THE GEOMETRY OF TIME-SERIES DIFFUSION: WHY LATENT SPACE DIFFUSION WORKS FOR GENERATION AND IMPUTATION

*Afshin Asadi, Ramin Moghaddass
Department of Industrial & Systems Engineering
University of Miami
Coral Gables, FL 33146, USA
{axa3418,ramin}@miami.edu

ABSTRACT

Diffusion models for time series are often trained in observation space, where autocorrelation and heteroskedasticity induce anisotropic and curved data geometry that complicates score estimation. We argue that the effectiveness of latent diffusion stems from geometric alignment, in which temporal encoders reshape sequences into smoother and more Gaussian-like manifolds on which diffusion is better conditioned. We formalize this view by connecting encoder-induced isotropy to reduced score-field complexity. Empirically, we validate this hypothesis using a geometry diagnostic suite consisting of denoising stability under increasing noise, PCA structure, local isotropy, and finite-difference smoothness. Experiments on two large-scale electricity datasets, ECL and LD2011, consistently exhibit improved conditioning and smoother score behavior under appropriate representation capacity in latent space, providing a principled explanation of when and why latent diffusion is preferable for time-series modeling.

Track: Research

1 INTRODUCTION AND RELATED WORK

Diffusion and score-based models have emerged as powerful generative frameworks for time-series data, enabling applications such as long-horizon synthesis and conditional imputation (Ho et al., 2020; Song et al., 2020). Most existing approaches, however, apply diffusion directly in *observation space*, treating a multivariate temporal window $x \in \mathbb{R}^{T\times d}$ as a point in a high-dimensional Euclidean space. This implicitly assumes that raw time-series data exhibit locally isotropic and approximately Gaussian structure, assumptions that are rarely satisfied in real-world temporal signals.

Time series exhibit strong autocorrelation, nonstationarity, regime changes, and heavy-tailed noise, inducing anisotropic and highly curved data manifolds. Injecting i.i.d. Gaussian noise during the forward diffusion process therefore rapidly destroys temporal structure, forcing the reverse process to learn complex and irregular score fields. Empirically, in our experiments, observation-space diffusion exhibits less stable denoising behavior and weaker uncertainty calibration relative to its latent-space counterpart.

Several diffusion-based methods have been proposed for time-series modeling. TimeGrad (Rasul et al., 2021) and CSDI (Tashiro et al., 2021) adapt diffusion to forecasting and imputation through autoregressive or conditional formulations, while subsequent works explore non-autoregressive conditioning, self-guided diffusion, and architectural refinements (Shen & Kwok, 2023; Kollovieh et al., 2023). Despite their empirical success, these approaches largely retain diffusion in the original observation space and focus on architectural design rather than the suitability of the representation space itself.

A related line of work performs diffusion or score-based modeling in learned latent spaces, including recent latent diffusion approaches for time-series modeling (Qian et al., 2024; Ruan et al., 2025).

*Corresponding Author

Latent diffusion models (Rombach et al., 2022) and latent score-based generative models (Vahdat et al., 2021) demonstrate that representation-space diffusion can simplify score learning and improve efficiency in vision and audio domains. However, these works are primarily empirical and do not provide a principled explanation for *why* latent diffusion is better aligned with the data geometry, particularly for time series where autocorrelation and nonstationarity play a central role.

We argue that the limitations of observation-space diffusion arise from a fundamental *geometric mismatch* between diffusion processes and the intrinsic structure of temporal data. Diffusion is well behaved on manifolds that are locally smooth, approximately linear, and near-Gaussian. While raw time series violate these properties, latent representations learned by temporal encoders often satisfy them. Representation learning compresses temporal structure, filters measurement noise, and induces low-dimensional manifolds where distances reflect dynamical similarity rather than raw scale.

In our approach, time-series windows are first encoded into a latent representation where diffusion is performed, and the resulting latent samples are then decoded back to observation space.

