Geometric Decoupling: Diagnosing the Structural Instability of Latent Diffusion Models

Yuangbang Liang, Zhengwen Chen, Yu-Kun Lai · Cardiff University

Key Concept: Riemannian Geometric Framework for LDM Latent Space Diagnosis

Motivation & Problem

Latent Diffusion Models (LDMs) achieve remarkable image synthesis, yet their latent spaces are notoriously **brittle**:

- Minor latent perturbations $\Rightarrow$ **discontinuous semantic jumps**
- Inversion and interpolation frequently fail
- Existing explanations are *qualitative* — they cannot locate where or why the manifold breaks

Central Question: Why does a model capable of photo-realistic synthesis fail to maintain *semantic continuity* over microscopic latent distances?

We open this black box with a **Riemannian geometric lens**, decomposing the generative Jacobian into two orthogonal descriptors of **capacity** and **curvature**.

Main Contributions

- We identify **Geometric Decoupling**: under OOD(Out-of-Distribution) generation, LC(Local Complexity) no longer tracks useful high-frequency detail, while LS(Local Scaling) remains predictive.
- We quantify an **efficiency collapse**: OOD samples expend much more geometric effort without proportional perceptual gain.
- We reveal **pathological manifold structure** under OOD: spectral isolation, dimensional collapse, and tortuous interpolation trajectories.

Geometric Framework

Let $G: \mathcal{Z} \to \mathcal{X}$ be the generator. We approximate the **subspace Jacobian** via finite differences on a random P-dimensional subspace $\mathbf{W} \in \mathbb{R}^{E \times P}$:

$$\mathbf{J}_{\text{sub}} \cdot [\mathbf{w}_i] \approx \frac{G(\mathbf{z} + \epsilon \mathbf{w}_i) - G(\mathbf{z})}{\epsilon}$$

The **local metric tensor** $\mathbf{A} = \mathbf{J}_{\text{sub}}^{\top} \mathbf{J}_{\text{sub}}$ is eigen-decomposed as $\mathbf{A} = \mathbf{V} \boldsymbol{\Lambda} \mathbf{V}^{\top}$.

Local Scaling (LS) — Capacity

Measures *volume expansion* of the latent manifold:

$$\psi_{\boldsymbol{\omega}}(\mathbf{z}) = \frac{1}{2} \sum_{i=1}^{P} \log(\lambda_i) \cdot \mathbf{1}_{\{\lambda_i > 0\}}$$

High LS $\Rightarrow$ dense information encoding in that region.

Local Complexity (LC) — Curvature

Measures the *rotation rate* of the principal direction $\mathbf{V}_1$:

$$\delta(\mathbf{z}) = \mathbb{E}_{\mathbf{z}' \sim \mathcal{N}_r(\mathbf{z})} \left[\frac{\|\mathbf{V}_1(\mathbf{z}) - \mathbf{V}_1(\mathbf{z}')\|_2}{\|\mathbf{z} - \mathbf{z}'\|_2} \right]$$

High LC $\Rightarrow$ manifold twists violently; traversal is unstable.

Projected High-Frequency Energy (PHFE)

Projects the principal axis into image space and measures its *functional content*:

$$\mathbf{P}_1 = \mathbf{J}_{\text{sub}} \mathbf{V}_1, \qquad \text{PHFE}(\mathbf{z}) = \text{Var}(\nabla^2 \mathbf{P}_1)$$

$\mathbf{P}_1$ as bridge: $\mathbf{P}_1$ is the instantaneous visual change induced by $\mathbf{V}_1$, revealing what the manifold's principal direction actually changes in image space.
HFE vs. PHFE: HFE measures static image detail in $\mathbf{x}$, whereas PHFE measures the high-frequency content of the *principal tangent change* $\mathbf{J}_{\text{sub}} \mathbf{V}_1$.

