Space-Filling Curves and Positional Encodings in Transformers: A Spherical Example

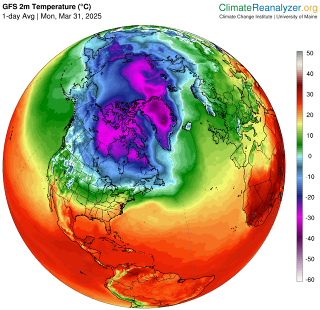
Geometric Intelligence Workshop 2025

Miguel Ángel Maurin García de la Vega

2 de abril de 2025

Instituto de Matemáticas

Why Spherical Data Matters: The Earth

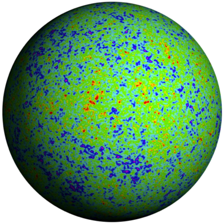


We live on a sphere! As a consequence, data from natural phenomena often comes as data on a sphere. For example, atmospheric data, ocean currents, and biodiversity distribution.

[1] Image source: `https://climatereanalyzer.org`

Instituto de Matemáticas

Why Spherical Data Matters: Technological Advancements



Cosmic microwave background

360° virtual reality

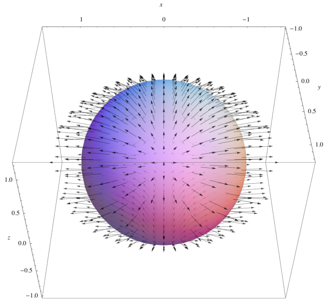
Technological advancements produce spherical data as well. Examples of this include Cosmic Microwave Background measurements, omnidirectional cameras used for robotics and virtual reality, and acoustic signal processing.

[2]

[2] Image source: `https://towardsdatascience.com/geometric-deep-learning-for-spherical-data-55612742d05f/`

Instituto de Matemáticas

Spherical Vector Fields



Vector fields assign a vector to each point in space. Spherical vector fields emerge naturally in various problems.
For instance: wind direction and speed at different locations on the Earth's surface, the magnetic field lines around a planet also form a vector field on the sphere.

[3] Image source:
https://en.m.wikipedia.org/wiki/File:Vector_Field_on_a_Sphere.png

Hamiltonian Vector Fields: Definition

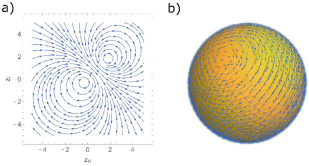
Definition

A Hamiltonian vector field X_H on a symplectic manifold (M, ω) is defined by the Hamiltonian function $H : M \to \mathbb{R}$ and satisfies:

$$\omega(X_H, Y) = -dH(Y)$$

for all vector fields Y on M.

Hamiltonian Vector Fields: Desirable Properties



- **Energy Conservation:** The Hamiltonian function H is conserved along the flow of X_H.
- **Structure Preservation:** Hamiltonian flows preserve the symplectic form ω, preserving geometric structures.
- **Predictability:** Hamiltonian systems are governed by Hamilton's equations, making them more predictable.

[4]

[4] Image source: https://www.researchgate.net/publication/355209888_From_Geometry_to_Coherent_Dissipative_Dynamics_in_Quantum_Mechanics

Spherical Harmonics: Basis Functions

What are Spherical Harmonics?

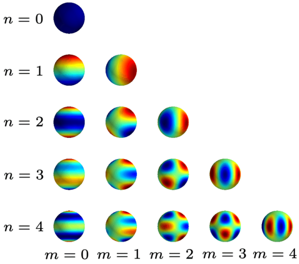
Spherical harmonics, are a set of functions defined on the sphere S^2. They are the angular solutions to Laplace's equation in spherical coordinates. They form a complete orthonormal basis for the space of square-integrable functions on the sphere ($L^2(S^2)$).