2 Geometric Perspective and Latent Diffusion Framework

Let $X \in \mathbb{R}^{T \times d}$ denote a multivariate time-series window drawn from an unknown data distribution supported near a low-dimensional structured manifold embedded in a high-dimensional observation space. We consider a temporal encoder E that maps sequences to a latent representation $Z = E(X) \in \mathbb{R}^k$, where $k \ll Td$. The image of this mapping defines a latent manifold on which temporal structure is compactly represented. All input features are standardized to zero mean and unit variance using training-set statistics prior to representation learning and diffusion modeling.

Diffusion and score-based models operate by learning the score field $\nabla \log p_t(\cdot)$ of progressively noised data. The complexity of this score field strongly depends on the geometry of the underlying representation space. In observation space, temporal autocorrelation, nonstationarity, and heteroskedastic noise induce highly curved and anisotropic densities, leading to irregular and poorly conditioned score fields and unstable reverse dynamics. In contrast, latent representations learned by temporal encoders are empirically smoother and closer to Gaussian, yielding score fields with reduced curvature and improved conditioning.

Motivated by this geometric perspective, and supported by the theoretical analysis in Appendix A, we perform diffusion directly **in latent space rather than on raw observations.** Our framework consists of three components: (i) a temporal encoder that maps input sequences to a smooth latent manifold, (ii) a diffusion process defined over latent trajectories, and (iii) a decoder that reconstructs or imputes observations from generated latent samples. This formulation decouples diffusion from raw temporal irregularities while preserving the underlying dynamics through the encoder–decoder mapping.

3 Experiments

This section evaluates the proposed latent-space diffusion framework on real-world electricity consumption data in the context of generative modeling and conditional reconstruction. Implementation details and hyperparameter settings are provided in Appendix C. In particular, our models are trained for unconditional window generation and conditional denoising, which serve as building blocks for downstream forecasting and imputation tasks. Rather than focusing solely on end-task prediction accuracy, we emphasize diagnostic experiments that isolate the geometric and dynamical properties governing diffusion behavior. Specifically, we aim to answer the following questions:

- **Q1** Does latent diffusion exhibit improved denoising stability compared to observation-space diffusion under increasing noise levels?
- **Q2** Are score fields in latent space empirically smoother than in observation space?
- **Q3** How does latent capacity affect the trade-off between geometric regularization and representation error?
- **Q4** Does latent diffusion preserve the global structure of temporal load manifolds?

Datasets. We evaluate on two publicly available electricity consumption datasets: ECL (Zhou et al. (2021)), consisting of hourly substation measurements, and LD2011 (Trindade (2015)), containing 15-minute consumption traces from over 300 clients. Both datasets exhibit strong seasonality, cross-variable correlations, and heterogeneous dynamics, making them suitable for geometric analysis. Additional validation on synthetic chaotic dynamical systems is provided in Appendix B.

3.1 Denoising Stability Analysis

To evaluate diffusion conditioning, we corrupt clean samples with Gaussian noise of scale σ and measure the mean squared denoising error. Figure 1 reports validation MSE as a function of σ.

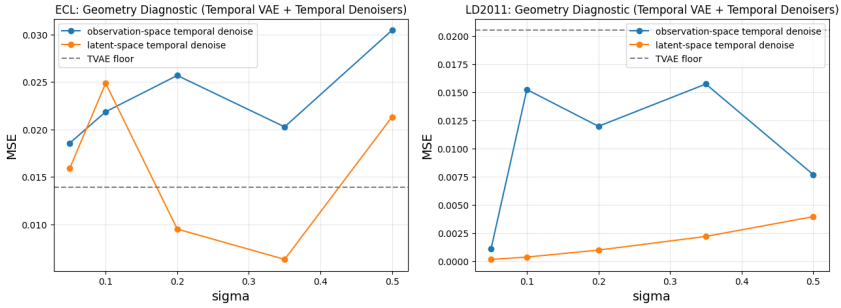

Figure 1: Denoising stability under increasing noise levels. Left: ECL. Right: LD2011.

Across both datasets, observation-space denoising exhibits greater variability across noise levels. In contrast, latent-space denoising remains more stable over moderate noise magnitudes when representation capacity is sufficient, although performance degrades under extreme corruption.

