Random Projection Flows for Efficient Manifold Density Estimation

Ahmad Ayaz Amin
Department of Computer Science
Toronto Metropolitan University
Toronto, Canada
ahmadayaz.amin@torontomu.ca

Baha Uddin Kazi
Faculty of Applied Science and Technology
Humber Polytechnic
Toronto, Canada
bahauddin.kazi@humber.ca

Abstract

Accurate density estimation is crucial for understanding complex high-dimensional data, but it becomes challenging when the data lies on or near low-dimensional manifolds. Random projections provide a natural way to reduce dimensionality while approximately preserving geometric structure, enabling effective density estimation in these settings. We introduce *Random Projection Flows* (RPFs), a principled framework for injective normalizing flows that leverages tools from random matrix theory and the geometry of random projections. RPFs employ random semi-orthogonal matrices, drawn from Haar-distributed orthogonal ensembles via QR decomposition of Gaussian matrices, to project data into lower-dimensional latent spaces for the base distribution. Unlike principal component analysis flows or learned injective maps, RPFs are plug-and-play, efficient, and yield closed-form expressions for the Riemannian volume correction term. We demonstrate that RPFs are both theoretically grounded and practically effective, providing a strong baseline for generative modeling and a bridge between random projection theory and normalizing flows.

1 Introduction

Normalizing flows [1, 2] are a class of generative models that provide exact likelihood evaluation and flexible density estimation by transforming a simple base distribution (typically Gaussian) into a complex data distribution via a sequence of invertible transformations. Formally, given a data sample $\mathbf{x} \in \mathbb{R}^D$ and an invertible transformation f_θ with parameters θ, the change-of-variables formula gives the log-likelihood:

$$\log p_\theta(\mathbf{x}) = \log p_Z(f_\theta(\mathbf{x})) + \log \left| \det \frac{\partial f_\theta(\mathbf{x})}{\partial \mathbf{x}} \right| \tag{1}$$

where p_Z is the density of the base distribution and the second term is the log determinant of the Jacobian.

Although fully invertible flows can be made arbitrarily flexible [3], high-dimensional data often motivate the use of *injective flows*, which map data into a lower-dimensional latent space $d < D$. For injective flows [4, 5], the log-likelihood must account for the dimensionality reduction:

$$\log p_\theta(\mathbf{x}) = \log p_Z(f_\theta(\mathbf{x})) - \frac{1}{2} \log \left| \det \left(J_\theta(f_\theta(\mathbf{x}))^\top J_\theta(f_\theta(\mathbf{x})) \right) \right| \tag{2}$$

where $J_\theta(\mathbf{x})$ is the $D \times d$ Jacobian of the inverse of the injective transformation. A central difficulty is computing the Riemannian volume element, as it involves determinants and matrix inverses.

Although dimensionality reduction for density estimation via random projections is a classic idea, the design of random projections explicitly tailored for density estimation back in the ambient space,

Accepted at third SPIGM workshop @ NeurIPS 2025.

with guarantees of correctness and scalability, has not been deeply explored. Our approach makes this explicit: rather than learning a subspace, we fix a Haar-random projection [6]. Thanks to the semi-orthogonal nature of the projection matrix, the corresponding Jacobian volume term has a closed form. An optional constant scaling factor motivated by Johnson–Lindenstrauss (JL) theory [7] can be applied to make projected norms unbiased in expectation, but this is not required for correctness.

Importantly, the JL-inspired scaling is best viewed as a convenient add-on: one could omit it (yielding a strictly isometric map onto the random subspace), or even replace it with a learned scale as part of the model. We show empirically that including the JL-motivated constant often improves likelihood calibration and density estimates compared to PCA flows, but the random projection flow itself remains valid regardless.

Our contributions are:

1. We propose Random Projection Flows (RPFs), a class of injective flows using Haar-distributed semi-orthogonal projections with a simple, optional JL-motivated volume correction.
2. We analyze this correction: although not necessary, it provides a constant closed-form Jacobian term and can improve calibration; it could also be dropped or learned.
3. We position RPFs conceptually between two-stage density estimators (fixed encoder + separate density model) and fully end-to-end latent-variable models, enjoying benefits of both while avoiding manifold overfitting common in VAEs.
4. We demonstrate that RPFs are composable: they can be used standalone (e.g. with a GMM) or as building blocks inside arbitrary normalizing flow or SurVAE [8] architectures while preserving exact likelihoods.
5. Empirically, RPFs outperform PCA-based injective flows on UC Irvine (UCI) benchmarks and preserve manifold geometry in synthetic experiments.