Diagnostic probe:
$\delta \uparrow +$ PHFE $\uparrow$ $\rightarrow$ curvature *encodes* detail
$\delta \uparrow +$ PHFE $\downarrow$ $\rightarrow$ **Geometric Decoupling** (pathology)

Spectral & Trajectory Metrics

Spectral Isolation Score (SIS)

Grounded in Davis-Kahan perturbation theory:

$$\text{SIS} = \frac{\text{CosSim}(\mathbf{V}_1, \mathbf{v}_1')}{\sum_{k=2}^{P} \text{CosSim}(\mathbf{V}_1, \mathbf{v}_k')}$$

High SIS $\Rightarrow$ *"Tunnel Vision"* geometry; manifold locks onto a single axis.

$$\rho_{\text{dim}}^{(k)} = 1 - \frac{\text{CosSim}(\mathbf{V}_1, \mathbf{v}_k')_{\text{OOD}}}{\text{CosSim}(\mathbf{V}_1, \mathbf{v}_k')_{\text{Normal}}}$$

Higher ρ_{dim} indicates stronger loss of secondary-axis coupling under OOD.

Interpolation Path Metrics

Noise-space SLERP $\mathbf{z}(t) \to \mathbf{h}_c(t_k)$ via diffusion sampler under condition c:

$$L^{(c)} = \sum_k \Delta_k, \quad \tau^{(c)} = \frac{L}{D + \varepsilon}, \quad E^{(c)} = L - D$$

Upper-tail increment $\Delta^{0.95,(c)}$ detects *Geometric Hotspots* — rare, severe local shocks along the trajectory.

Qualitative Visualisation

LC-Map: spatial distribution of manifold curvature. **PHFE-Map:** functional utility of that curvature.
Red hotspots = regions of extreme curvature ("Geometric Hotspots").

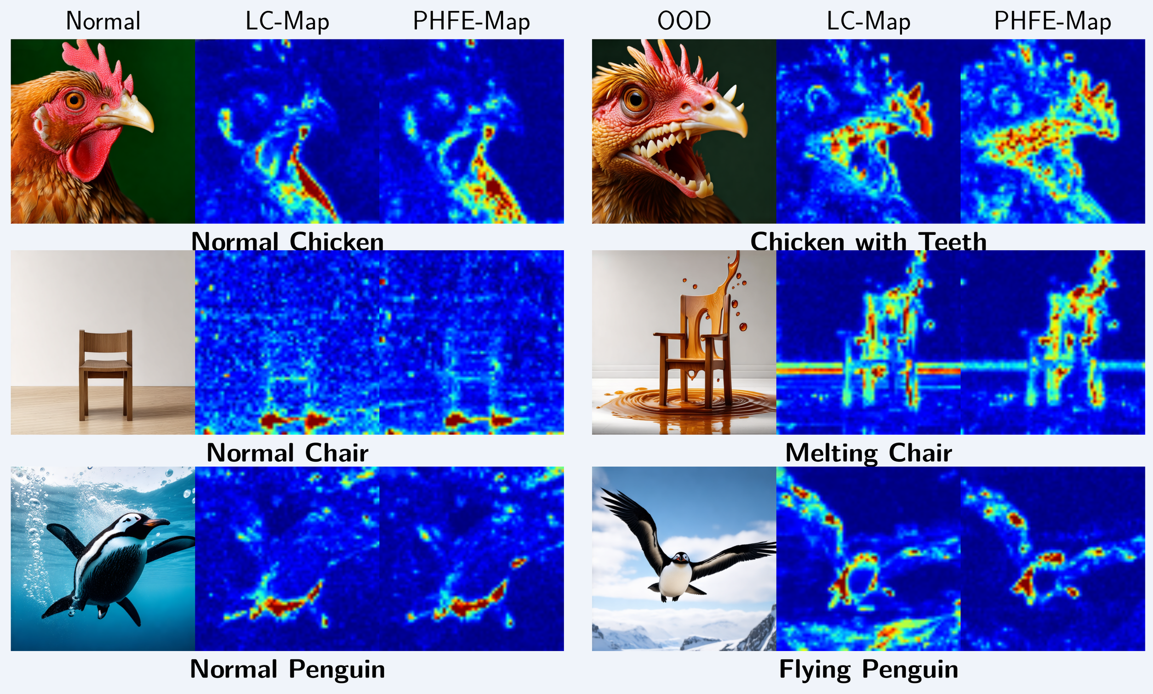

Figure. Normal (left column) and OOD (right column) samples are shown with their LC-Maps and PHFE-Maps. Red LC hotspots spatially align with semantic anomalies, illustrating geometric decoupling when curvature no longer supports useful high-frequency detail.