3.2 Global Geometry Visualization

To examine manifold structure, we project observation-space windows and latent encodings onto their first two principal components. As shown in Figure 2, raw windows exhibit pronounced radial anisotropy, hollow-center structure, and directional stretching, reflecting correlated temporal variation and scale imbalance.

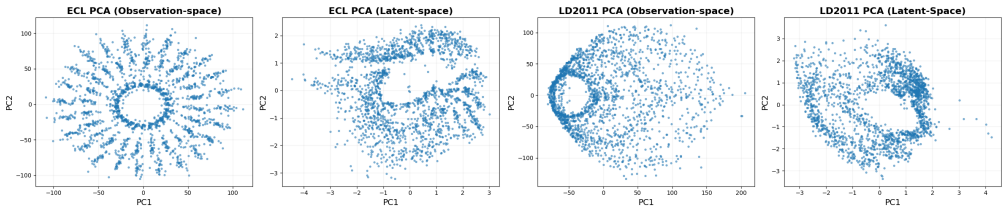

Figure 2: PCA projections of windowed electricity data. Left: ECL. Right: LD2011.

In contrast, latent representations compress variance and reduce extreme anisotropy, yielding more centralized and balanced embeddings. While not strictly Gaussian, the learned representations are better aligned with diffusion assumptions and exhibit smoother global structure.

3.3 Local Geometry and Smoothness

We analyze local geometric regularity using neighborhood covariance conditioning and finite-difference smoothness diagnostics. As shown in Figure 3 (kNN=30, PCA dim=5), latent-space neighborhoods exhibit lower condition numbers than observation-space neighborhoods, indicating smoother and more isotropic geometry that is better aligned with Gaussian diffusion assumptions.

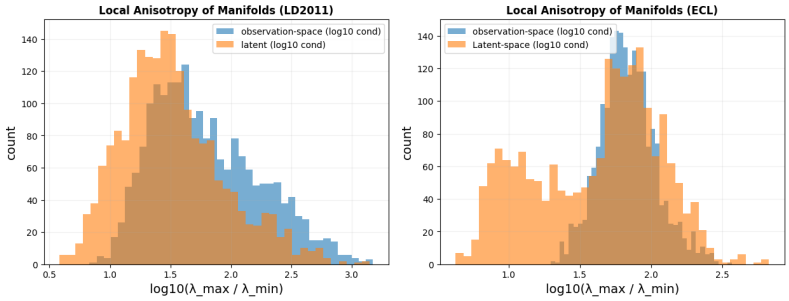

Figure 3: Local isotropy diagnostics based on k-nearest neighbor covariance analysis.

The finite-difference Lipschitz estimates in Figure 4 show that latent denoisers are less sensitive to local perturbations, reflecting smoother and better-conditioned score fields. Together, these results indicate that latent representations regularize both the data manifold and diffusion dynamics.

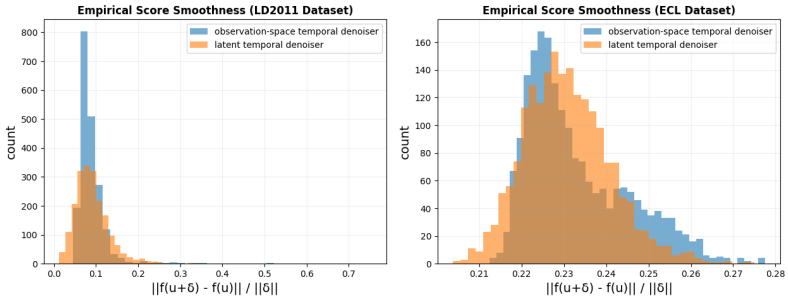

Figure 4: Smoothness diagnostics at $\sigma = 0.35$ based on finite-difference Lipschitz estimates. Distributions of $\|f(u+\delta) - f(u)\|/\|\delta\|$ for obs-space and latent denoisers.