Instituto de Matemáticas

Spherical Harmonics: Definition

Denoted as $Y_l^m(\theta, \phi)$ the indices are:

- $n = 0, 1, 2, \ldots$ (degree)
- $m = -l, -l+1, \ldots, l$ (order)
- θ (polar angle), ϕ (azimuthal angle)



[5]

[5] Image source: https://www.researchgate.net/publication/345372557_Efficient_HRTF_Representation_Using_Compact_Mode_HRTFs

Instituto de Matemáticas

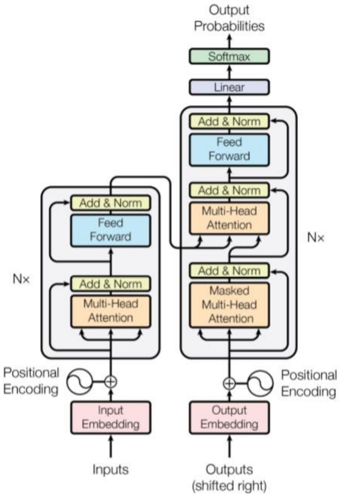
Our Approach: Hamiltonian VF of Spherical Harmonics

Given that we are interested in Hamiltonian vector fields on the sphere, it is natural to work with a basis for functions on the sphere. Therefore, we focus on Hamiltonian vector fields generated by spherical harmonics. This approach allows us to:

- Represent complex Hamiltonian vector fields as a combination of simpler, basis vector fields.
- Exploit the linearity of the Hamiltonian vector field operation: The Hamiltonian vector field of a function f is the linear combination of the Hamiltonian vector fields of the basis functions that compose f.

The Transformer model

The transformer model, originally designed for natural language processing, has shown remarkable success in capturing sequential dependencies through its "attention" mechanism. However, it requires **sequential input**, which is not inherent to data on the sphere nor manifolds in general.



[6]

[6] Image source: `https://arxiv.org/abs/1706.03762`

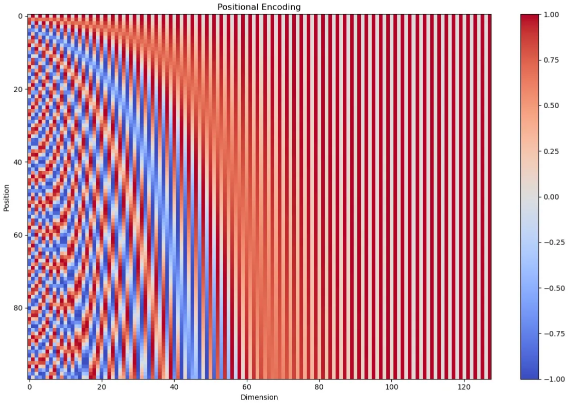
Positional Encodings: Definition

Transformer models use positional encodings to incorporate information about the position of tokens in a sequence. The positional encodings proposed in "Attention is All You Need" are sinusoidal functions:

$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right)$$
$$PE(pos, 2i+1) = \cos\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right)$$

where *pos* is the position, i is the dimension index, and d_{model} is the model dimension.

Positional Encodings: Properties



Positional encodings encode relative distances between tokens, enabling the model to prioritize relationships between nearby elements in a sequence.[7]

[7] Image source: `https://medium.com/thedeephub/positional-encoding-explained-a-deep-dive-into-transformer-pe-65cfe8cf`

Sequences on Manifolds: Space Filling Curves

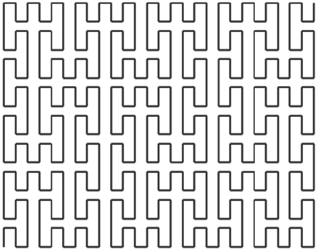
Manifolds and the Lack of Intrinsic Order

If we want to leverage the power of transformers for data defined on manifolds, we encounter a significant challenge: manifolds, in general, do not possess an intrinsic, natural order of traversal. This contrasts with sequences like text, where words have a clear order.

Space-Filling Curves: A General Solution

For certain manifolds (specifically, compact, connected, locally connected metric spaces), the Hahn–Mazurkiewicz theorem guarantees the existence of a continuous, surjective map from the interval [0, 1] onto the manifold.