2 Random Projections and Random Matrix Theory

2.1 Haar Orthogonal Matrices from Gaussian QR

A standard method to generate a Haar-distributed orthogonal matrix $Q \in \mathbb{R}^{D\times D}$ is to sample a Gaussian matrix G with i.i.d. $\mathcal{N}(0,1)$ entries and apply QR decomposition $G = QR$. The orthogonal factor Q is Haar distributed [6]. Taking the first d rows of Q yields a semi-orthogonal matrix $V \in \mathbb{R}^{d\times D}$ with $VV^\top = I_d$. This induces a uniform distribution over d-dimensional subspaces (the Grassmannian).

2.2 The Johnson–Lindenstrauss (JL) Lemma

The JL lemma [7] provides a theoretical foundation for dimension reduction via random linear projections. It states that a set of N points $\{x_i\} \subset \mathbb{R}^D$ can be embedded into $\mathbb{R}^d$ with $d = O(\epsilon^{-2}\log N)$ such that all pairwise distances are approximately preserved:

$$(1-\epsilon)\|x_i - x_j\|_2 \le \|Rx_i - Rx_j\|_2 \le (1+\epsilon)\|x_i - x_j\|_2, \quad \forall i, j. \tag{3}$$

Gaussian and Haar-distributed semi-orthogonal matrices satisfy this property with high probability. These embeddings are approximate isometries, with a scaling factor that concentrates tightly; this underpins our choice of volume correction in RPFs.

2.3 Singular Value Concentration

The singular values of random rectangular Gaussian matrices follow the Marchenko–Pastur law [9], implying concentration of $\|Rx\|_2^2$ around its expectation. This explains both the JL distance-preservation guarantees and why the Jacobian volume term in RPFs reduces to a simple constant.

3 Random Projection Flows

We introduce *Random Projection Flows (RPF)*, an efficient class of injective flows that leverage random linear projections to compress high-dimensional data into a lower-dimensional latent space, followed by a tractable latent density model such as a Gaussian mixture model (GMM).

3.1 Random Linear Projection and Volume Element

Let $\mathbf{x} \in \mathbb{R}^D$ and $V \in \mathbb{R}^{d \times D}$ be Haar semi-orthogonal with $VV^\top = I_d$. We define a scaled projection:

$$W = \sqrt{\frac{D}{d}}\, V, \qquad \mathbf{z} = W\mathbf{x}. \tag{4}$$

This JL-inspired scaling keeps norms approximately unbiased. The Jacobian is $J = W$ and the Riemannian metric is

$$J^\top J = W^\top W = \frac{D}{d} V^\top V. \tag{5}$$

$V^\top V$ projects onto a d-dimensional subspace with d eigenvalues 1 and $D - d$ zeros. Hence

$$\det(J^\top J) = \left(\frac{D}{d}\right)^d, \quad \sqrt{\det(J^\top J)} = \left(\frac{D}{d}\right)^{d/2}, \tag{6}$$

yielding a constant log-volume correction:

$$\log \sqrt{\det(J^\top J)} = \frac{d}{2} \log \frac{D}{d}. \tag{7}$$

This constant appears with opposite signs when encoding versus decoding, reflecting the symmetry of the change-of-variables formula. One could drop it (making the map strictly isometric) or even learn the scale along with the shift, resulting in an injective analogue of the affine flow in RealNVP [10] and Glow [11]. We do not pursue the latter, leaving that for future work, but we do show that the JL-derived constant can yield better results than the fully isometric model in certain scenarios (and vice versa).

3.2 Latent Density Modeling and Positioning

In the latent space, we fit a tractable density such as a GMM:

$$p_Z(\mathbf{z}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{z}; \mu_k, \Sigma_k).$$

Then, the data log-likelihood is

$$\log p_X(\mathbf{x}) = \log p_Z(W\mathbf{x}) + \frac{d}{2} \log \frac{D}{d}. \tag{8}$$

RPFs thus sit conceptually between classic two-stage methods (PCA+GMM) and end-to-end latent-variable models. Like the former, the projection is fixed; like the latter, the whole map is differentiable with exact likelihood. RPFs avoid manifold overfitting [12] and post-hoc density correction [13] often needed in variational autoencoders [14] and other maximum likelihood models with manifold latent dimensionality.

Furthermore, because the projection has a closed-form volume element, RPF layers and their trainable counterparts can be dropped into arbitrary normalizing flow or SurVAE [8] architectures while preserving exact likelihoods.