3.4 Latent Capacity and Diffusion Performance

To address Question Q3, we analyze how the latent dimension K affects the balance between representation fidelity and diffusion conditioning. While diffusion is performed in latent space, the resulting latent variables must ultimately be decoded back into observation space where the original data resides. Consequently, the effectiveness of latent diffusion depends not only on the conditioning of the latent manifold but also on the decoder's ability to reconstruct observations from perturbed latent representations. This introduces an important tradeoff between representation capacity and diffusion stability.

Latent dimensionality plays a critical role in balancing geometric regularization and representation accuracy. When the latent dimension K is small, strong regularization toward Gaussian latent distributions can induce posterior collapse and information loss, causing encoder–decoder distortion to dominate and offset geometric benefits. In this regime, latent diffusion provides limited advantage. As the latent dimension increases, the representation becomes more expressive and the reconstruction error decreases, allowing diffusion to operate on a smoother and lower-distortion latent manifold that is better aligned with Gaussian diffusion assumptions.

At the same time, diffusion in latent space does not operate in isolation. After the diffusion process, the perturbed latent variables must be mapped back to observation space through the decoder. If the decoder is highly sensitive to small latent perturbations, errors introduced during diffusion can be amplified during reconstruction. This creates a practical limit on how much increasing latent dimensionality can improve performance, as excessively large latent spaces weaken the regularizing effect of the latent bottleneck.

To examine the role of latent capacity more directly, we perform an ablation on the latent dimension K on two datasets. Table 1 reports the Temporal VAE reconstruction floor, denoising errors in

latent and observation spaces, and two geometry diagnostics: an isotropy score derived from local covariance condition numbers and a smoothness proxy based on an empirical Lipschitz estimate of the denoiser.

Table 1: Latent-dimension ablation on ECL and LD2011.

Dataset	K	TVAE recon floor ↓	Latent denoise MSE ↓	Obs denoise MSE ↓	Latent isotropy score ↑	Latent smoothness proxy ↓
ECL	16	0.01223	0.02469	**0.02019**	0.04349	0.27870
ECL	32	0.02120	0.02000	0.02568	**0.07539**	0.26660
ECL	64	0.01390	**0.01774**	0.02581	0.04212	0.22900
ECL	128	**0.01220**	0.10500	0.02110	0.02918	**0.20340**
LD2011	16	0.02038	0.00315	0.07748	0.10110	0.10639
LD2011	32	0.02050	0.00153	**0.01008**	0.09484	0.08428
LD2011	64	**0.01737**	0.00092	0.02476	0.08236	0.06243
LD2011	128	0.01868	**0.00047**	0.01558	**0.10995**	**0.04260**

The results confirm that latent capacity strongly affects the tradeoff between representation fidelity and diffusion-friendly geometry. When K is small, the latent bottleneck limits reconstruction quality and encoder–decoder distortion reduces the advantage of latent diffusion. As K increases, latent denoising error generally decreases and latent diffusion becomes more favorable relative to observation-space diffusion, indicating that the geometric benefits of latent diffusion emerge once the representation is sufficiently expressive.

However, increasing latent dimensionality does not guarantee monotonic improvements. While larger K typically improves the smoothness of the learned score field and can improve isotropy, excessively large latent spaces weaken the regularizing effect of the latent bottleneck. In such cases, the decoder may become more sensitive to small perturbations in the latent code, so errors introduced during the diffusion process can be amplified when mapping back to the observation space.

These results highlight that latent dimensionality is not merely a modeling choice, but a critical factor governing the geometric conditioning of diffusion. This behavior can be interpreted as a bias–variance tradeoff in representation space: small latent dimensions over-regularize and distort the data manifold, while excessively large latent spaces reduce geometric regularization and degrade diffusion conditioning. Optimal performance therefore arises in an intermediate regime where the latent space is expressive enough to preserve structure yet sufficiently constrained to maintain diffusion-friendly geometry.