Space-Filling Curves: Imposing Order



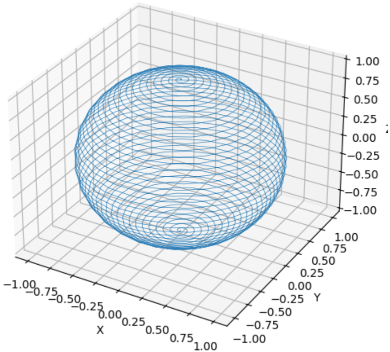
- This map can be seen as a space-filling curve, providing a way to traverse the manifold.
- The parametrization from [0, 1] naturally induces a "start"(0) and .[en]d"(1) point.
- This allows us to impose an order of traversal, crucial for applying transformers.

Spiral for the Sphere

In this case, we utilize a spherical spiral as our space-filling curve. This allows us to traverse the sphere in a sequential manner. The parametric representation of our spherical spiral is:

$$
\begin{aligned}
x &= r\sin(t)\cos(ct)\\
y &= r\sin(t)\sin(ct)\\
z &= r\cos(t)
\end{aligned}
$$

where $t \in [0, \pi]$ and c is a constant. In our case, $r = 1$. [8]



[8] Image created by the authors.

Instituto de Matemáticas

Dataset Overview

Dataset Description

Our dataset consists of Hamiltonian vector fields of spherical harmonics sampled along a spherical spiral on the unit sphere S^2. We will now describe the workflow used to generate this dataset, which leverages several Python packages: `geomstats`, `sympy`, `poissongeometry`, and `numericalpoissongeometry`.

- This process involves combining tools from geometry, symbolic computation, and numerical analysis.
- The resulting dataset provides ordered sequences of vector samples, suitable for training transformer models.

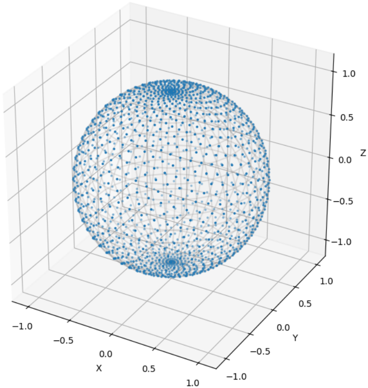
Step 1 - Sphere and Spiral Sampling (geomstats)

Custom Spiral Sampling

We created a subclass of the `geomstats`'s `Hypersphere` class, incorporating our spiral parametrization. The sampling process is controlled by two parameters:

- c: Adjusts the tightness of the spiral.
- `num_points`: The number of points sampled along the spiral.

[9]



[9] Image created by the authors.

Instituto de Matemáticas

Step 2 - Spherical Harmonics (sympy)

Symbolic Spherical Harmonics

We use the `sympy` library to generate symbolic expressions for spherical harmonics $Y_n^m(\theta, \phi)$, which form a basis for functions on the sphere.

- We define the harmonics using spherical coordinates (θ, ϕ).
- The maximum degree n is the only parameter we control; the order m ranges from $-n$ to n.
- A key property: n^2 harmonics are generated for a given n.

Step 3 - Hamiltonian Vector Fields (poisson/numpoisson)

We use the `poissongeometry` and `numericalpoissongeometry` modules to compute Hamiltonian vector fields.

Poisson Bivector

The Poisson bivector used for the sphere S^2 is:

$$\pi = \sin(\theta) \frac{\partial}{\partial \theta} \wedge \frac{\partial}{\partial \phi}$$

Result Dictionary Layout

The computed Hamiltonian vector fields are stored in a dictionary with the following structure:

$$\text{result\_dict}[(n, m)] = \{\text{'symbolic'} : H_{\text{symbolic}}, \text{'numeric'} : H_{\text{numeric}}\}$$

10

[10] Modules were developed by M. A. Evangelista-Alvarado, J. C. Ruíz-Pantaleón, and P. Suárez-Serrato.

