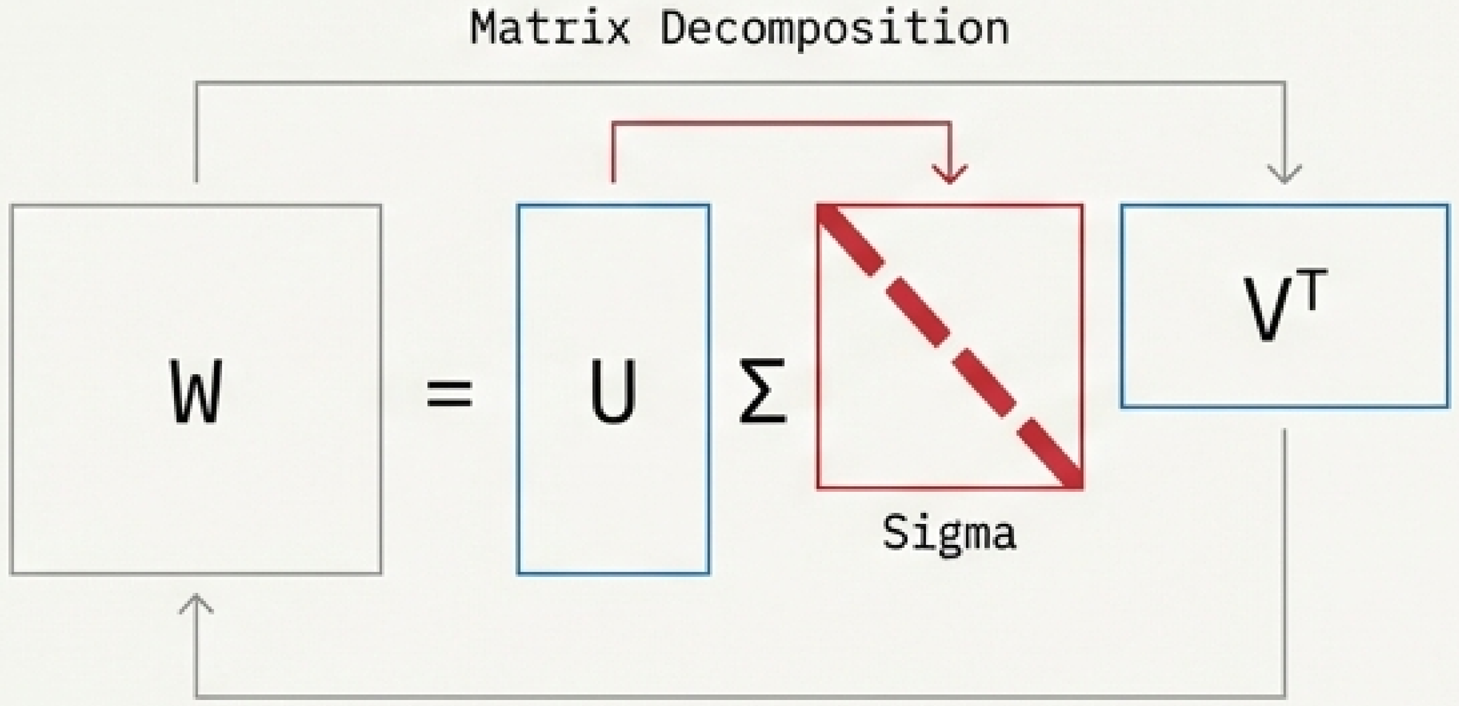
The Unlearned Flaw: Identifying & Correcting Spectral Optimization Errors

Why Neural Networks Fail to Learn Condition Numbers and How to Fix It.



Based on "Condition Number as a Non-Learned Matrix Property" (Anonymous Authors).
An investigation into the 13,000x error amplification in deep networks.

Executive Summary: The ‘Silent Killer’ in Optimization

The Vitals

- **The Symptom:** Optimizers learn Scale (Norms) well (~0.13 error) but fail to learn Shape (Condition Number).