4 Discussion

Across two large-scale real-world electricity datasets, we show that latent diffusion achieves improved denoising stability, geometric isotropy, and score smoothness when representation capacity is sufficient. Additional validation on chaotic dynamical systems is provided in Appendix B. By combining theoretical analysis with empirical diagnostics, we provide a geometric explanation for why diffusion in latent space outperforms observation-space modeling. Our results highlight the role of well-conditioned, low-distortion representations in shaping diffusion dynamics. When these conditions are met, latent diffusion induces smoother probability landscapes and more stable reverse-time dynamics, improving robustness under noise. At the same time, strong regularization toward diffusion-friendly latent geometry can introduce representation loss and posterior collapse, degrading reconstruction fidelity. Balancing geometric regularization and expressive capacity therefore remains a key challenge. Future work will explore adaptive representation learning strategies that jointly optimize conditioning and reconstruction accuracy, as well as extensions to multiscale and long-horizon forecasting. Overall, this work establishes geometric regularity of learned representations as a key factor governing diffusion performance in time-series data.

Acknowledgments

This material is based on work supported by the National Science Foundation under Grant No. 2544544.

REFERENCES

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Marcel Kollovieh, Abdul Fatir Ansari, Michael Bohlke-Schneider, Jasper Zschiegner, Hao Wang, and Yuyang Bernie Wang. Predict, refine, synthesize: Self-guiding diffusion models for probabilistic time series forecasting. *Advances in Neural Information Processing Systems*, 36:28341–28364, 2023.

Jian Qian, Bingyu Xie, Biao Wan, Minhao Li, Miao Sun, and Patrick Yin Chiang. Timeldm: Latent diffusion model for unconditional time series generation. *arXiv preprint arXiv:2407.04211*, 2024.

Kashif Rasul, Calvin Seward, Ingmar Schuster, and Roland Vollgraf. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In *International conference on machine learning*, pp. 8857–8868. PMLR, 2021.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Weilin Ruan, Siru Zhong, Haomin Wen, and Yuxuan Liang. Vision-enhanced time series forecasting via latent diffusion models. *arXiv preprint arXiv:2502.14887*, 2025.

Lifeng Shen and James Kwok. Non-autoregressive conditional diffusion models for time series prediction. In *International Conference on Machine Learning*, pp. 31016–31029. PMLR, 2023.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. Csdi: Conditional score-based diffusion models for probabilistic time series imputation. *Advances in neural information processing systems*, 34:24804–24816, 2021.

Artur Trindade. ElectricityLoadDiagrams20112014. UCI Machine Learning Repository, 2015. DOI: https://doi.org/10.24432/C58C86.

Arash Vahdat, Karsten Kreis, and Jan Kautz. Score-based generative modeling in latent space. *Advances in neural information processing systems*, 34:11287–11302, 2021.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Conference*, volume 35,12, pp. 11106–11115. AAAI Press, 2021.

A Geometric Perspective on Latent Diffusion

This appendix provides a preliminary theoretical perspective supporting the geometric arguments developed in the main paper. Our goal is not to present a complete formal theory of latent diffusion for time series, but rather to clarify the mathematical mechanisms through which representation learning can improve diffusion conditioning.

A.1 Setup and Notation

Let $X \in \mathbb{R}^{T\times d}$ denote a multivariate time-series window and let $D = Td$ be its ambient dimension after vectorization. We assume that the data distribution P_X is supported in a small neighborhood of a compact, smooth manifold $\mathcal{M}_X \subset \mathbb{R}^D$ and admits a twice continuously differentiable density p_X.

Let $E : U \subset \mathbb{R}^D \to \mathbb{R}^k$ be a differentiable temporal encoder defined on a neighborhood U of $\mathcal{M}_X$, and define the latent representation

$$Z = E(X) \in \mathbb{R}^k, \quad k \ll D.$$

For stochastic encoders such as VAEs, we consider the mean mapping $E(X) = \mathbb{E}[Z|X]$.

The image $\mathcal{M}_Z = E(\mathcal{M}_X)$ defines a latent representation manifold in $\mathbb{R}^k$. Under the regularity conditions of Assumption A.1, the pushforward distribution $P_Z = E_\# P_X$ admits a twice continuously differentiable density p_Z.

We consider diffusion processes in both observation and latent spaces, and denote by $p_{X,t}$ and $p_{Z,t}$ the corresponding marginal densities at diffusion time t.

