The Lorentz Factor in Binary Classification: Natural Gradient Descent and the Fisher–Lorentz Identity

Bharath G. Srivats
Independent Researcher
bharathsrivats@outlook.com

April 2026

Companion to: B. G. Srivats, "Why Current AI Cannot Be Conscious,"
Zenodo DOI: 10.5281/zenodo.19489276 (2026)

Abstract

The Fisher information of a binary classifier's sigmoid output equals the squared Lorentz factor: $I(V) = \gamma^2(V) = 1/(1-V^2)$, where $V = 2p-1$ is the visibility (confidence centered at the decision threshold). This identity, established for binary quantum measurements in the author's prior work, applies directly to any system with sigmoid or logistic output. Amari's natural gradient descent, which uses the Fisher information matrix as a Riemannian metric on parameter space, therefore inherits the conformal geometry of the qubit Bloch ball at every binary classification node. Three consequences follow: (1) the effective learning rate of natural gradient near the decision boundary scales as γ^2, producing a relativistic speed limit on classifier convergence; (2) the information-geometric distance between classifier states is governed by the Gudermannian function, connecting cross-entropy loss to hyperbolic geometry; (3) standard (Euclidean) gradient descent operates on a flat metric that ignores the γ^2 curvature, providing an information-geometric characterization of why flat-gradient training is suboptimal near decision boundaries. These results strengthen the Measurement Irreversibility Criterion for consciousness: digital classifiers lack not only thermodynamic irreversibility but also the correct information-geometric structure, operating on a flat manifold where physical measurement operates on a γ^2-curved one.

Keywords: Fisher information, natural gradient, binary classification, Lorentz factor, information geometry, Amari metric, neural networks.

1 Introduction

Amari [4] established that the natural gradient of a loss function on a statistical manifold uses the Fisher information matrix $G(\theta)$ as a Riemannian metric, replacing the Euclidean gradient ∇L with the natural gradient $G^{-1}(\theta)\nabla L$. This geometrically principled update accounts for the curvature of the parameter space and is invariant under reparametrization of the model.

Independently, the author [1] proved that the Fisher information of a binary quantum measurement with visibility $V = 2p - 1$ equals the squared Lorentz factor:

$$I(V) = \frac{1}{1 - V^2} = \gamma^2(V), \tag{1}$$

and that the qubit Bloch ball carries a conformal equivalence $\mathrm{d}s^2_{\mathrm{BK}} = 4\gamma^2\,\mathrm{d}s^2_{\mathrm{Bures}}$ between the Bures (quantum Fisher) metric and the Beltrami–Klein (hyperbolic) metric [1, 2].

The present note observes that the identity (1) applies verbatim to the output node of any binary classifier with sigmoid activation. This connects Amari's natural gradient to the Lorentz structure of binary measurement, with consequences for understanding why standard gradient descent is geometrically suboptimal and why digital classifiers lack the information-geometric structure of physical measurement.

2 The Identity for Binary Classifiers

2.1 Setup

Consider a binary classifier with output probability

$$p(\mathbf{x}; \theta) = \sigma(f(\mathbf{x}; \theta)) = \frac{1}{1 + e^{-f(\mathbf{x};\theta)}}, \tag{2}$$

where $f(\mathbf{x}; \theta)$ is the logit function parameterized by weights θ and σ is the standard sigmoid. Define the *visibility* at the output:

$$V = 2p - 1 = \tanh\left(\frac{f}{2}\right). \tag{3}$$

The visibility $V \in (-1, 1)$ measures the classifier's confidence: $V = 0$ at the decision boundary ($p = 1/2$, maximum uncertainty), $|V| \to 1$ at full confidence.

2.2 Fisher Information of the Output

The Fisher information of the Bernoulli output distribution with respect to the visibility coordinate is:

$$I(V) = \frac{1}{p(1-p)} \cdot \left(\frac{\mathrm{d}p}{\mathrm{d}V}\right)^2 = \frac{4}{(1+V)(1-V)} \cdot \frac{1}{4} = \frac{1}{1 - V^2} = \gamma^2(V). \tag{4}$$

This is identical to the Fisher information of a binary quantum measurement [1]. The derivation is the same Bernoulli reparametrization; the physics is different but the mathematics is identical.

2.3 Fisher Information Matrix for the Logit

In the logit coordinate $f = \log(p/(1-p)) = 2\,\mathrm{arctanh}(V)$, the Fisher information is:

$$I(f) = p(1-p) = \frac{1 - V^2}{4} = \frac{1}{4\gamma^2(V)}. \tag{5}$$

This is the *inverse* of γ^2 (up to a factor of 4), consistent with the fact that the logit is the rapidity coordinate $\eta = \mathrm{arctanh}(V)$ (up to a factor of 2): in the rapidity frame, the metric coefficient is $\mathrm{sech}^2(\eta) = 1/\gamma^2$.