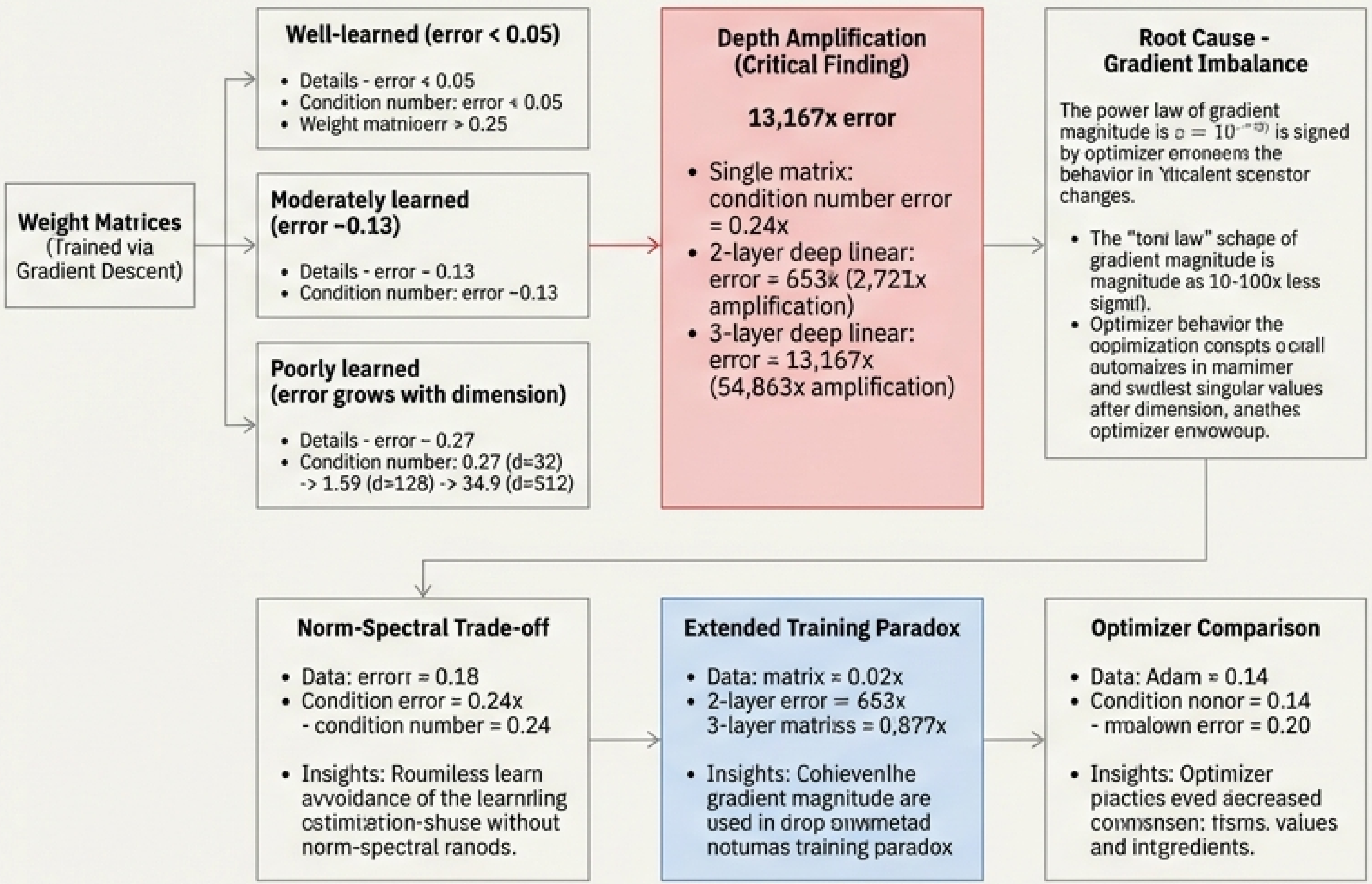
- **The Severity:** Errors amplify exponentially. A 3-layer linear net sees 13,167x error amplification.

- **The False Cure:** Adam performs worse than SGD (0.60 vs 0.14 overall error). Weight decay does not help.

- **The Root Cause:** Gradient imbalance. The optimizer is “deaf” to the smallest singular values (10-100x less signal).

- **The Solution:** Hybrid Regularization. Combine Learnable Multipliers (for Scale) + Spectral Regularization (for Shape).