A.2 Regularity of the Encoder

Our analysis relies on mild regularity conditions on the encoder.

Assumption A.1 (Bi-Lipschitz Regularity). The encoder E is continuously differentiable on a neighborhood of $\mathcal{M}_X$ and bi-Lipschitz on $\mathcal{M}_X$. That is, there exist constants $0 < m \le L < \infty$ such that

$$m\|x_1 - x_2\| \le \|E(x_1) - E(x_2)\| \le L\|x_1 - x_2\|, \quad \forall x_1, x_2 \in \mathcal{M}_X.$$

While strong, this assumption is approximately satisfied in practice by well-trained representation models that avoid posterior collapse and maintain well-conditioned Jacobians in regions of high data density. In intuitive terms, Assumption A.1 requires that the encoder neither collapses distinct time-series windows into nearly identical latent representations nor excessively amplifies small perturbations. It ensures that distances and local geometry on the data manifold are preserved up to constant factors, yielding well-conditioned and approximately invertible representations.

A.3 Probability Flow ODE

Following Song et al. (2020), the probability flow ODE associated with a diffusion process satisfies

$$\frac{dX_t}{dt} = f_X(t, X_t) - \frac{1}{2}g(t)^2 s_{X,t}(X_t), \tag{1}$$

$$\frac{dZ_t}{dt} = f_Z(t, Z_t) - \frac{1}{2}g(t)^2 s_{Z,t}(Z_t), \tag{2}$$

where f_X and f_Z denote the drift functions in observation and latent spaces, respectively, and

$$s_{X,t}(x) = \nabla_x \log p_{X,t}(x), \quad s_{Z,t}(z) = \nabla_z \log p_{Z,t}(z)$$

denote the corresponding score fields.

Under standard regularity conditions ensuring well-posedness, learning a diffusion model corresponds to approximating these score functions.

A.4 Score Curvature and Conditioning

We quantify the local geometric complexity of a density through the curvature of its log-density.

Definition A.2 (Score Curvature). For a twice continuously differentiable density p on $\mathbb{R}^d$, define

$$\kappa_p(x) = \|\nabla_x^2 \log p(x)\|_{\mathrm{op}}, \tag{3}$$

and the global curvature

$$K(p) = \sup_{x \in \mathrm{supp}(p)} \kappa_p(x). \tag{4}$$

Large values of $K(p)$ correspond to rapidly varying score fields and ill-conditioned reverse dynamics, leading to stiff probability flow ODEs and increased sensitivity to perturbations.

Intuitively, $\kappa_p(x)$ measures how abruptly the denoising directions change in a neighborhood of x, reflecting the local roughness of the underlying probability landscape. Smaller values correspond to smoother score fields and more stable reverse-time dynamics.

A.5 Gaussian Reference Case

The Gaussian distribution provides a minimal-complexity baseline for score-based diffusion.

Lemma A.3. Let $p(x) = \mathcal{N}(x|\mu, \Sigma)$ on $\mathbb{R}^d$ with $\Sigma \succ 0$. Then

$$\nabla_x \log p(x) = -\Sigma^{-1}(x - \mu), \tag{5}$$

$$\nabla_x^2 \log p(x) = -\Sigma^{-1}, \tag{6}$$

and $K(p) = \|\Sigma^{-1}\|_{\mathrm{op}}$.

Proof. Direct differentiation. □

This result shows that Gaussian distributions induce globally linear score fields and constant curvature, yielding optimally conditioned reverse-time dynamics.

A.6 Transformation of Score Fields

We relate observation-space and latent-space scores via a change of variables.

Lemma A.4 (Score Transformation). Assume that E is a diffeomorphism between $\mathcal{M}_X$ and $\mathcal{M}_Z$, with Jacobian J_E and inverse E^{-1}. Then

$$s_Z(z) = J_{E^{-1}}(z)^\top s_X(x) + \nabla_z \log |\det J_{E^{-1}}(z)|, \qquad x = E^{-1}(z), \tag{7}$$

for all $z \in \mathcal{M}_Z$.

