We would like to thank all the reviewers for their thoughtful and generally positive comments. We address their concerns below, and will make corresponding clarifications in the revision.

To Reviewer 1:

Q: Related works.

A: Thanks for pointing this out. We will include more related work and cite them more precisely. We have already discussed in the paper the related kernel-based DGMs, such as MMD-GAN[1], SMMD-GAN[2], Repulsive loss [3]. We will add more kernel learning methods such as IKL [4]. Even comparing to these results, ours is still much better.

Q: Experimental results.

A: Our proposed method learns the kernel using WGF. The method itself is independent of the kernel network construction. HK and HK-DK differ only at their network constructions, and HK-DK is based on kernel approximation using Fourier features [5]. The problem of HK-DK comes from sampling; it needs much more samples to ensure a good kernel approximation in high-resolution case, which can lead to training instability. Thus, the convergence issue of HK-DK does not affect the novelty and effectiveness of our proposed method. As for evaluation, we agree that there is no theoretical guarantee on the metrics, and the metrics may not be perfect. However, there is no better metric available as far as we know. Following existing works, we evaluated all the methods on both FID and IS, under the same settings, with the same model structures and datasets. We believe that it is fair enough to show the effectiveness of our method. Although Auto-GAN achieves comparable IS, its FID score is much worse than ours (i.e., 20.4% higher on CIFAR-10 and 29.5% higher on STL-10). Furthermore, Auto-GAN needs architecture search, which costs 43 hours according to the authors, while ours needs only 12 hours to obtain better results.

To Reviewer 2:

Q: Theorem, initial condition and explanation.

A: Thanks for the suggestions. We agree that all the PDFs can be seen as instances with different initial conditions and it is pointwise in practice. By approximating the heat kernel, we mean approximating the function with one specific initial condition (e.g. with a specific data point as the 'heat source'). We will rephrase the sentences accordingly to avoid confusion.

Q: Negative entropy v.s. relative entropy.

A: We use negative entropy because relative entropy is often referred to KL-divergence in some research domains, which may lead to confusion. We will emphasize this in the revision.

To Reviewer 3:

Q: About the assumption.

A: The assumption is used for proving the exponential convergence. It is related to the convexity property of the functional. We agree that the manifold could be much more complicated in practice, and our proposed method is still effective for those general manifolds (as demonstrated in the experiments). The algorithm is not biased. It only means that we cannot provide guarantees on the convergence rate and approximation error without the assumption.

Q: Reference and broader impact.

A: Thanks for the suggestion. We will discuss more recent works and impact.

To Reviewer 4:

Q: Experiments, pseudo-code and reproducibility.

A: Due to space limit, we provided pseudo-code of the deep generative model, experimental settings and hyper-parameters in the Appendix. The code is contained in the supplementary files with running scripts. We would like to emphasize that our model structures and data resolutions all follow previous works [2, 6] for fair comparisons. These works provided detailed information of their experimental settings and code online. Our settings and resolutions are consistent with theirs. Please kindly check our Appendix and code to verify this. The reported experimental results of all compared methods are under exactly the same settings. We copied some results directly from their papers whenever we cannot reproduce their results using their provided methods/codes. Thus, we believe that the comparisons are fair and our claims are conclusive.

References

[1] Li *et al.* Mmd gan: Towards deeper understanding of moment matching network. In *NeurIPS*, 2017.

[2] Michael Arbel, Dougal Sutherland, Mikołaj Bińkowski, and Arthur Gretton. On gradient regularizers for mmd gans. In *NeurIPS*, pages 6700–6710, 2018.

[3] Wei Wang, Yuan Sun, and Saman Halgamuge. Improving mmd-gan training with repulsive loss function. In *ICLR*, 2018.

[4] Chun-Liang Li, Wei-Cheng Chang, Youssef Mroueh, Yiming Yang, and Barnabas Poczos. Implicit kernel learning. In *AISTATS*, pages 2007–2016, 2019.

[5] Yufan Zhou, Changyou Chen, and Jinhui Xu. Kernelnet: A data-dependent kernel parameterization for deep generative modeling, 2019.

[6] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *ICLR*, 2018.