The Diagnosis - Visual



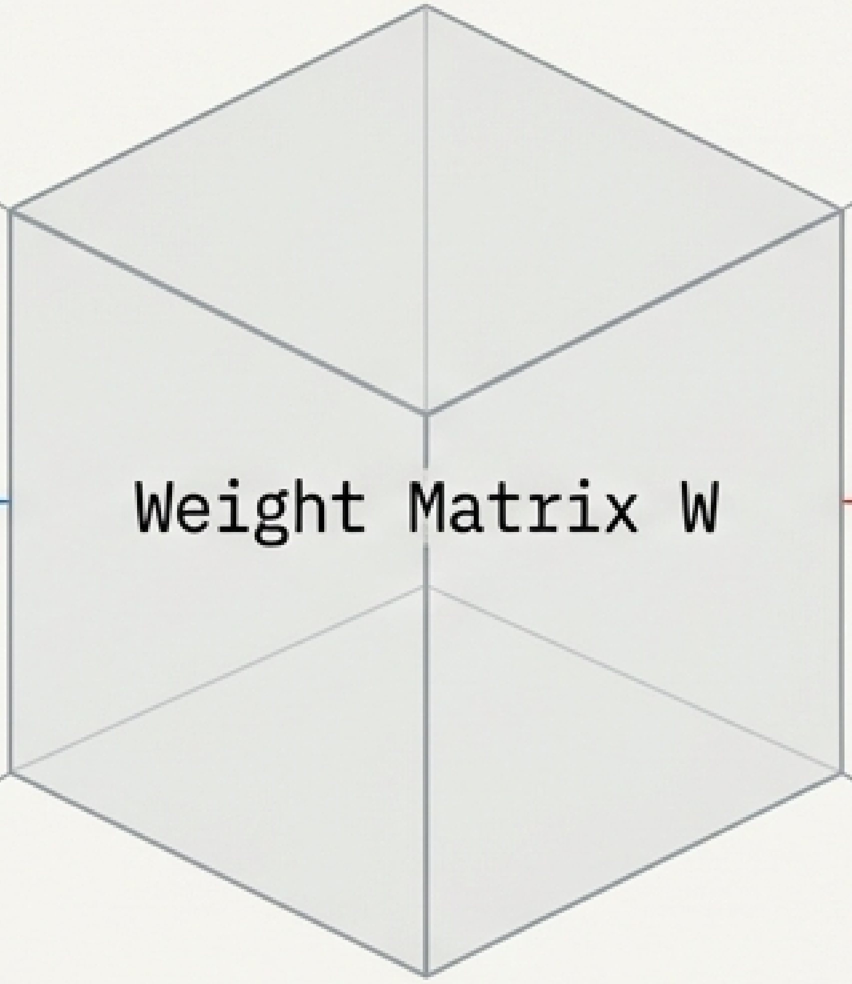
Anatomy of a Weight Matrix

Scale (Norms)

Volume / Size

- Row Norms
- Column Norms
- Frobenius Norm

└ Easily Learned



Spectrum (Shape)

