ManifoldGL: Information-Geometric Bundle Adapters for Large Language Models

Jesús Vilela Jato

Department of Advanced Agentic Coding

Abstract

*The Euclidean assumption inherent in dot-product attention mechanisms fundamentally limits the ability of Large Language Models (LLMs) to capture the hierarchical and polysemous nature of natural language. We propose the **Information-Geometric Bundle (IGBundle)**, a novel adapter architecture that reinterprets neural activations as local sections of a fiber bundle over a smooth base manifold. By integrating Information Geometry (IG) and Sheaf Theory, we construct a loss function based on the Jensen-Shannon divergence of overlapping patch distributions. We demonstrate that this architecture, when injected into a 7B parameter model (Qwen2.5), learns to maintain a consistent non-Euclidean geometry, effectively capturing semantic curvature (σ) that correlates with linguistic ambiguity.*

1. Introduction

Contemporary Transformers operate on the premise that semantic meaning can be encoded as a vector in a flat space $\mathbb{R}^d$. However, linguistic phenomena such as polysemy, entailment, and negation suggest a geometry closer to hyperbolic or mixed-curvature manifolds. We argue that a word embedding is not a point, but a fiber F_x over a base structural manifold M. The "context" selects a specific point $p \in F_x$, and the attention mechanism approximates parallel transport along a geodesic connecting token locations.

2. Mathematical Preliminaries

2.1. Fiber Bundles

A fiber bundle is a tuple (E, M, π, F), where E is the total space, M is the base space, and for every $p \in M$, the fiber $\pi^{-1}(p)$ is homeomorphic to F. In our framework:

- M: The syntactic or structural manifold of the sentence.
- F: The semantic space of possible meanings (polysemy).
- E: The bundle of all contextualized meanings.

2.2. Information Geometry

We treat the fibers not as vector spaces, but as statistical manifolds equipped with the Fisher Information Metric. The distance between two semantic states is defined by the Kullback-Leibler (KL) divergence between their probability distributions:

$$D_{KL}(P \,||\, Q) = \sum P(x) \log \frac{P(x)}{Q(x)}$$

3. The IGBundle Adapter

3.1. Bottleneck Architecture

To manage computational complexity on consumer hardware (8GB VRAM), we implement a bottleneck architecture. The hidden state $h \in \mathbb{R}^H$ is projected into a lower-dimensional tangent space $T_p M$ of dimension $D_{bot}=256$:

$$z_{bot} = W_{in} h + b_{in}, \quad z_{bot} \in \mathbb{R}^{256}$$

3.2. Manifold Gaussian Mixture

Within the bundle space, we model the semantic state as a Mixture of Gaussians. The adapter predicts parameters for P components:

- Means μ: Location in the bundle.
- Precision $\lambda = e^{-2\sigma}$: Inverse variance, representing semantic certainty.
- Weights w: Importance of each component.

3.3. Sheaf Consistency Loss

A key innovation is the **Sheaf Loss**, effectively a Laplacian regularizer that enforces consistency between overlapping "semantic patches" (covering sets of the manifold). We utilize the Jensen-Shannon Divergence as a symmetric metric for gluing conditions:

```
\mathcal{L}_{Sheaf} = \sum_{r,s} \Omega_{rs} \cdot JS(\mathcal{P}_r ||
                              \mathcal{P}_s)
```

where Ω_{rs} is the overlap strength between patches r and s. This forces the model to learn a globally consistent topology from local charts.

4. Experiments & Results

We integrated the IGBundle adapter into **Qwen/Qwen2.5-7B**.

Training Configuration:
- Dataset: Alpaca (Cleaned)
- Optimization: Low-Rank Adaptation (LoRA) + Bundle Injection
- Hardware: NVIDIA RTX 3060 Ti (8GB VRAM)
- Precision: 4-bit Normal Float (NF4)

4.1. Curvature Analysis

We monitored the **Average Internal Sigma (σ)** across layers. In a flat Euclidean model, we would expect $\sigma \to 0$ (collapse to point masses).

Observed Result: $\sigma \approx 2.2 - 2.3$

This sustained non-zero curvature indicates the model is actively utilizing the fiber spread to represent ambiguity, confirming the "Proof of Life" of the bundle architecture.

5. Conclusion

We successfully formulated, implemented, and validated an Information-Geometric Bundle Adapter for LLMs. By explicitly modeling the statistical manifold structure of semantic spaces, we offer a path toward more robust, interpretable, and mathematically rigorous language models. Future work will focus on scaling the patch count and integrating hyperbolic attention kernels.

References

[1] Amari, S. (2016). Information Geometry and Its Applications. Springer.
[2] Ehresmann, C. (1950). Les connexions infinitésimales dans un espace fibré différentiable.
[3] Bronstein, M. et al. (2021). Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges.