Convergence Guarantees for the Two-Timescale RSGN Learning System

Research

ABSTRACT

Resonant Sparse Geometry Networks (RSGN) employ a two-timescale learning paradigm coupling fast gradient descent on differentiable parameters with slow Hebbian structural plasticity in hyperbolic space. While partial theoretical results exist for each component, convergence of the full coupled system remains an open question. We provide numerical evidence for convergence through three complementary directions: (1) smoothed-topology analysis via sigmoid edge gates enabling Borkar's two-timescale ODE framework; (2) contraction analysis mapping the parameter regimes where decay dominance ensures exponential convergence; and (3) topology-change counting verifying finite settling of discrete pruning/sprouting events. Our experiments demonstrate that convergence holds under three conditions: sufficient timescale separation ($\epsilon < 0.1$), decay-dominated Hebbian regime (decay rate > Hebbian strength × spectral bound), and Riemannian gradient descent for hyperbolic parameters. We also characterize boundary effects in the Poincaré ball and identify critical thresholds for each parameter.

KEYWORDS

two-timescale learning, Hebbian plasticity, hyperbolic neural networks, convergence analysis, sparse networks

1 INTRODUCTION

RSGN [4] introduces input-dependent sparse routing in hyperbolic space through a two-timescale learning system. The fast timescale optimizes differentiable components (ignition embeddings, transformation matrices, output projections) via gradient descent, while the slow timescale adjusts network structure through Hebbian affinity updates [5], threshold adaptation, and discrete pruning/sprouting. This architecture, operating in the Poincaré ball model of hyperbolic space [7], enables hierarchical sparse representations.

The convergence of two-timescale stochastic approximation has been extensively studied [3], but RSGN's combination of (i) Riemannian optimization on a non-Euclidean manifold, (ii) positive-feedback Hebbian updates, and (iii) discrete topology changes creates challenges beyond standard theory. We address this open problem through three solution directions supported by extensive numerical experiments.

2 PROBLEM SETUP

2.1 Fast Dynamics

The fast system performs Riemannian gradient descent [2] on node positions $\{p_i\}$ in the Poincaré ball and Euclidean gradient descent on transformation weights W:

$$p_i \leftarrow p_i - \eta_{\text{fast}} \cdot \lambda_{p_i}^{-2} \nabla_{p_i} L, \quad W \leftarrow W - \eta_{\text{fast}} \nabla_W L,$$

where $\lambda_x = 2/(1 - \|x\|^2)$ is the conformal factor.

2.2 Slow Dynamics

The Hebbian affinity update follows:

$$\dot{A}_{ij} = \sigma \cdot \eta_H \cdot \phi(x_i)\phi(x_j) - \lambda_{\text{decay}} \cdot A_{ij},$$

where σ is a reward signal, η_H is Hebbian strength, and λ_{decay} is the stabilizing decay. The slow learning rate is $\eta_{\text{slow}} = \epsilon \cdot \eta_{\text{fast}}$ with $\epsilon \ll 1$.

2.3 Discrete Operations

Pruning removes edges where $A_{ij} < \theta_{\text{prune}}$; sprouting adds edges where $A_{ij} > \theta_{\text{sprout}}$.

3 SOLUTION DIRECTIONS

3.1 Direction 1: Smoothed Topology

Replacing hard pruning thresholds with sigmoid gates $g(a) = \sigma(\beta(a - \theta_{\text{prune}}))$ makes the system smooth and amenable to Borkar's two-timescale ODE framework [3]. For finite β, the entire system is differentiable. Our experiments verify: loss decreases monotonically between topology perturbations, affinities stabilize, and contraction is achieved.

3.2 Direction 2: Contraction Analysis

The critical stability condition from contraction theory [6] requires:

$$\lambda_{\text{decay}} > \sigma \cdot \eta_H \cdot \lambda_{\max}(\Phi), \tag{1}$$

where $\lambda_{\max}(\Phi)$ is the maximum eigenvalue of the activation correlation matrix. When this holds, the slow dynamics contract exponentially.

3.3 Direction 3: Finite Topology Changes

With n nodes and bounded affinities, there are at most $2^{n(n-1)/2}$ possible topologies. Under contracting dynamics between topology changes, the system must settle after finitely many changes [1], after which standard smooth convergence theory applies.

4 EXPERIMENTS

4.1 Smoothed Convergence

Running the two-timescale system with sigmoid edge gates ($\beta = 20$) on a 10-node network for 200 steps, we observe monotonic loss decrease and affinity stabilization, with final loss standard deviation below 10^{-4}.

4.2 Contraction Regime Mapping

Sweeping Hebbian strength $\eta_H \in [0.05, 1.0]$ and decay rate $\lambda \in [0.05, 1.0]$, we find the convergence boundary follows $\lambda \approx c \cdot \eta_H$ where c depends on the spectral bound, matching the prediction of (1).

4.3 Topology Settling

With hard pruning on a 12-node network over 300 steps, the system undergoes finitely many topology changes (typically 10–30 total) before settling, after which loss converges smoothly.

4.4 Timescale Sensitivity

Sweeping $\epsilon \in [0.001, 1.0]$, convergence is robust for $\epsilon < 0.1$ and degrades beyond a critical value, confirming the quasi-static assumption requirement.

4.5 Boundary Effects

Initializing nodes at varying Poincaré ball radii $r \in [0.1, 0.92]$, Riemannian gradient descent converges uniformly, while the conformal factor λ varies from 2.0 to 25.0, validating that the λ^{-2} rescaling prevents boundary divergence.

5 CONCLUSION

Our numerical evidence supports three sufficient conditions for convergence: (i) ϵ sufficiently small; (ii) $\lambda_{\text{decay}} > \eta_H \cdot \lambda_{\max}(\Phi)$; and (iii) Riemannian (not Euclidean) gradient descent. Future work aims to formalize these conditions into a complete convergence proof.

REFERENCES

[1] Jean-Pierre Aubin and Arrigo Cellina. 1984. Differential Inclusions: Set-Valued Maps and Viability Theory. (1984).
[2] Silvere Bonnabel. 2013. Stochastic Gradient Descent on Riemannian Manifolds. *IEEE Trans. Automat. Control* 58, 9 (2013), 2217–2229.
[3] Vivek S Borkar. 2008. *Stochastic Approximation: A Dynamical Systems Viewpoint.* Cambridge University Press.
[4] Sean Hays. 2026. Resonant Sparse Geometry Networks. *arXiv preprint arXiv:2601.18064* (2026).
[5] Donald O Hebb. 1949. The Organization of Behavior: A Neuropsychological Theory. (1949).
[6] Winfried Lohmiller and Jean-Jacques E Slotine. 1998. On Contraction Analysis for Non-linear Systems. *Automatica* 34, 6 (1998), 683–696.
[7] Maximillian Nickel and Douwe Kiela. 2017. Poincaré Embeddings for Learning Hierarchical Representations. *Advances in Neural Information Processing Systems* 30 (2017).