Internal Structure

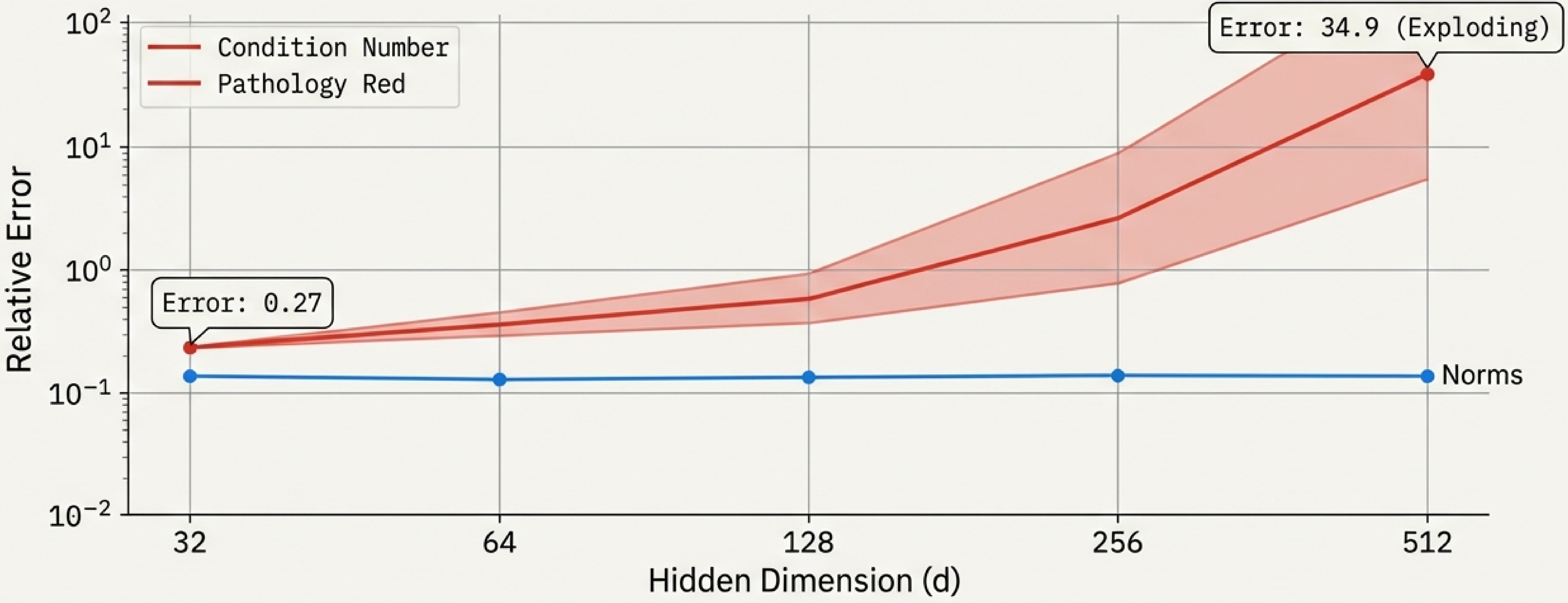
- Singular Values (Sigma)
- Spectral Gap
- Condition Number (Kappa)

└ Condition Number:
Kappa = Sigma_max / Sigma_min

The Research Question: We know Gradient Descent struggles with Scale.
Does it get the Internal Structure right?