3.3 Properties

The key properties of RPFs are:

- **Injectivity:** W is injective almost surely.
- **Constant volume term:** additive constant, independent of x.

- **JL guarantees:** high-probability distance preservation; scaling matches expected norms.
- **Haar invariance:** rotationally unbiased.

Although having many benefits, RPFs do have a key disadvantage in being unable to scale to very complex datasets, which we show in the CIFAR-10 experiment.

4 Related Works

Normalizing flows [1, 2] provide invertible maps with tractable likelihoods. Injective extensions [4, 5] handle dimensionality reduction but incur expensive Riemannian volume terms. PCA flows [15] exploit semi-orthogonal projections learned from data. RPFs generalize these by using random Haar projections: unbiased across subspaces and with closed-form volume correction.

Random projection theory underpins RPFs. The JL lemma [7, 16, 17, 18] shows Gaussian and sub-Gaussian maps preserve distances. Haar matrices from Gaussian QR [19] have well-characterized spectra [20, 21]. Concentration results [22] explain the near-isometry.

Structured orthogonal embeddings further motivate RPFs. Orthogonal Random Features [23] reduce kernel approximation variance using random orthogonal matrices. Felix et al. [24] demonstrate the "unreasonable effectiveness" of structured embeddings (e.g. Hadamard), achieving JL-like guarantees at lower cost. These lines of work show random orthogonal maps are powerful primitives across ML; RPFs leverage them for generative modeling with exact likelihoods.

5 Experiments

5.1 UCI Density Esimation

We evaluate RPFs on the UCI density estimation benchmarks introduced in [25], including `POWER`, `GAS`, `HEPMASS`, and `MINIBOONE`. These datasets are widely used to assess the quality of density estimators under moderate-dimensional structured data.

5.1.1 Setup

Following standard practice, we preprocess the datasets using the protocol of [25] and split them into training, validation, and test sets.

We compare RPFs against flows constructed with PCA projections [15], which serve as a strong baseline for injective flow layers. Each method is combined with a Gaussian mixture model (GMM) base density. For fair comparison, we train all models under identical optimization settings, including the number of manifold dimensions and the number of Gaussian mixture components, and report test log likelihoods. The results are shown in Table 1.

5.1.2 Results

Table 1 reports test log-likelihoods on the UCI benchmark datasets for three models: JL-scaled Random Projection Flows (JL), isometric RPFs without the JL scaling factor (ISO), and PCA-based injective flows.

Both RPF variants clearly outperform PCA across all datasets, often by several nats, despite using data-independent random projections. This highlights their competitiveness as low-cost, data-agnostic alternatives. The JL scaling offers a small, dataset-dependent benefit, but is not essential as ISO performs similarly well. The key advantage is that even fixed random projections provide strong density estimates without the need for learned projections as in PCA.

5.2 Projection Comparison

In addition to quantitative evaluation on synthetic 2D datasets, we investigated how RPFs and PCA behave on higher-dimensional data. Specifically, we examined three benchmark 3D datasets from scikit-learn [26]: the Swiss roll, the S-curve, and clustered blobs.

Table 1: Density estimation results (test log-likelihood) on UCI datasets. JL denotes RPFs with Johnson–Lindenstrauss-inspired scaling, ISO denotes isometric RPFs (no scaling), and PCA denotes PCA-based injective flows. Higher is better.

Model	POWER	GAS	HEPMASS	MINIBOONE
RPF (JL)	-1.72 ± 0.14	-1.57 ± 0.37	-20.08 ± 0.14	-14.68 ± 0.30
RPF (ISO)	-1.99 ± 0.21	-1.40 ± 0.04	-19.97 ± 0.20	-14.63 ± 0.47
PCA Flow [15]	-2.51 ± 0.13	-2.32 ± 0.04	-20.71 ± 0.38	-20.66 ± 0.05

5.2.1 Setup

For each dataset, we first display the true distribution in 3D, with a color gradient assigned according to one coordinate axis. This coloring allows us to track local structure after projection. We then visualize the corresponding 2D embeddings produced by PCA and RPF. The results are shown in Figure 1.

5.2.2 Results

Swiss roll PCA flattens the manifold into a nearly linear band, losing the roll's spiral geometry. In contrast, the RPF retains much of the nonlinear curvature, producing a projection that still resembles the original manifold.

S-curve PCA reduces the manifold to a simple arc, obscuring the double-banded shape. RPF better preserves the two-layered structure, suggesting a stronger correspondence with the underlying geometry.

Blobs Both methods maintain separability between clusters, but PCA aligns the blobs along its principal axes, while RPF introduces greater variability in density and orientation, highlighting differences in local structure.