- **Chicken with teeth:** curvature hotspot at beak only
- **Melting chair:** geometric stress at liquefying structure
- **Flying penguin:** instability mapped to wings

Hotspots *spatially align* with semantic anomalies — confirming that peak geometric budget is spent resolving factual conflicts, not encoding detail.

Experiment Setup

- **Models:** Stable Diffusion 3.5 Medium (SD3.5) and FLUX.1
- **Normal prompts:** Physically realizable descriptions
- **OOD prompts:** Physics violations, material–behavior mismatches, biological chimeras (subjects from COCO object set)
- **Evaluation:** 10 independent sub-sampling runs of $N = 500$ each; Spearman rank correlations reported as mean $\pm$ std

Normal prompt	OOD prompt
A domestic chicken with a smooth, toothless beak …	A chicken with teeth, close-up showing small sharp teeth …
A freshwater fish swimming underwater …	A fish with four muscular legs walking on sand …
A penguin swimming with flippers …	A penguin flying high in the sky with large outstretched wings …

The Correlation Gap

Table 1. Spearman correlations (mean $\pm$ std, n=500, 10 runs)

	ρ(LC, PHFE)	ρ(LS, PHFE)	ρ(LS, LC)
Stable Diffusion 3.5			
Normal	0.413 ± 0.036	0.836 ± 0.016	0.606 ± 0.023
OOD	$\mathbf{0.083 \pm 0.027}$	0.824 ± 0.011	0.296 ± 0.018
FLUX.1			
Normal	0.448 ± 0.014	0.827 ± 0.024	0.612 ± 0.031
OOD	$\mathbf{0.269 \pm 0.021}$	0.780 ± 0.009	0.395 ± 0.078

Key Finding — Geometric Decoupling:
ρ(LC, PHFE) **collapses by** $\Delta\rho \approx -0.33$ **under OOD, while** ρ(LS, PHFE) **remains stable** (≈ 0.82).
$\Rightarrow$ **LS faithfully encodes capacity; LC becomes *non-functional* under semantic stress.**

High-Frequency Transfer Collapse

Table 2. Median HF-transfer statistics (SD3.5)

	PHFE_{lat}	$\text{Var}(\nabla^2 \mathbf{x})$	Top5-HF	Top10-HF
Normal	158.5	0.0120	0.530	0.691
OOD	**252.8**	0.0131	**0.493**	**0.663**

- Latent geometric energy surges $+59.5\%$ under OOD
- Image-space HFE barely changes ($+9\%$)
- **Transfer efficiency** η drops — wasted geometric effort
- High-frequency patterns become *spatially diffuse* (lower Top$k$-HF), consistent with noise/artifact rather than structured detail

Geometric Resource Misallocation: OOD stress amplifies latent curvature expenditure without commensurate perceptual return.

Manifold Rigidity & Dimensional Collapse

Table 3. Spectral coupling profile (SD3.5)

Chair	$(\mathbf{V}_1, \mathbf{v}_2')$	$(\mathbf{V}_1, \mathbf{v}_3')$	$(\mathbf{V}_1, \mathbf{v}_4')$	SIS
Normal	0.140	0.062	0.042	4.91
OOD	0.084	0.040	0.027	**6.45**
ρ_{dim}	0.401	0.355	0.357	$+31\%$

- OOD severs secondary spectral coupling (-17% to -40%)
- SIS rises sharply $\Rightarrow$ *Hyper-Rigid* state
- Manifold degenerates toward a single isolated axis
- Explains "stiffness": OOD generation resists correction and editing

Interpolation Trajectory Analysis

Table 4. Global path metrics (Monte Carlo, 50 trials, n=100 seed pairs)