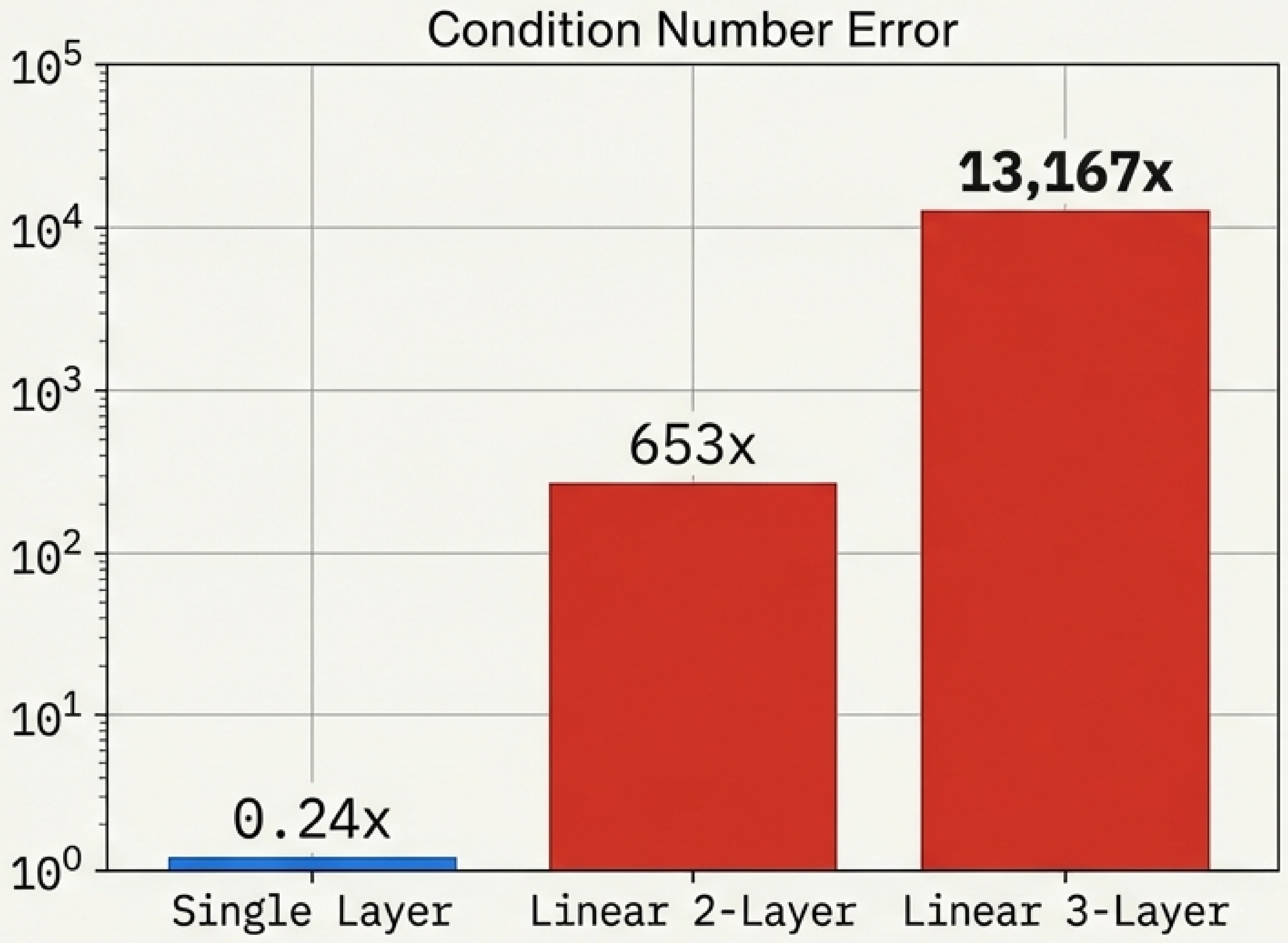
Clinical Swiss Editorial

The Diagnosis: Condition Number is Broken



While Norm errors stay low (~0.13), Condition Number errors explode super-linearly with dimension.

The Progression: Errors Amplify Exponentially with Depth



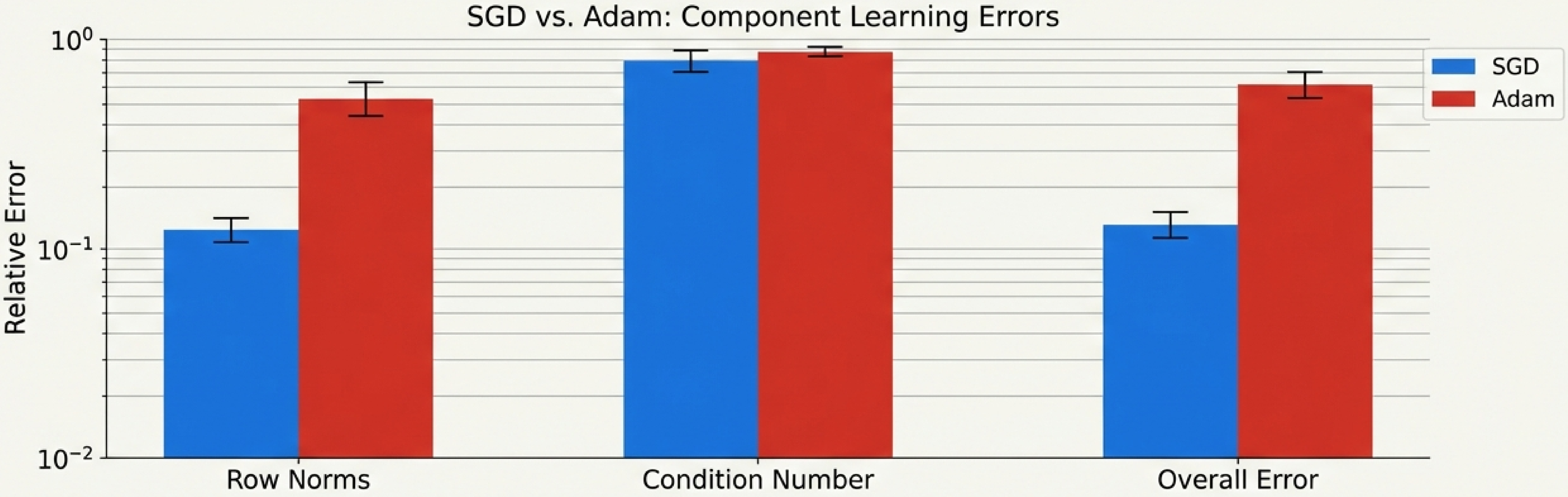
The Pathology of Depth:

Errors in individual layers do not add up; they multiply.