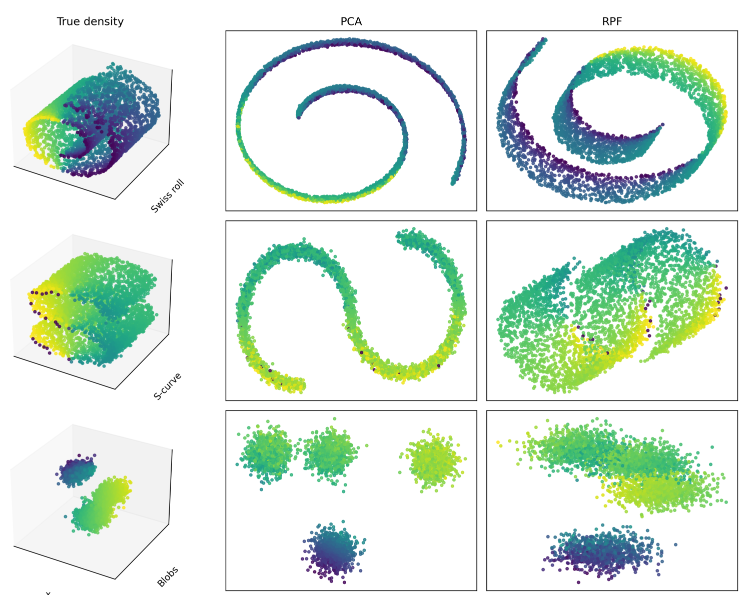

Figure 1: Projections of 3D benchmark datasets. Left: true 3D density with color gradients. Middle: PCA projection. Right: RPF projection.

Implications for density estimation. These projections illustrate why RPFs outperform PCA in UCI data. PCA collapses nonlinear manifold features by aligning with global variance, as seen in the Swiss roll and S-curve. Random orthogonal projections instead preserve local geometry more uniformly in expectation, leading to embeddings that better respect intrinsic structure — crucial for accurate likelihood estimation.

5.3 MNIST and CIFAR-10 Density Estimation

5.3.1 Setup

We perform unconditional density estimation on the MNIST and CIFAR-10 datasets, following the preprocessing methodology of [25]. We report log-likelihoods in logit space and compare to the results presented in [25].

We train a Random Projection Flow (RPF) with JL scaling, using a Gaussian Restricted Boltzmann Machine (GRBM [27]) prior trained via persistent contrastive divergence (PCD [28]). Log-likelihoods are estimated using annealed importance sampling (AIS [29]) to approximate the partition function.

5.3.2 Results

As shown in Table 2, the RPF model substantially outperforms MADE [30] on MNIST. On CIFAR-10, however, RPF performs worse than MAF [25] and the Gaussian model. This may reflect the difficulty of modeling projected densities with a GRBM for complex, high-dimensional natural images. While MADE also underperforms the Gaussian model, it suggests that RPFs combined with GRBMs are effective for simpler datasets like MNIST but face limitations on CIFAR-10, likely due to the need for more expressive latent models or deeper architectures.

Table 2: Unconditional density estimation results (test log-likelihood) on MNIST and CIFAR-10. RPF denotes RPFs with JL scaling. Higher is better; best-performing model is highlighted in **bold**.

Model	MNIST	CIFAR-10
RPF (ours)	$\mathbf{-149.7}$	-4885
Gaussian	-1366.9	2367
MADE MoG	-1038.5	-397
MAF (10)	-1313.1	$\mathbf{3049}$

6 Conclusion

We introduced *Random Projection Flows (RPFs)*, a simple and efficient class of injective normalizing flows that leverage Haar-distributed semi-orthogonal projections to compress high-dimensional data into a lower-dimensional latent space. By exploiting the constant volume element of these projections, RPFs provide tractable likelihood evaluation without per-sample determinant computations.

Empirically, RPFs achieve competitive density estimation performance on UCI benchmarks and synthetic manifolds, often outperforming PCA-based injective flows despite being entirely data-independent. Our experiments highlight that the geometry preserved by random projections can improve density estimation, though generation tasks—particularly on high-dimensional natural images like CIFAR-10—remain challenging due to the limitations of simple latent models.

Overall, RPFs offer a low-cost, composable, and theoretically grounded alternative for injective flow modeling, providing a promising building block for future latent-variable and hybrid generative architectures. Future work could explore more expressive latent densities, deeper projections, or hybrid methods to improve generation performance on complex datasets.

References

[1] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 1530–1538, Lille, France, 07–09 Jul 2015. PMLR.

