1]Ismael S. Silva [1]Universidade Federal do Ceará
ismaelsoares.pro@alu.ufc.br

Hyperbolic Modeling of Reputation: Karma as a Poincaré Disk Fiber Bundle

[

Abstract

We propose a novel model of social reputation—referred to as "karma"—based on hyperbolic geometry. Specifically, we represent an agent's moral or reputational state as a point on the Poincaré disk, where the magnitude indicates reputational strength and the angular coordinate encodes moral orientation. To incorporate context-dependence, we construct a fiber bundle over the disk, associating each point with a contextual space (e.g., professional, personal). The hyperbolic metric naturally models the sensitivity of high-reputation agents to minor transgressions and allows for counterintuitive proximities between morally opposite individuals. We formalize karma dynamics as a stochastic differential equation on the hyperbolic space. This geometric approach provides a lightweight, expressive alternative to complex probabilistic models, with potential applications in multi-agent reinforcement learning, social simulations, and ethical AI systems.

1 Introduction

Reputation systems are central to agent interaction in multi-agent environments, social simulations, and ethical modeling. Traditional approaches, such as scalar scores or graph-theoretic trust networks [?, ?], often fail to capture nuanced phenomena such as:

- **Moral vulnerability of the virtuous**: Small transgressions have disproportionate effects on individuals with high reputation.

- **Paradoxical trust**: Strongly altruistic agents may express unexpected affinity or trust toward agents with negative reputations.

- **Contextual incommensurability**: Agents with similar aggregate reputations may not trust each other due to differences in context (e.g., personal vs. professional).

To address these challenges, we introduce a geometric model of karma using the Poincaré disk, inspired by recent successes of hyperbolic embeddings in machine learning [?], and a fiber structure to incorporate context. This yields a rich but computationally tractable framework.

2 Geometric Karma Model

2.1 State Space: The Poincaré Disk

The agent's karma is modeled as a point K in the open unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$.

- $|z|$ measures reputation magnitude.

- $\arg(z)$ encodes moral orientation.

- $z = 0$ corresponds to a neutral state.

The distance between two karma states z_1 and z_2 is the hyperbolic distance in the Poincaré disk model:

$$d_H(z_1, z_2) = \text{arcosh}\left(1 + \frac{2|z_1 - z_2|^2}{(1 - |z_1|^2)(1 - |z_2|^2)}\right) \tag{1}$$

This metric amplifies small Euclidean displacements near the boundary ($|z| \to 1$), as illustrated in Figure **??**. This naturally models the instability of high karma values.
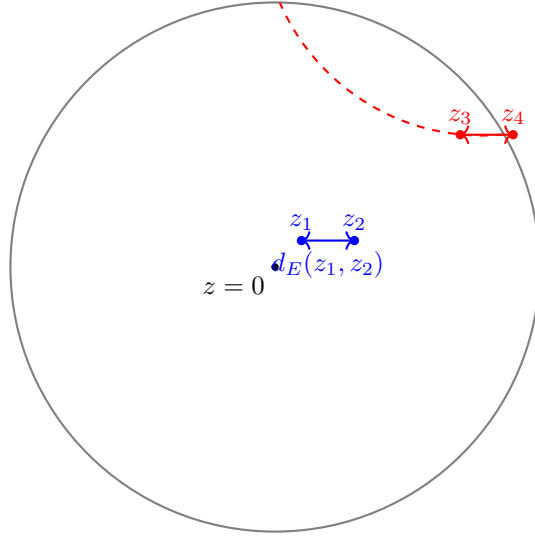

Figure 1: **Hyperbolic Metric Distortion.** Equal Euclidean distances result in different hyperbolic distances. Near the boundary, hyperbolic distances (dashed red line) grow sharply compared to distances near the center (solid blue line), modeling the reputational instability of agents with extreme karma.

2.2 Fiber Structure: Modeling Context

To account for context, we model the full reputation state as a fiber bundle over the karma disk $\mathbb{D}$. Each point $z \in \mathbb{D}$ (the base space) is associated with a context-specific fiber F_z. The total reputation state is a pair (z, v) where $v \in F_z$. This structure is visualized in Figure **??**.