1-Layer: 0.24x
2-Layer: 653x
3-Layer: 13,167x

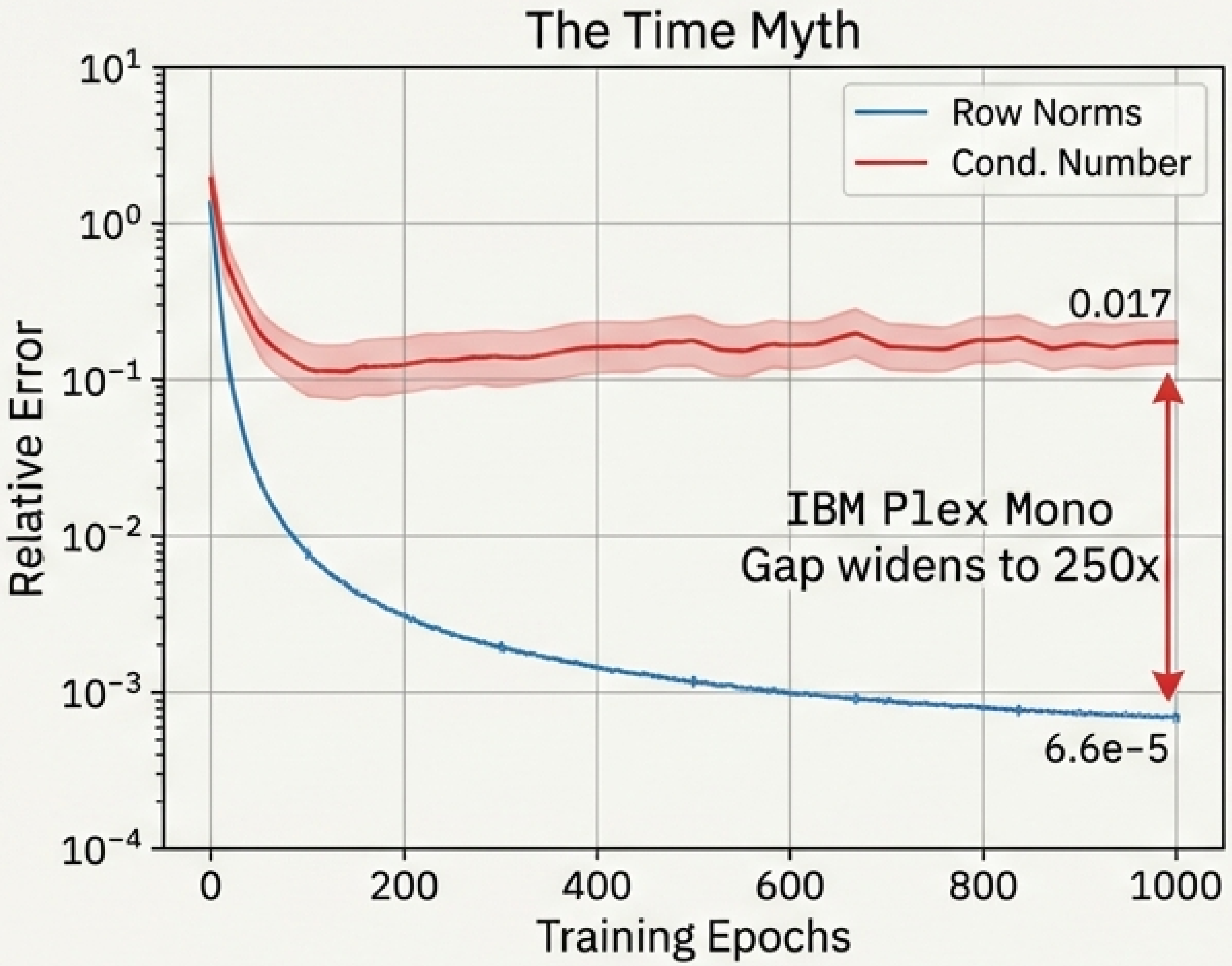
If a simple 3-layer net has a a 13,000x distortion, deep Transformers are likely severely compromised.

