STANLEY: Stochastic Gradient Anisotropic Langevin Dynamics for Learning Energy-Based Models

We would like to thank the three reviewers for their feedback. Upon acceptance, we will include in the final version (a) *a clearer presentation of the algorithms* and (b) *additional experiments*.

We would like to address common concerns shared by the reviewers, noted R2, R3 and R4 for conciseness.

– Notations (R2/R3): Upon acceptance, the revised paper will fix and include more comprehensive notations, particularly used throughout the theory section of our contribution. The algorithm typos, in line 6, has been fixed (the stepsize does follow the EBM iteration index and not the MCMC iterations). We thank the reviewers for having pointed it out.

– Longer training procedures (R2/R3): Longer training procedure is a direction we will consider. Two options are possible: either increase the number of MCMC transitions per EBM iteration to make every chain convergent or increase the number of EBM iterations to reach better image quality. For the former, our method aims at reducing the number of MCMC iterations which we exhibit by running non-convergent Markov chains. The latter option is doable and would increase the quality of the images generated with our method but also of all the other baselines. Our purpose was to compare all the algorithms at the same iteration threshold and compare their accuracies (FID and visually) regardless of reaching the optimal image quality after training.

– Additional numerical experiments (R2/R3/R4):

Image completion:

More baselines:

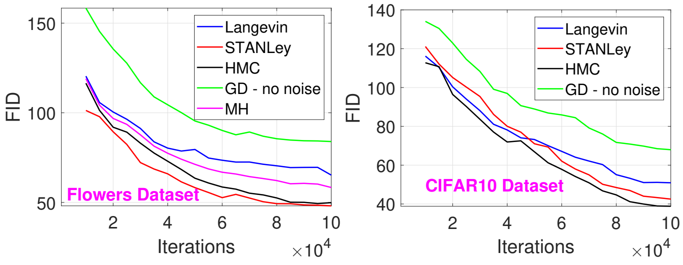

Figure 1. (FID values per method against 100k iterations elapsed). Left: Oxford Flowers dataset. Right: CIFAR-10 dataset.

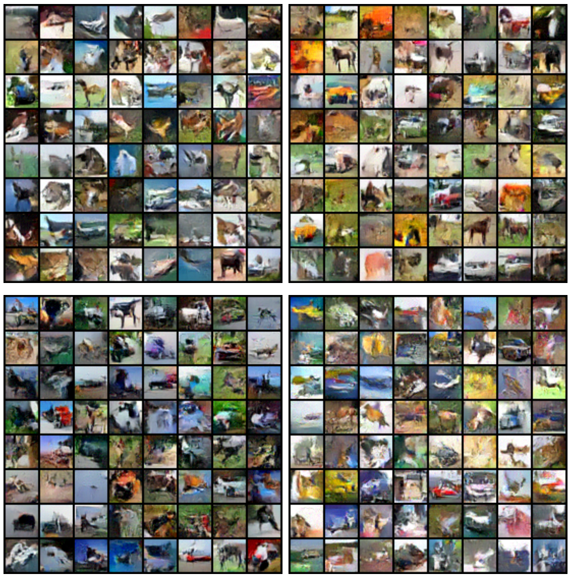

Figure 2. (CIFAR Dataset). 1: STANLey 2: MH. 3: HMC 4: GD without noise. After 100k iterations.

– Originality of our contributions (R2/R3/R4): We would like to use a paragraph of our rebuttal to address all reviewers consideration of our paper. We develop STANLey in order to more efficiently sample from the Gibbs potential. Hence, our goal is not to reach an optimal and high resolution generated image, but rather to decrease the number of kernel transitions need at each EBM iteration in order to obtain relatively good samples. Drastically reducing this number would have a great impact on the energy consumption and speed of the whole training process. Besides, we stress on the important theoretical contribution that is presented along our algorithm. To the best of our knowledge, EBM methods are presented mainly using empirical insights on their respective contribution. In this paper, we wanted to show the benefit of using adaptive stepsize for learning a convent-based EBM where the energy landscape is highly nonconvex, not only via experiments but with a rigorous non-asymptotic convergence analysis.

This also echoes with R4's remark on ”how the approach fits within the broader landscape of energy-based modeling approaches”. Indeed, we specifically design STANLey update to take into account the curvature of the nonconvex energy landscape by embedding a dimension gradient informed stepsize.