Dataset Parameters

When generating our datasets, we varied two key parameters:

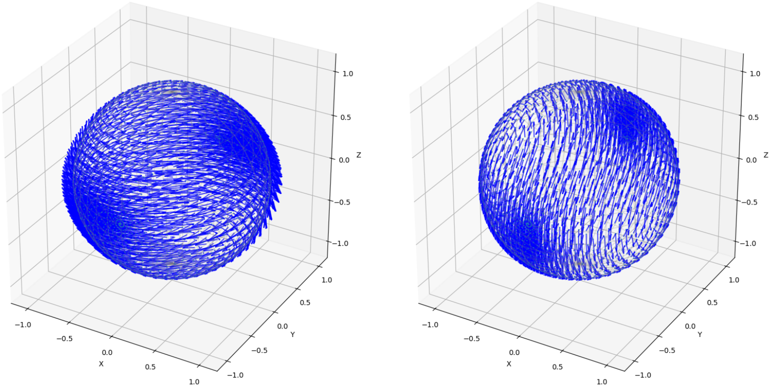
- `range_n`: The number of spherical harmonics considered.
- `num_points`: The number of points sampled along the spiral.

The combination of these parameters determines the size and resolution of the dataset.

Dataset Sizes

`range_n \n_samples`	100	250	500	1000
10	10,000	25,000	50,000	100,000
20	40,000	100,000	200.000	400,000
32	102,400	256,000	512.000	1,024,000

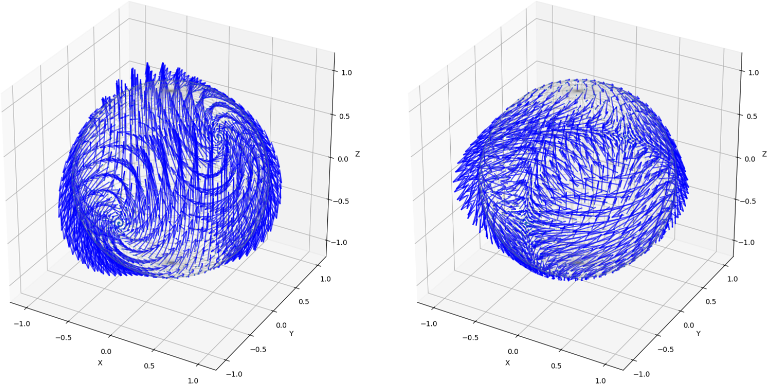
Dataset Visualization



Two vector fields for `n_samples = 1500` [11]

[11] Image created by the authors.

Dataset Visualization

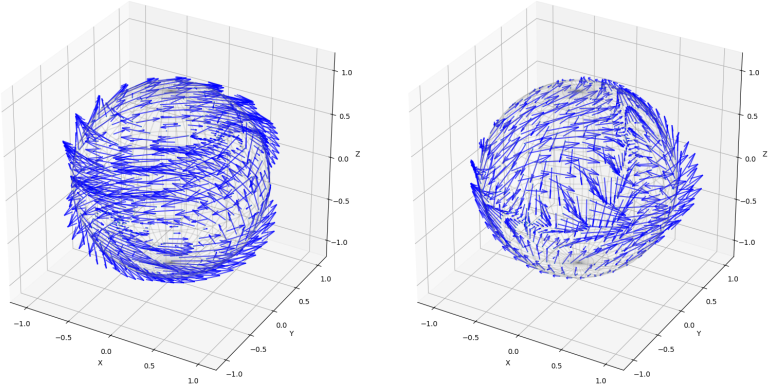


Two vector fields for `n_samples = 1000` [12]

[12] Image created by the authors.

Instituto de Matemáticas

Dataset Visualization



Two vector fields for `n_samples = 500` [13]

[13] Image created by the authors.

Analogy with Natural Language Processing

Sequence-to-Sequence Interpretation

With our dataset setup, we can draw a powerful analogy with Natural Language Processing (NLP):

- **Spirals are analogous to sentences.** Each spiral trajectory represents a sequence of vector field samples.
- **Vector field samples are analogous to words.** Each vector in the sequence corresponds to a "tokenïn the language of vector fields.

