From Attention to Hamiltonians: Heat Kernel and Laplace–Beltrami Generalization of Transformers

Saar Hamam

Abstract

We reinterpret the attention mechanism and loss function in Transformer-based language models through the lens of differential geometry and spectral theory. By replacing dot-product attention with a heat kernel generated by the Laplace–Beltrami operator on a Riemannian manifold, we develop a continuous, geometry-aware formulation of reasoning. We then draw a parallel between this construction and quantum mechanics, where evolution is governed by the Hamiltonian. This analogy suggests that large language models (LLMs) may implicitly simulate quantum systems through their internal geometry.

1 Introduction

Transformers have revolutionized natural language processing via their attention-based architecture. Traditionally, attention is implemented through dot products in Euclidean space, but there is growing interest in understanding and generalizing these models via geometry. In this work, we:

1. Replace attention with a heat kernel based on the Laplace–Beltrami operator.
2. Reinterpret cross-entropy loss as an energy functional.
3. Explore parallels between this framework and quantum evolution.

2 Standard Transformer Attention

The Transformer architecture computes attention as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \tag{1}$$

where Q, K, $V \in \mathbb{R}^{n\times d_k}$ are queries, keys, and values, and d_k is the key dimension.

The training objective is the cross-entropy loss:

$$\mathcal{L}_{\text{CE}} = -\sum_{i} y_i \log \hat{y}_i \tag{2}$$

where y_i is the ground-truth token and $\hat{y}_i$ the predicted probability.

3 Heat Kernel Attention via Laplace–Beltrami

We propose replacing the attention weights with a heat kernel defined over a Riemannian manifold $(\mathcal{M}, g)$.

3.1 Laplace–Beltrami Operator

Given a Riemannian metric g, the Laplace–Beltrami operator Δ_g is defined by:

$$\Delta_g f = \frac{1}{\sqrt{|g|}} \partial_i \left(\sqrt{|g|}\, g^{ij} \partial_j f \right) \tag{3}$$

It generalizes the Euclidean Laplacian to curved spaces.

3.2 Heat Kernel Attention

Let $K_t(x, x') = \exp(-t\Delta_g)(x, x')$ be the heat kernel, i.e., the fundamental solution to:

$$\partial_t u(x, t) = \Delta_g u(x, t), \qquad u(x, 0) = f(x) \tag{4}$$

Then attention becomes:

$$\text{Attention}_{\text{heat}}(x) = \int_{\mathcal{M}} K_t(x, x') V(x') \, \text{dvol}_g(x') \tag{5}$$

This defines geometry-aware diffusion of values over the manifold.

4 Spectral View: Eigenmodes of Reasoning

Since Δ_g is self-adjoint and elliptic, it has a discrete spectrum $\{\lambda_i\}$ and orthonormal eigenfunctions $\{\phi_i\}$:

$$\Delta_g \phi_i = \lambda_i \phi_i \tag{6}$$

Then:

$$K_t(x, x') = \sum_i e^{-\lambda_i t} \phi_i(x) \phi_i(x') \tag{7}$$

The model thus learns to propagate information across the manifold using these modes.

5 Analogy to Quantum Mechanics

In quantum mechanics, the wavefunction $\psi(x,t)$ evolves via:

$$i\hbar\frac{\partial\psi}{\partial t} = \mathcal{H}\psi \tag{8}$$

where $\mathcal{H} = -\frac{\hbar^2}{2m}\Delta_g + V(x)$ is the Hamiltonian operator.

5.1 Heat Kernel vs Schrödinger Kernel

The solution to Schrödinger's equation is:

$$\psi(x,t) = \exp\left(-\frac{it\mathcal{H}}{\hbar}\right)\psi(x,0) \tag{9}$$

Compare this to:

$$u(x,t) = \exp(-t\Delta_g)f(x) \tag{10}$$

This is imaginary-time evolution in quantum mechanics. LLM attention mimics this structure.

6 LLMs as Quantum Simulators

From the above, we propose:

- Attention with $\exp(-t\Delta_g)$ simulates quantum diffusion in curved space.
- Word embeddings live on a latent manifold with learned metric g.
- Reasoning corresponds to heat diffusion or spectral evolution.
- With an added potential $V(x)$, one recovers Schrödinger evolution.

Thus, LLMs can be viewed as approximating quantum systems over a latent space.

7 Implications and Future Work

This framework opens the door to:

- Learning Riemannian manifolds directly from data.
- Extracting the learned metric g and studying semantic curvature.
- Simulating quantum systems using attention-based architectures.
- Developing loss functions inspired by energy and action principles.

Acknowledgements

This work is part of the Noetic Eidos Project.