[2] Laurent Dinh, David Krueger, and Yoshua Bengio. NICE: non-linear independent components estimation. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Workshop Track Proceedings*, 2015.

[3] Will Grathwohl, Ricky T. Q. Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *International Conference on Learning Representations*, 2019.

[4] Mevlana C. Gemici, Danilo Rezende, and Shakir Mohamed. Normalizing flows on riemannian manifolds, 2016.

[5] Anthony L. Caterini, Gabriel Loaiza-Ganem, Geoff Pleiss, and John P. Cunningham. Rectangular flows for manifold learning. In *none*, 2021.

[6] Francesco Mezzadri. How to generate random matrices from the classical compact groups, 2007.

[7] William B. Johnson and Joram Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. In Richard Beals, Anatole Beck, Alexandra Bellow, and et al., editors, *Conference in Modern Analysis and Probability (New Haven, Conn., 1982)*, volume 26 of *Contemporary Mathematics*, pages 189–206. American Mathematical Society, Providence, RI, 1984.

[8] Didrik Nielsen, Priyank Jaini, Emiel Hoogeboom, Ole Winther, and Max Welling. Survae flows: Surjections to bridge the gap between vaes and flows. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 12685–12696. Curran Associates, Inc., 2020.

[9] V A Marčenko and L A Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, apr 1967.

[10] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real NVP. In *International Conference on Learning Representations*, 2017.

[11] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[12] Gabriel Loaiza-Ganem, Brendan Leigh Ross, Jesse C Cresswell, and Anthony L. Caterini. Diagnosing and fixing manifold overfitting in deep generative models. *Transactions on Machine Learning Research*, 2022. Expert Certification.

[13] Bin Dai and David Wipf. Diagnosing and enhancing VAE models. In *International Conference on Learning Representations*, 2019.

[14] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *CoRR*, abs/1312.6114, 2013.

[15] Eike Cramer, Alexander Mitsos, Raul Tempone, and Manuel Dahmen. Principal component density estimation for scenario generation using normalizing flows, 2022.

[16] Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Structures & Algorithms*, 22(1):60–65, 2003.

[17] Dimitris Achlioptas. Database-friendly random projections: Johnson-lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66(4):671–687, 2003. Special Issue on PODS 2001.

[18] Santosh Vempala. *The Random Projection Method*, volume 65 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. American Mathematical Society, 2005.

[19] Francesco Mezzadri. How to generate random matrices from the classical compact groups. *Notices of the American Mathematical Society*, 54(5):592–604, 2007.

[20] Alan Edelman and N. Raj Rao. Random matrix theory. *Acta Numerica*, 14:233–297, 2005.

[21] Persi Diaconis and Mehrdad Shahshahani. On the eigenvalues of random matrices. *Journal of Applied Probability*, 31:49 – 62, 1994.

[22] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

[23] Felix Xinnan X Yu, Ananda Theertha Suresh, Krzysztof M Choromanski, Daniel N Holtmann-Rice, and Sanjiv Kumar. Orthogonal random features. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

[24] Krzysztof M Choromanski, Mark Rowland, and Adrian Weller. The unreasonable effectiveness of structured random orthogonal embeddings. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[25] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[27] Renjie Liao, Simon Kornblith, Mengye Ren, David J. Fleet, and Geoffrey Hinton. Gaussian-bernoulli rbms without tears, 2022.

[28] Tijmen Tieleman. Training restricted boltzmann machines using approximations to the likelihood gradient. In *Proceedings of the 25th International Conference on Machine Learning*, pages 1064–1071. ACM, 2008.

[29] Radford M. Neal. Annealed importance sampling, 1998.

[30] Mathieu Germain, Karol Gregor, Iain Murray, and Hugo Larochelle. Made: Masked autoencoder for distribution estimation. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 881–889, Lille, France, 07–09 Jul 2015. PMLR.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We derive theory and apply the proposed method with success.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The approach does not greatly improve over other density estimators in every case, which we explain.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.

- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: See paper for complete derivation.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Yes, see paper for complete derivation.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.

- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: The datasets used are publicly available, and the methods described are straight forward to implement.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: See Experiments section.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: See Experiments section.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: See Experiments section.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: Paper does not violate any of the Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [No]

Justification: No particular societal impact besides those standard to generative models.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: No models are released, and the datasets used are publicly available.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.

- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Benchmarks and datasets are properly cited.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: No new assets released.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: No crowdsourcing was done.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.

- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: No human subjects nor crowdsourcing involved.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: No LLMs used for core research.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.