Reconstruction as Next Word Prediction

Our task of reconstructing a Hamiltonian vector field on the sphere then becomes analogous to predicting the next word in a sentence. Given a sequence of vectors along the spiral, the model predicts the subsequent vector.

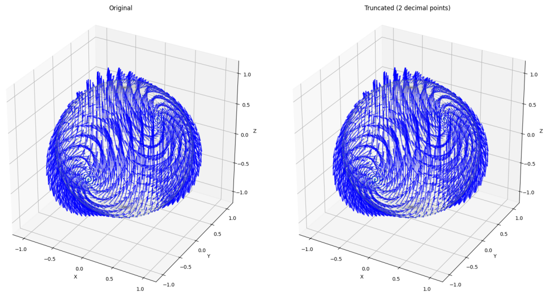
Resolution Parameter Adjustment: Controlling Vector Precision

Controlling Vector Precision

To mitigate numerical issues and manage the "vocabulary" size of our vector samples, we truncate the vector components during dataset generation.

- Numerical calculations can lead to very similar, but not exactly identical, vectors (e.g., (1, 1, 1) vs. (1.00001, 1.00001, 1.00001)).
- Truncating to a fixed number of decimals (e.g., 2 decimals) effectively treats these as the same vector.
- This reduces the number of unique vector samples, analogous to reducing the vocabulary size in NLP.

Resolution Parameter Adjustment: Vocabulary Reduction



Vocabulary Reduction and Lexical Diversity

For $n = 32$ and $\texttt{num\_points} = 250$, the ratio of unique vectors to total vectors is approximately 0.281 (28.1 %). a significant improvement from the original 77 % ratio.

Instituto de Matemáticas

Our Transformer Architecture

Our model utilizes a standard transformer encoder with the following key components:

- **Embedding Layer:** Maps input vector sequences to a higher-dimensional space.
- **Positional Encodings:** Sinusoidal encodings added to embeddings to provide sequence order information.
- **Multi-Head Attention:** Allows the model to attend to different parts of the input sequence.
- **Masking (Padding Mask):** Used to handle sequences of varying lengths by masking padding tokens.
- **Feed-Forward Network:** Point-wise fully connected network for non-linear transformations.

Training Parameters

To find suitable model parameters, we conducted a search while keeping the following fixed:

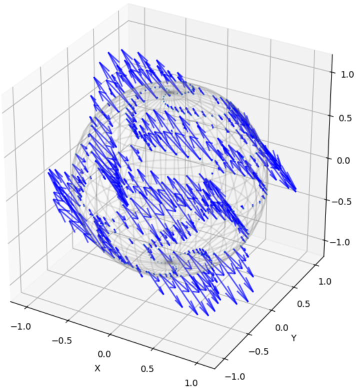
- `range_n` = 10 (number of spherical harmonics)
- `n_samples` = 50 (number of spiral points)
- `epochs` = 200
- `batch_size` = 128

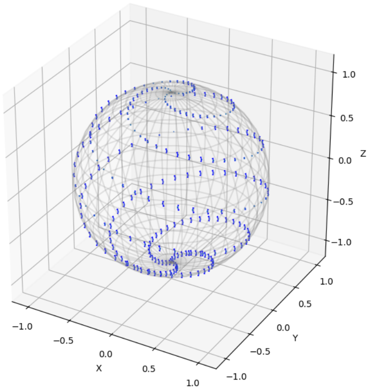
We explored the following model configurations:

Parameter	Config 1	Config 2	Config 3
`num_layers`	4	2	8
`d_model`	128	96	256
`num_heads`	8	4	16
`dff`	512	512	512
`dropout_rate`	0.1	0.2	0.2
`accuracy`	34 %	35 %	5 %

Instituto de Matemáticas

Reconstructions





Reconstruction Details

These reconstructions were generated with:

- `range_n` $= 10$ (number of spherical harmonics)
- `n_samples` $= 250$ (number of spiral points)

Instituto de Matemáticas