Proof. By the change-of-variables formula, $p_Z(z) = p_X(x)|\det J_{E^{-1}}(z)|$ with $x = E^{-1}(z)$. Taking logarithms and differentiating with respect to z yields the result. □

A.7 Implications for Score Conditioning

Lemma A.4 implies that the geometry of the encoder directly influences the regularity of the latent score field. In particular, if the Jacobian J_E and its inverse are well conditioned, then variations in s_X are not strongly amplified in latent space.

Proposition A.5 (Conditioning Transfer). Suppose E satisfies Assumption A.1 and has uniformly bounded second derivatives on a neighborhood of $\mathcal{M}_X$. Then there exists a constant $C > 0$, depending only on these bounds, such that

$$\|\nabla s_Z(z)\|_{\mathrm{op}} \leq \frac{L}{m}\|\nabla s_X(x)\|_{\mathrm{op}} + C, \qquad x = E^{-1}(z). \tag{8}$$

Proof Sketch. The result follows by differentiating Lemma A.4 using the chain rule and applying operator norm bounds to the Jacobian and Hessian terms. □

This bound indicates that encoders with well-conditioned Jacobians limit the curvature of the latent score field.

A.8 Stability of Reverse-Time Dynamics

Let L_s^X and L_s^Z denote Lipschitz constants of the observation-space and latent-space score fields, respectively.

Proposition A.6 (ODE Stability). Consider two solutions of the probability flow ODE in latent space with initial conditions z_0 and $\tilde{z}_0$. If $L_s^Z \leq \alpha L_s^X$ for some $\alpha < 1$, then perturbations grow at a reduced exponential rate:

$$\|z_t - \tilde{z}_t\| \leq e^{L_s^Z t}\|z_0 - \tilde{z}_0\|, \tag{9}$$

and hence more slowly than in observation space.

Proof. This follows from standard Grönwall inequalities for ordinary differential equations. □

A.9 Connection to Empirical Diagnostics and Limitations

The bounds above suggest that encoders with approximately isotropic and well-conditioned Jacobians reduce score curvature and improve stability by controlling $K(p_Z)$ and the Lipschitz constant of the latent score field. Our empirical diagnostics in Section 3 (denoising stability, local isotropy, and finite-difference smoothness) provide quantitative evidence supporting this prediction on real-world datasets.

We emphasize that this analysis is preliminary. Establishing sharp, non-asymptotic bounds on score complexity under learned representations remains an open problem, and our results should therefore be viewed as theoretical motivation consistent with, rather than replacing, empirical validation.

B TOY DYNAMICAL SYSTEM VALIDATION

To further validate the geometric effects of latent diffusion, we evaluate the framework on trajectories generated from two chaotic dynamical systems: the Lorenz system and the Rössler system. These systems produce trajectories that lie on low-dimensional nonlinear manifolds with well-studied geometric structure, making them suitable toy benchmarks. For consistency, we use the same model architecture and hyperparameters as in the real-world experiments.

Consistent with our observations on electricity consumption datasets, latent representations exhibit reduced local anisotropy and smoother score fields compared to observation-space representations.

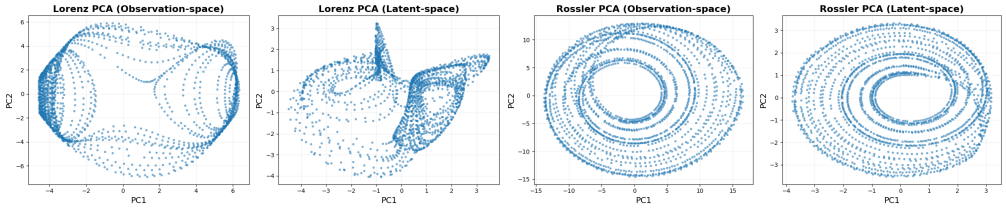

Figure 5: PCA projections of Lorenz and Rössler trajectories in observation and latent spaces.

Figure 5 visualizes the geometry of the two chaotic attractors. In observation space, both systems exhibit highly curved manifolds reflecting their nonlinear dynamics. After encoding, the latent representations preserve the underlying structure while producing a more compact and geometrically regular manifold.