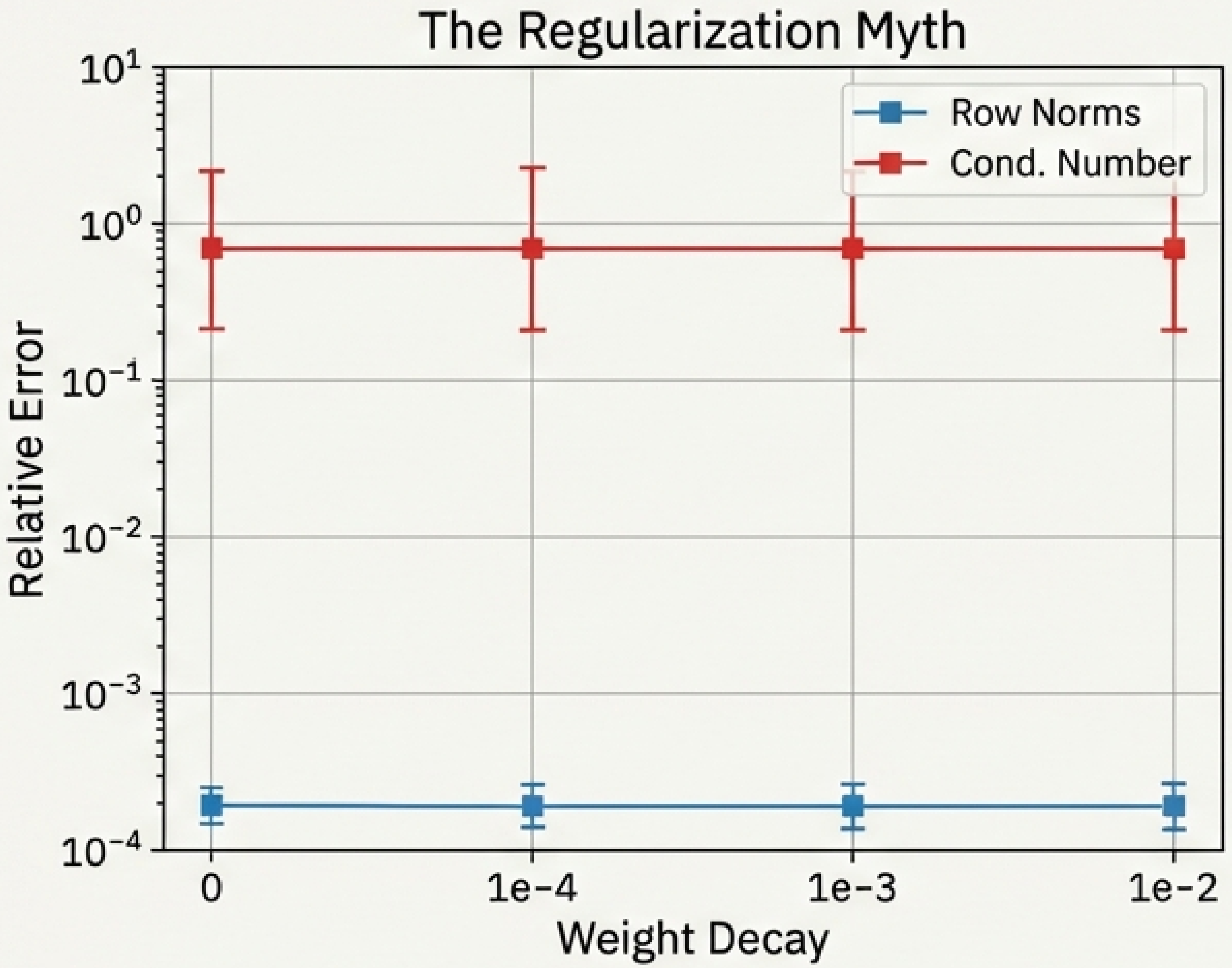
Standard 'Cures' Fail: The Case Against Adam



- Contrary to popular belief, Adam performs worse than SGD across all structural components.
- Mechanism: Adam's element-wise adaptive rates disrupt the coherent spectral structure.

