Milestone 008 — What does the sphere actually do?

Tyler J.H.G.

Abstract

Which proposed mechanisms survive frozen tests of the model's dynamics?

Research period: September 3–5, 2026
Historical anchors: #201–#206

The normalization derivative constrained the local behavior, but several explanations remained possible. The unit constraint could remove an unhelpful magnitude channel, suppress a shortcut, preserve useful tangent directions, or change how perturbations accumulate through recurrent depth.

I could test those explanations on the model I already had. Change its radius, compare radial and tangent sensitivity, propagate perturbations through recurrent steps, and measure which changes reach the output. Information probes could then test what was recoverable from different parts of the state.

If the frozen model contradicted a proposed mechanism, that would give me a reason to reject the mechanism before training another model around it.

1. Visual intuition

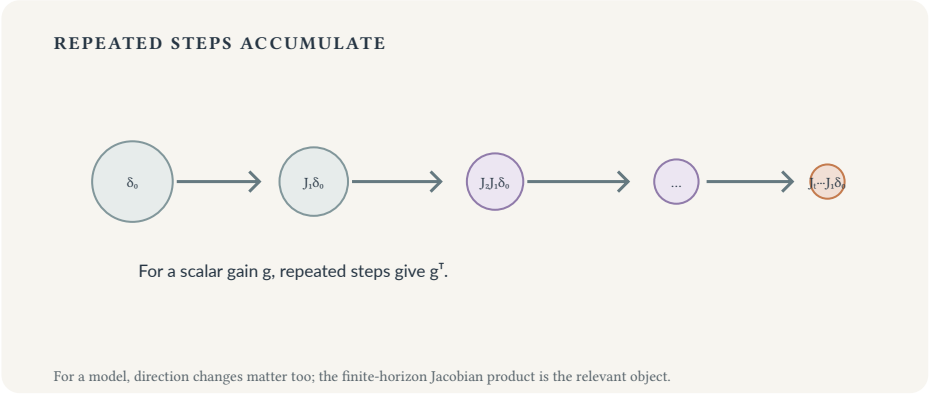

Figure 1: A toy scalar contraction compounds across steps. In a recurrent model, the full calculation uses a sequence of Jacobians.

Published Sep 29, 2026

2. The mechanism campaign

The tests returned:

- radius interventions: **MIXED**
- radial/tangent Jacobian analysis: **RADIAL-WEAK**
- finite-horizon Jacobian products: **NO-SPECTRAL-SEPARATION**
- output sensitivity: **RADIAL-LOCALLY-ACTIVE**
- information probes: **MIXED**

Each label answers a different question. They need to be interpreted together, with their interventions kept explicit.

Changing radius affected some measurements without producing a simple monotonic effect. The local derivatives showed weaker radial response than tangent response, consistent with the normalization theorem. Across recurrent steps, the accumulated Jacobians lacked the clean spectral separation expected from a strong selective-contraction explanation.

Radial perturbations could still affect the output locally. Removing radial variation at one normalization step therefore did not make every radius-related degree of freedom elsewhere in the computation irrelevant. The probes likewise supplied mixed evidence about radius as useful signal or harmful shortcut.

> **Sticky note — Jacobian:** a matrix describing how small changes at an input or internal state affect the output locally. Comparing directions reveals which perturbations have the strongest local effect.

> **Sticky note — finite-horizon Jacobian:** the product of local Jacobians along a finite sequence of recurrent updates. It describes how a small perturbation propagates through that sequence.

= Try a small example

Two hand-chosen matrices show how local Jacobians combine over a finite horizon.

```
import numpy as np
J0=np.diag([0.5,1.0]); J1=np.array([[1.,0.2],[0.,0.8]])
J10=J1@J0
print('finite-horizon Jacobian:\n',J10)
print('singular values:',np.linalg.svd(J10,compute_uv=False))
```

Saved output:

```
finite-horizon Jacobian:
 [[0.5 0.2]
 [0.  0.8]]
singular values: [0.83792489 0.47736976]
```

3. Why there wasn't a sixth experiment

The mechanism plan allowed another training experiment if the diagnostics exposed a question that required one. The proposed control would ignore radius while remaining meaningfully different from the existing unit model.

Once I specified how it would ignore radius, the construction reduced to essentially the model already tested. Another training run would produce another measurement of that construction, without separating the competing explanations.

I rejected the experiment for that reason. A control needs to create a comparison whose possible outcomes distinguish the hypotheses; specifying this one exposed that it could not.

4. Research context

Dynamical systems and sensitivity analysis supplied the tools for this pass. A one-step derivative describes a local response; products of derivatives describe how the response accumulates along a trajectory. Recurrence can amplify, rotate, cancel, or redistribute a perturbation after its first step.

The distinction matters when interpreting the result. The normalization theorem, a finite sequence of trained Jacobians, and an asymptotic claim about recurrence each concern a different object and require their own assumptions.

5. Interpretation

The diagnostics gave mixed support for the radial-shortcut and selective-contraction explanations. An intervention could establish local influence, while held-out prediction was still needed to establish whether that influence was useful for the task.

The proposed extra control reduced to a comparison I had already made. Recognizing that equivalence eliminated a training run and left the unresolved mechanisms stated more precisely.

6. Public formal sources

- NormalizationComposition
- RadialMemory

These pinned sources state the assumptions and checked mathematical conclusions.

7. Sources and chronology

- Milestone notebook 008 — preserved chronological source
- Research chronology
- Publication dispositions and Atlas-to-article map

The matching Atlas theme is **Finite-horizon Jacobian products**. The earlier Visual Intuition Atlas notebook preserves the source visualization.

The next chronological article is **Milestone 009 — From hypothesis sprawl to a theorem ledger**.