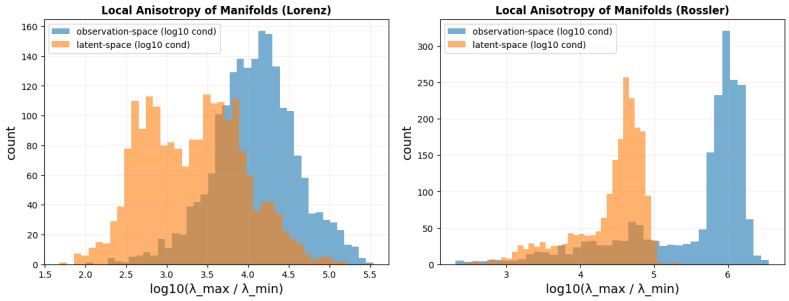

Figure 6: Local anisotropy measured via condition numbers of local covariance matrices.

Figure 6 quantifies the geometric conditioning of the manifolds. The latent representations exhibit significantly smaller condition numbers, indicating improved local isotropy relative to observation space.

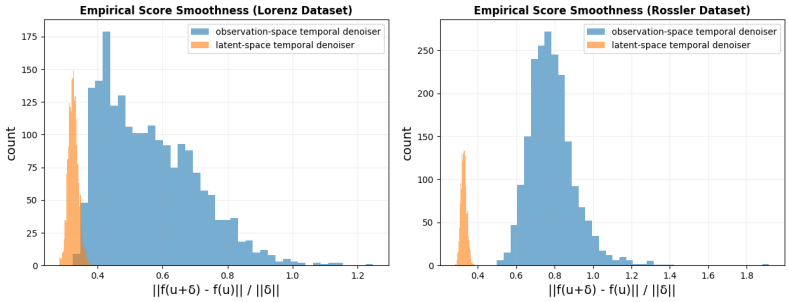

Figure 7: Empirical score smoothness measured via a Lipschitz proxy.

Finally, Figure 7 compares the smoothness of the learned score fields. Latent-space denoisers produce substantially smoother score fields than observation-space denoisers for both dynamical systems.

These results demonstrate that the geometric advantages of latent diffusion persist even in controlled dynamical systems with well-understood attractor structure.

C Experimental Details

Experiments are conducted on two real-world electricity consumption datasets: the Electricity Consumption Load (ECL) dataset and the Electricity Load Diagrams 2011–2014 (LD2011) dataset. Both datasets contain multivariate hourly electricity usage time series from multiple customers.

All multivariate time series are segmented into overlapping windows of length $T = 48$ with stride 1 and globally standardized using training statistics. The datasets are split temporally into 80% training and 20% validation sets.

Latent representations are learned using a temporal β-VAE based on single-layer LSTM encoders and decoders with hidden size 512 and latent dimension $K = 64$. The VAE is trained for 80 epochs using Adam with learning rate 10^{-3} and KL weight $\beta = 10^{-3}$.

Observation-space and latent-space denoisers are implemented as temporal Transformer encoders with identical architectures, consisting of 4 layers, 8 attention heads, and hidden dimension 256. Noise-level conditioning is applied through learned σ embeddings.

All models are trained using Adam with learning rate 10^{-3} and batch size 256. For each noise level $\sigma \in \{0.05, 0.1, 0.2, 0.35, 0.5\}$, denoisers are trained for 15 epochs to minimize mean squared error between clean and noisy sequences, with parameters reinitialized for each run.

Representation geometry is evaluated using three diagnostics: (i) validation mean squared denoising error, with the posterior-mean VAE reconstruction error used as a reference floor; (ii) local anisotropy measured by condition numbers of k-nearest-neighbor covariance matrices after PCA reduction ($k = 30$, $r = 5$); and (iii) score-field smoothness estimated using a finite-difference Lipschitz proxy with perturbation scale $\delta = 0.02$ and noise level $\sigma = 0.35$.

All experiments are implemented in PyTorch with fixed random seeds and executed on NVIDIA A100 GPUs (Google Colab environment).