We thank all reviewers for their time and constructive comments.

Common Concerns. We first address concerns that were brought up by multiple reviewers.

- **(R 1, 2) Tractability and Runtime**. Since we use the CNF/ Neural ODE solver, our runtime is comparable to that of CNF, and thus we can use advances such as FFJORD [18] to scale with manifold dimension. Also note the log-determinant formulas for tested manifolds scale linearly in dimension, giving us tractability. PRNVP is faster as it uses the (non-continuous) RealNVP in the projected space. Also, our experiments show that NMODE is more sample efficient than other methods (Appendix C.2, first paragraph), so for density estimation we can use substantially fewer batches than other methods; this results in lower overall runtime.
- **(R 1, 2) Choice of ϵ.** Appendix D.1 discusses controlling the dynamics with ϵ. With a bound on the derivative we enforce a Lipschitz constraint on the NMODE solutions, in which case ϵ can be chosen less than r/L, where L is the Lipschitz constant and r is the local radius of injectivity. ϵ can be arbitrary for hyperbolic space (since the radius of injectivity is infinite) and should be less than π/L for the sphere. We find that most reasonably small choices work in practice. The quantifier for Prop 5.1 should be "for some"; this will be fixed.
- **(R 2, 3, 4) Applications and Concerns about Results** First note that aside from the general use cases presented in the paper (to data that does not require a manifold constraint), our approach gives a way to learn manifold-constrained flows for problems that require this constraint. In physics, we allow learning tractable densities on unitary Lie groups for Lattice Quantum Field Theory [37], in biology we need these densities for protein-structure prediction [21], and in robotics we need these for path navigation/motion estimation (e.g. modeling motion of a robot arm as a density in $\mathbb{T}^6$) [13]. Regarding our results, the lower margins with respect to Euclidean baselines as we increase dimension are due to the fact that it is possible to compensate for geometric distortion in higher Euclidean dimensions (Nickel et al., 2017). Note that for small dimensions (e.g. 2) our results significantly exceed the baselines and this is important for application via faster inference (more efficient modeling of data in low dimensions). We also note that although the example densities are simplified cases, we outperform all state-of-the-art baselines, highlighting existing flaws that we resolve.
- **(R 2, 4) Choice and Construction of Local Charts** The manifold is determined a priori in many applications, as in all applications above. Thus the local charts are determined a priori when the requisite manifold is selected and the construction follows from the manifold definition. Regarding "...choice of the charts for non-Riemannian manifolds...", note that the manifolds considered in our paper are smooth, hence admit a Riemannian metric, and are thus Riemannian. Therefore all manifolds we consider have well-defined charts.

Reviewer 1 Thank you for your kind words on our approach and the potential future impact of our paper.

- **"...the log likelihood...worse than those reported in...[2]."** We used [2]'s code (we had access before they open sourced it), fixed bugs in their implementation (that the authors confirmed to be actual bugs), and also fix evaluation (we run with a validation set, unlike [2], which gives misleadingly high numbers).
- **"Regarding the claims of generality..."** and **"I was confused by some parts of the motivations."** Yes, both (exp map and chart) are equally general. Projecting from the ambient space is not as principled (projecting $\mathbb{R}^3$ to $\mathbb{S}^2$ is not well-defined) whereas our exponential map helps learn over the space directly. We claim to be the first general principled method; we will augment the paper to make this more clear.

Reviewer 2 We are glad you appreciated the theoretical aspects and motivation of our paper.

- **"Does the method really benefit from using more complex solvers and the adjoint method in this case?"** Note that other methods do not provide the ability to incorporate more complex solvers from the classical literature, yet we do. Certainly, switching to one of these solvers will not always help (as in the case you mentioned). However, ϵ can frequently be fairly large; in these cases, a more complex solver would help.

Reviewer 3 Thank you for recognizing our theoretical and algorithmic merits.

- **"determinant of the derivative of the exponential map is needed...this may be difficult beyond the most simple manifold..."** This is possible given a parameterization of the manifold via local charts.
- **"...exponential maps are used...this should be stated much more explicitly..."** This will be done.
- **"Thm 4.1 follows from [4] due to the ambient assumption?"** Yes, the proof is analogous to that of [4].
- **"...the Whitney embedding Theorem is not isometric..."** NMODE itself is not affected by the change of metric. For MCNF, note that for relevant manifolds (e.g. those for applications given above: hyperboloid, sphere, unitary Lie groups, etc.) the data has an explicit embedding.

Reviewer 4 Thank you for your comments and appreciation of our usage of charts.

- **"...whether VAE can learn the local chart structure and the Riemannian metric..."** The VAE does not learn the manifold structure; the hyperbolic manifold is chosen a priori.