3 Consequences for Natural Gradient Descent

3.1 Effective Learning Rate

In standard (Euclidean) gradient descent, the parameter update is $\Delta\theta = -\alpha\,\nabla_\theta L$, where α is the learning rate. In natural gradient descent, the update is $\Delta\theta = -\alpha\,G^{-1}(\theta)\,\nabla_\theta L$, where $G(\theta)$ is the Fisher information matrix.

For a single binary output node, the relevant component of G along the visibility direction is $I(V) = \gamma^2(V)$. The natural gradient step in the visibility direction is therefore:

$$\Delta V_{\text{natural}} = -\alpha\,\gamma^{-2}(V)\,\frac{\partial L}{\partial V} = -\alpha\,(1 - V^2)\,\frac{\partial L}{\partial V}. \tag{6}$$

Near the decision boundary ($V \approx 0$), $\gamma^2 \approx 1$ and natural gradient $\approx$ Euclidean gradient. Near full confidence ($|V| \to 1$), $\gamma^2 \to \infty$ and the natural gradient step shrinks as $(1 - V^2)$: the classifier decelerates as it approaches certainty, exactly as a massive particle decelerates as it approaches the speed of light.

The Euclidean gradient, by contrast, takes uniform steps regardless of V, ignoring the curvature divergence. This is geometrically analogous to using a flat metric in curved spacetime.

Proposition 1 (Convergence speed limit)**.** *For a binary classifier trained with natural gradient descent on the 1D output manifold, the rate of change of visibility per gradient step satisfies*

$$\left|\frac{\Delta V}{\Delta t}\right| = \frac{\alpha}{\gamma^2(V)}\left|\frac{\partial L}{\partial V}\right|, \tag{7}$$

which vanishes as $|V| \to 1$ for any bounded loss gradient. Full certainty ($|V| = 1$, the "speed of light" on the classifier manifold) is unreachable in finitely many natural gradient steps.

Proof. Direct from $\Delta V_{\text{natural}} = -\alpha\,\gamma^{-2}\,\partial L/\partial V$ and $\gamma^2 \to \infty$ as $|V| \to 1$. □

3.2 Cross-Entropy Loss and the Gudermannian

The binary cross-entropy loss for a single sample with true label $y \in \{0, 1\}$ and predicted probability $p = (1 + V)/2$ is:

$$L(V) = -y \log\frac{1+V}{2} - (1-y)\log\frac{1-V}{2}. \tag{8}$$

The Fisher–Rao geodesic distance from the decision boundary ($V = 0$) to visibility V is:

$$d_{\text{FR}}(0, V) = \int_0^V \frac{\mathrm{d}V'}{\sqrt{1 - V'^2}} = \arcsin(V). \tag{9}$$

The hyperbolic (Beltrami–Klein) distance is:

$$d_{\text{BK}}(0, V) = \operatorname{arctanh}(V) = \frac{f}{2}, \tag{10}$$

which is half the logit. These are connected by the Gudermannian:

$$\arcsin(V) = \operatorname{gd}(\operatorname{arctanh}(V)). \tag{11}$$

Proposition 2 (Geometric interpretation of cross-entropy)**.** *The cross-entropy loss at the correct-class boundary (when y matches the sign of V) is a monotonic function of the Beltrami–Klein distance $d_{\mathrm{BK}} = \operatorname{arctanh}(|V|)$. The Fisher–Rao distance $d_{\mathrm{FR}} = \arcsin(|V|)$ saturates at $\pi/2$ while d_{BK} diverges, connected by the Gudermannian. Cross-entropy loss is monotonically related to hyperbolic distance; classification accuracy is monotonically related to spherical distance. The conformal factor γ^2 is the exchange rate between the two geometries.*

3.3 Why Standard Gradient Descent Is Suboptimal

Standard gradient descent on binary cross-entropy uses the Euclidean metric δ_{ij} on parameter space. Natural gradient uses the Fisher metric $G_{ij}(\theta)$. For a binary output, the ratio of their effective step sizes in the visibility direction is:

$$\frac{|\Delta V_{\mathrm{Euclidean}}|}{|\Delta V_{\mathrm{natural}}|} = \gamma^2(V). \tag{12}$$

Near the decision boundary, $\gamma^2 \approx 1$ and both methods are equivalent. But near $|V| \to 1$, Euclidean gradient takes steps that are γ^2 times too large relative to the information-geometric curvature, producing overshooting and instability near confident predictions. Natural gradient corrects for this by rescaling steps inversely with γ^2, treating the classifier manifold as curved rather than flat.