	L	τ	E
Normal	5.374 ± 0.044	5.695 ± 0.040	4.428 ± 0.039
OOD	6.243 ± 0.033	6.446 ± 0.031	5.295 ± 0.038
Ratio R	**1.183**	**1.150**	**1.219**

Table 5. Extremal trajectory increments

	$\Delta^{0.90}$	$\Delta^{0.95}$	$\Delta^{\max}$
Normal	0.406	0.453	0.530
OOD	0.450	0.496	0.568
Ratio	1.133	1.118	1.110
frac	0.728	0.680	0.568

$\text{frac}(m) = \Pr(m_{\text{OOD}} > m_{\text{Normal}})$

- Cumulative path length $+18.3\%$; excess length $+21.9\%$
- OOD expands and **warps** the traversal space
- Mean increment: $0.259 \to 0.304$ ($+17.4\%$) across *all* steps
- Endpoint distance barely changes ($R_D \approx 1.04$): OOD is not merely "further apart" — it makes the journey more tortuous
- $\text{frac}(\Delta^{0.95}) = 0.68$: the OOD 95th-percentile increment exceeds the normal one in 68% of trials

Causal Nature of Geometric Decoupling

It compared two models with identical architectures but distinct training regimes: Stable Diffusion 3.5 (SD3.5) Base and SD3.5 Turbo.

Table 6. Impact of Training Intervention on Geometric Coupling.

Model	$\rho(\text{LC, PHFE})_{\text{Normal}}$	$\rho(\text{LC, PHFE})_{\text{OOD}}$
SD3.5 Base	0.413	0.083
SD3.5 Turbo	0.342	0.156

The metric's sensitivity to training interventions combined with cross-architecture replication provides strong evidence that geometric decoupling is a fundamental structural mechanism rather than a mere statistical coincidence.

Validated Practical Utility

Testing on 500 Normal and 500 OOD samples using SD3.5, we compared the raw LC metric, the raw LS metric, and our proposed geometric efficiency ratio (LC/PHFE).

Table 7. OOD Detection Performance ($N = 1000$).

Score	AUROC	Interpretation
LS (raw)	0.199	Capacity fails
LC (raw)	0.427	Curvature fails
LC/PHFE (ours)	**0.816**	Geometric succeeds

Core Decoupling Hypothesis Validation: OOD stress does not systematically shift absolute LC values, but it destroys the functional relationship between LC and PHFE.

Conclusions

1. **Geometric Functional Decoupling**: First empirical evidence that LC's coupling with image detail (PHFE) collapses under OOD conditions — while LS remains stable throughout.
2. **Geometric Resource Misallocation**: OOD stress triggers extreme curvature that encodes *no perceptual gain* — wasted on resolving semantic conflicts.
3. **Manifold Rigidity / Dimensional Collapse**: SIS surges and secondary spectral coupling breaks down, locking the latent path into a hyper-rigid, low-dimensional trajectory.
4. **Trajectory Pathology**: Smooth noise-space interpolations map to tortuous, heavy-tailed paths in x_0-latent space under OOD conditions.

Diagnostic Paradigm Shift: LDM instability is not random noise — it is a *structural cost* of semantic generalisation, locatable via Riemannian geometry.

Limitations

- Jacobian approximation is compute-intensive — not yet real-time
- Diagnosis requires *realised* hallucinations; latent semantic intent alone is insufficient

Selected References

1. Rombach *et al.* High-Resolution Image Synthesis with LDMs. *CVPR* 2022.
2. Humayun *et al.* Local Scaling & Local Complexity. *ICLR* 2024.
3. Shao *et al.* Riemannian Geometry in Deep Generative Models. *NeurIPS* 2018.
4. Kwon *et al.* Diffusion-based Image Translation. *CVPR* 2023.
5. Mokady *et al.* Null-Text Inversion. *CVPR* 2023.
6. Davis & Kahan. Rotation of Eigenvectors. *SIAM J. Numer. Anal.* 1970.
7. Esser *et al.* Scaling Rectified Flow Transformers (SD3). *ICML* 2024.