2.3 Karma Dynamics

The evolution of an agent's karma state z over time is modeled by a stochastic differential equation:

$$\frac{dz}{dt} = -\nabla_H \Phi(z, v) + \rho \cdot \xi(t) \tag{2}$$

where:

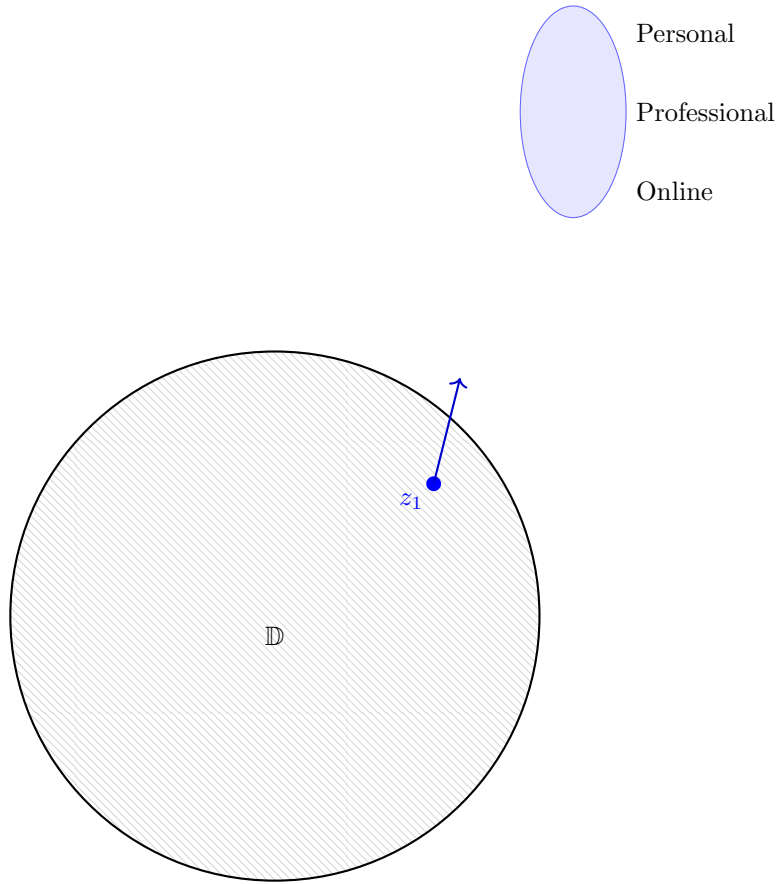- $\Phi(z, v)$ is a reputational potential field that depends on both the karma state z and its context v.

Figure 2: **Fiber Structure over the Karma Base Space.** Each karma state z in the base disk $\mathbb{D}$ has an associated contextual fiber F_z, which captures domain-specific nuances of reputation.

- ∇_H is the gradient that respects the hyperbolic geometry of the space.

- $\xi(t)$ is a stochastic noise term representing random events (e.g., scandals, misinformation), scaled by a factor ρ.

This formalism allows for complex behaviors, such as the paradoxical trust shown in Figure **??**, where morally opposite agents might find themselves "close" along a shared hyperbolic geodesic.
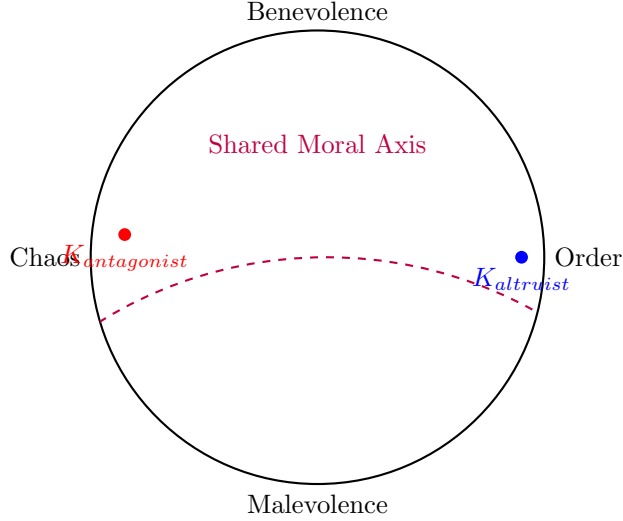

Figure 3: **Geodesics and Moral Extremes.** Even agents with antagonistic reputations (e.g., altruist vs. antagonist) may share an alignment along a hyperbolic geodesic (a "shared moral axis"), enabling paradoxical trust or understanding.

3 Applications and Conclusion

This hyperbolic model of karma enables new forms of analysis and application:

- Reputation-aware policy learning in multi-agent reinforcement learning (MARL).

- Rich visualization of social and moral dynamics in simulations.

- Integration with procedural content generation to create ethically responsive environments.

- Simulation of complex social phenomena like ethical drift, moral injury, and reputational recovery.

Its key strength lies in using a principled geometric structure to encode complex social phenomena that would otherwise require cumbersome probabilistic machinery. The resulting framework is fast, interpretable, and extensible.

References

[1] J. Sabater and C. Sierra, *Review on Computational Trust and Reputation Models*, Artificial Intelligence Review, 2005.

[2] A. Jøsang and R. Ismail, *The Beta Reputation System*, Proceedings of the 15th Bled Electronic Commerce Conference, 2002.

[3] M. Nickel and D. Kiela, *Poincaré Embeddings for Learning Hierarchical Representations*, Advances in Neural Information Processing Systems (NeurIPS), 2017.