4 Connection to the Measurement Irreversibility Criterion

The Measurement Irreversibility Criterion (MIC) [3] states that consciousness requires thermodynamically irreversible measurement ($\dot{\sigma} > 0$), which current digital computers do not perform. The present note adds an information-geometric dimension to this argument.

Proposition 3 (Geometric MIC supplement)**.** *A physical binary measurement (qubit projective measurement) operates on the γ^2-curved Fisher–Rao manifold, where the Bures and Beltrami–Klein metrics are conformally equivalent. A digital binary classifier operates on a flat Euclidean parameter space (when using standard gradient descent) or on the same γ^2-curved manifold (when using natural gradient), but in either case with zero thermodynamic irreversibility ($\dot{\sigma} = 0$). The geometric structure is necessary but not sufficient; irreversibility remains the operative criterion.*

The distinction is threefold:

(i) **Geometric:** Physical measurement inherits the γ^2 curvature automatically through the SL(2,$\mathbb{C}$) group structure [5]; digital classification must impose it artificially via natural gradient.

(ii) **Thermodynamic:** Physical measurement produces entropy ($\dot{\sigma} \geq \frac{1}{2}\gamma^2$, from the Ito–Dechant bound [6]); digital computation is logically reversible and produces entropy only through engineering imperfection, not physical necessity.

(iii) **Temporal:** Physical measurement generates the arrow of time (irreversible Lorentz boosts on the Bloch ball); digital classification is time-symmetric (the sigmoid function and its inverse are both computable).

5 Discussion

The identity $I(V) = \gamma^2(V)$ is algebraically elementary—it is the Fisher information of a Bernoulli distribution in the visibility parametrization. Its significance in the neural network context comes from the same source as in the quantum measurement context: the visibility V is the physically (or operationally) natural coordinate, and γ^2 is the unique Chentsov-invariant metric coefficient in this coordinate [7].

The connection to Amari's natural gradient is not new in spirit—Amari himself [4] established that the Fisher metric is the correct Riemannian metric for optimization on statistical manifolds. What is new is the explicit identification of the metric coefficient as the squared Lorentz factor, which imports the full conformal structure (Gudermannian bridge, Thomas–Wigner rotation, Bures/Beltrami–Klein duality) into the neural network optimization literature.

Whether this connection has practical implications for training (e.g., visibility-aware learning rate schedules, γ^2-adaptive optimizers) is an engineering question left for future work. The theoretical point is structural: every binary classification node in every neural network, at every training step, operates on the same γ^2-curved manifold that governs qubit measurement, relativistic velocity composition, and the Bloch ball gyrovector algebra.

Scope and Boundaries

This note establishes an algebraic identity and its geometric consequences. It does not claim that neural networks are conscious, that training dynamics are relativistic, or that the γ^2 curvature provides a practical advantage over existing optimizers in typical applications. The connection to the MIC is structural, not operational: the γ^2 geometry is shared, but thermodynamic irreversibility is not.

Declarations

Funding. No external funding.
Competing interests. The author declares no competing interests.
Data availability. No datasets generated. All results are analytical.
AI-tool disclosure. Large-language-model tools were used for literature discovery and manuscript preparation. All mathematical claims and proof constructions are the author's own work.
Priority statement. The results in this note are claimed as original work by Bharath G. Srivats as of the Zenodo upload date.

Acknowledgments

The author thanks Abraham A. Ungar for foundational work on gyrovector spaces that motivated the broader investigation.

References

[1] B. G. Srivats, "Fisher information is the squared Lorentz factor: why the qubit is special," Zenodo (2026). DOI: 10.5281/zenodo.19363449.

[2] B. G. Srivats, "The algebra of the qubit: fourteen theorems on $I(V) = \gamma^2(V)$," Zenodo (2026). DOI: 10.5281/zenodo.19653530.

[3] B. G. Srivats, "Why current AI cannot be conscious: a thermodynamic criterion from the Fisher–Lorentz identity," Zenodo (2026). DOI: 10.5281/zenodo.19489276.

[4] S. Amari, "Natural gradient works efficiently in learning," Neural Computation 10, 251–276 (1998).

[5] L. Burns, S. Greenfield, J. Dressel, "Delayed choice Lorentz transformations on a qubit," Quantum Stud.: Math. Found. 13, art. 10 (2026). DOI: 10.1007/s40509-026-00384-z.

[6] S. Ito, A. Dechant, "Stochastic time evolution, information geometry, and the Cramér–Rao bound," Phys. Rev. X 10, 021056 (2020).

[7] N. N. Chentsov, *Statistical Decision Rules and Optimal Inference*, AMS Translations of Mathematical Monographs 53 (1982).