==========Reply to R#1==========

Thanks for pointing out our paper is significantly novel with competitive performance. We address your concerns below.

Q1: Far behind SOTA. Is it because of computational limits?

A1: We would like to clarify on the originality and goal of our contribution. In this paper, we zqnt to show the benefits of using adaptive stepsize for learning a ConvNet-based EBM where the energy landscape is highly nonconvex, not only via experiments but with a rigorous non-asymptotic theoretical analysis. As our method aims at accelerating the convergence of the model in the first iterations, we argue that our paper does not aim at proving an additional SOTA in terms of generated outputs. Rather, we tackle this problem from an optimization point of view with the motivation of improving how the latent samples find rapidly a good enough maxima in the target conditional distribution. Our numerical experiments and theorem do provide insights on that regards as the first epochs show.

Q2: About inpainting the same images for better comparison in Figure 6

A2: We will follow your suggestion to revise Figure 6.

Q3: How to choose the anistropic step size?

A3: The simplicity of our stepsize comes in the fact that it is based on the gradient of the distribution at each iteration. The only tuning parameter is the threshold which can be found empirically.

Q4: How the STANLEY alleviates the issues of the existing methods?

A4: See A1. We mostly focus on the convergence of the sampling scheme in the first epochs. This contribution is focussing on the transition regime and do not tackle the asymptotic convergence nor the final accuracy of the model.

Q5: How would the NUTS ampler compare?

A5: NUTS sampler is based on HMC sampler. We refer the reviewer to check our results for HMC in the numerical experiments.

Q6: About Typos.

A6: We will correct the typos. Thanks.

==========Reply to R#3==========

Q1: Not clear if the assumptions of Allassonniere and Kuhn hold for EBM.

A1: Our main contribution theory-wise is to extend their proof to the case of EBM, i.e. in the nonconvex case. Hence their assumptions do not suffice as they are stronger that our case.

Q2: Figure 1 doesn't support the conclusion that the proposed method recovers the density earlier.

A2: We will improve the resolution of Figure 6 for the sake of clarity.

Q3: How robust are the curves to different model initializations?

A3: This is an interesting point that we believe all papers in that realm should tackle in the future. Most of our runs are single runs and studying the robustness to the initialization by averaging multiple runs could be interesting.

Q4: Could STANLEY be leading only at the beginning in Figure 5? How do the curves behaves as the training progresses and convergence?

A4: As we tackle early convergence speed, we do not focus on the heavy tail of the convergence and certainly, our contribution does not reside in the resulting model accuracy but more on the ability to improve the ability of the MCMC to obtain good samples in the first epochs, i.e. when the EBM model parameters are still far from the target parameters.

Q5: Why use FID rather than PSNR to evaluate image inpainting?

A5: Jianwen can you reply to that please?

Q6: It would be more informative to apply two algorithms to the same image in Figure 6.

A6: we will follow suggestion in our revision.

Q7: How does the Langevin step subscript k combine with the iteration step subscript t? Do we start the scheme with the input argument z_0^m all the time or draw random initial states?

A7: The two subscripts are independent. k monitors the MCMC chain and t monitors the EBM training. The initialization of the chain is random at each new parameter hence at each t.

Q8: In theorem 1 drift function V_θ does not explicitly occur nor in eq. 8, nor in eq. 9. Is it present somewhere implicitly? What does χ stand for? Is there supposed to be $\pi_t heta(\cdot)$ instead of $\pi_t hetau(\cdot)$ in corollary 1?

A8: We define the drift function in the proof in the appendix. We will fix the typo raised by the reviewer. Thank you.

Q9: Proof of geometric ergodicity seems to follow the proof Allassonni'ere and Kuhn leading to limited novelty from the theoretical viewpoint

*A9:*Our main contribution theory-wise is to extend their proof to the case of EBM, i.e. in the nonconvex case. Hence their assumptions do not suffice as they are stronger that our case.

==========Reply to R#4==========