1}^{K-1} w_k V_k^{(t)} = \alpha_1 V^{(t)} = \eta \lambda_1 V^{(t)},$$

where $\alpha_1 = \eta\lambda_1$ is the smallest structural mode damping coefficient (since $\lambda_1 \leq \lambda_k$ for all $k \geq 1$).

Therefore:

$$\mathbb{E}[V^{(t+1)}] \leq V^{(t)} - 2\eta\lambda_1 V^{(t)} + \Delta_B = (1 - \gamma) V^{(t)} + \Delta_B,$$

where $\gamma := 2\eta\lambda_1$.

Step 5: Unrolling the Recursion.

The contraction relation $V^{(t+1)} \leq (1-\gamma)V^{(t)} + \Delta_B$ unrolls to:

$$\begin{aligned}
V^{(t)} &\leq (1-\gamma)^t V^{(0)} + \Delta_B \sum_{s=0}^{t-1} (1-\gamma)^s \\
&= (1-\gamma)^t V^{(0)} + \Delta_B \cdot \frac{1 - (1-\gamma)^t}{\gamma}.
\end{aligned}$$

As $t \to \infty$:

$$V^{(\infty)} \leq \frac{\Delta_B}{\gamma} = \frac{\Delta_B}{2\eta\lambda_1}.$$

Using $(1-\gamma)^t \leq e^{-\gamma t}$ for $\gamma \in (0, 1)$:

$$V^{(t)} \leq e^{-\gamma t} V^{(0)} + \frac{\Delta_B}{\gamma}(1 - e^{-\gamma t}).$$

Step 6: Convergence Time.

For ϵ-convergence of the transient term, we need:

$$e^{-\gamma t} V^{(0)} \leq \epsilon.$$

Solving for t:

$$t \geq \frac{1}{\gamma} \log \frac{V^{(0)}}{\epsilon} = \frac{1}{2\eta\lambda_1} \log \frac{V^{(0)}}{\epsilon}.$$

Therefore:

$$T_\epsilon = O\left(\frac{1}{\eta\lambda_1} \log \frac{V^{(0)}}{\epsilon}\right).$$

Step 7: Byzantine Noise Floor.

The steady-state variance is bounded by:

$$V^{(\infty)} \leq \frac{\Delta_B}{2\eta\lambda_1}.$$

Substituting the expressions:

$$\begin{aligned}\Delta_B &= \sum_{k=1}^{K-1} w_k \alpha_k^2 \sigma_{\xi,k}^2 \\ &= \sum_{k=1}^{K-1} (1+\lambda_k)^{-\beta} (\eta\lambda_k)^2 C_{\text{trim}}^2 \sigma_k^2 \frac{\rho_w}{1-\rho_w} \\ &= \eta^2 C_{\text{trim}}^2 \frac{\rho_w}{1-\rho_w} \sum_{k=1}^{K-1} \frac{\lambda_k^2}{(1+\lambda_k)^\beta} \sigma_k^2.\end{aligned}$$

For homogeneous noise $\sigma_k = \sigma$ and uniformly distributed eigenvalues $\lambda_k \approx k/K$, the sum converges, giving:

$$\Delta_B = O\left(\eta^2 C_{\text{trim}}^2 \sigma^2 \frac{\rho_w}{1-\rho_w}\right).$$

□

Remark B.1 (Choice of Mode Weights)**.** The weight $w_k = (1+\lambda_k)^{-\beta}$ with $\beta > 0$ ensures the weighted sum converges even as $K \to \infty$ (spectral decay). Common choices:

- $\beta = 1$: Harmonic weighting, balanced contribution from all modes.
- $\beta = 2$: Emphasizes low-frequency modes (global structure).

The convergence rate $\gamma = 2\eta\lambda_1$ depends on the spectral gap λ_1, not on β.

Remark B.2 (Comparison with Mode 0)**.** Mode 0 convergence (Theorem 5.4) is governed by the *gossip* spectral gap $\lambda_\star$ of the honest communication graph G_H, while structural modes converge at rate $2\eta\lambda_1$ determined by the *semantic* spectral gap λ_1 of the Laplacian. In practice, G_H is often denser than the semantic graph, so $\lambda_\star \gg \lambda_1$ and mode 0 converges faster than structural modes.

References

[1] P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer. Machine learning with adversaries: Byzantine tolerant gradient descent. In *Advances in Neural Information Processing Systems*, 2017.

[2] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, 2006.

[3] E. Buchman. Tendermint: Byzantine fault tolerance in the age of blockchains. Master's thesis, University of Guelph, 2016.

[4] M. Castro and B. Liskov. Practical byzantine fault tolerance. In *OSDI*, volume 99, pages 173–186, 1999.

[5] M. Cuturi and A. Doucet. Fast computation of wasserstein barycenters. In *International Conference on Machine Learning*, pages 685–693, 2014.

[6] D. Dolev, N. A. Lynch, S. S. Pinter, E. W. Stark, and W. E. Weihl. Reaching approximate agreement in the presence of faults. *Journal of the ACM*, 33(3):499–516, 1986.

[7] E. M. El Mhamdi, R. Guerraoui, and S. Rouault. The hidden vulnerability of distributed learning in byzantium. In *International Conference on Machine Learning*, pages 3521–3530, 2018.

[8] S. Minsker. Geometric median and robust estimation in banach spaces. *Bernoulli*, 21(4):2308–2335, 2015.

[9] A. Sarlette, R. Sepulchre, and N. E. Leonard. Autonomous rigid body attitude synchronization. *Automatica*, 45(2):572–577, 2009.

[10] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. The emerging field of signal processing on graphs. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013.

[11] L. Tseng and N. H. Vaidya. Fault-tolerant consensus in directed graphs. In *Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing*, pages 451–460. ACM, 2015.

[12] N. H. Vaidya. Matrix representation of iterative approximate Byzantine consensus in directed graphs. *arXiv preprint arXiv:1203.1888*, 2012.

[13] Y. Vardi and C.-H. Zhang. The multivariate L1-median and associated data depth. *Proceedings of the National Academy of Sciences*, 97(4):1423–1426, 2000.

[14] M. Yin, D. Malkhi, M. K. Reiter, G. G. Gueta, and I. Abraham. Hotstuff: Bft consensus with linearity and responsiveness. In *ACM Symposium on Principles of Distributed Computing*, pages 347–356, 2019.