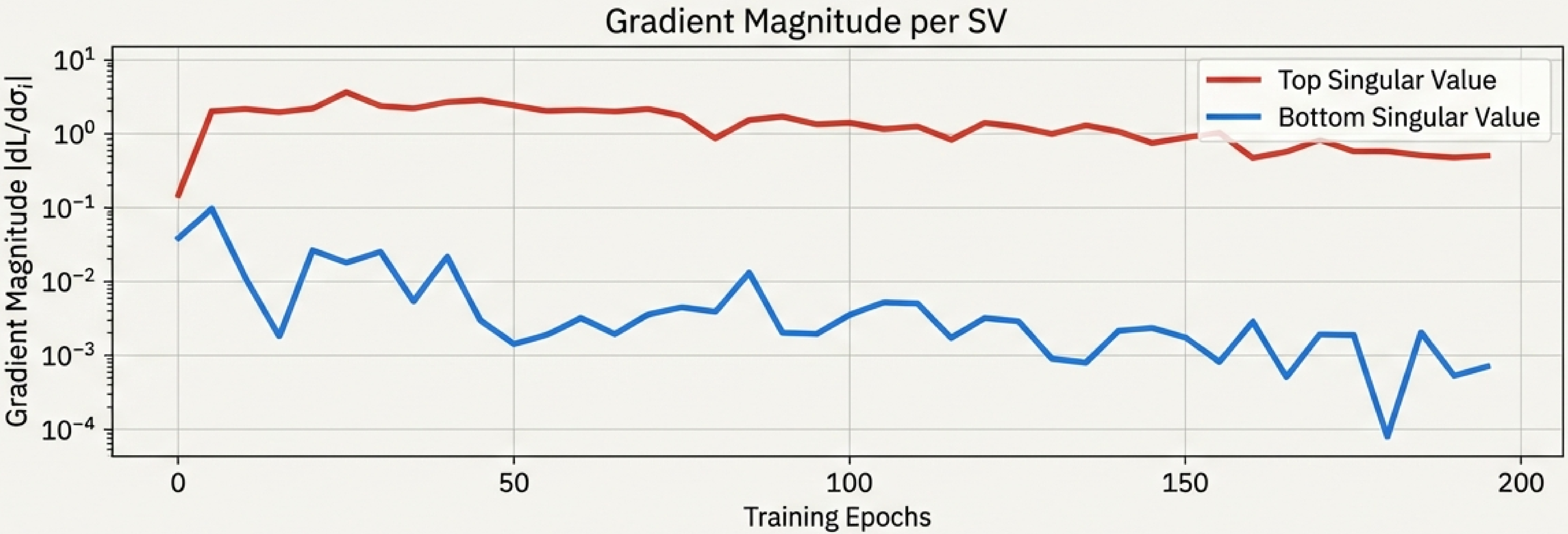
Chronic Condition: Why Time and Weight Decay Don't Help





Weight Decay has ZERO effect on spectral error.

The Root Cause: Gradient Signal Imbalance



Explanation:
The optimizer is 'deaf' to the smallest singular values.
Top singular values receive massive signal. Bottom values receive 10-100x less.
Consequence: Since Condition Number = Top / Bottom, and 'Bottom' never converges, the structure remains broken.

Clinical Trials: Four Corrective Strategies

Standard SGD The Baseline	Learnable Multipliers The Scale Fix	Spectral Regularization The Shape Fix	SVD Correction The Brute Force
Standard Gradient Descent (LR 0.01)	Per-row/column scaling. W_{eff} = diag(r) W diag(c)	Targeted penalty on condition number. Loss + lambda * (log Kappa(W) - log Kappa(target))^2	Periodic manual adjustment of singular values.

Pathology Red (#D93025) · Clinical Blue (#1A73E8) · Spectral Reg. · SVD Correction
IBM Plex Mono (#1A1A1A) · Neutral Grey (#9AA0A6)

The Discovery: A Fundamental Norm-Spectral Trade-off

Strategy	Norm Improvement	Condition Number Improvement
Learnable Multipliers	**67% (Great)**	10% (Fail)
Spectral Regularization	0% (Fail)	**77% (Great)**
SVD Correction	64%	32%

Distinct flaws require distinct correction mechanisms.
There is no “Silver Bullet”.

Prognosis for Large Language Models

- If a simple 3-layer net distorts structure by 13,000x, deep Transformers are operating with severely distorted spectral vitals.
- Symptoms in the Wild:
  - Training Instabilities
  - Need for Learning Rate Warmup
  - Inexplicable Loss Spikes
- Conclusion: Current LLMs are trained with 'unlearned' internal structures.

The Prescription: Hybrid Regularization

Solving the Dual Pathology

Future Architecture =

- **For Scale (Norms)** → **Learnable Multipliers**
- **For Structure (Shape)** → **Spectral Regularization**

To cure the model, we must treat both symptoms: Scale and Shape.