ICCV
#8604

ICCV
#8604

ICCV 2021 Submission #8604. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

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

We thank the three reviewers for their feedback.

– Originality of our contributions (R2/R3/R4): 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 benefits 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. *The goal is to propose an efficient algorithm with low computational cost and with theoretical guaranties of relevance. These are particularly hard tasks when dealing with high dimensional data and nonconvex models such as CNNs. Hence, we propose an optimized Langevin update and prove that this new sampler leads to a geometrically uniformly ergodic Markov chain.*

– 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 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):

We acknowledge the perfectibility of our contribution regarding numerical runs to convey the advantages brought by STANLey's update. While we want to insist on the theoretical aspects of our method and on the fact that our runs on two different datasets are meant to illustrate the faster convergence of our method, we will add to the revised paper additional experiments including image translation and image completion/inpainting experiments.
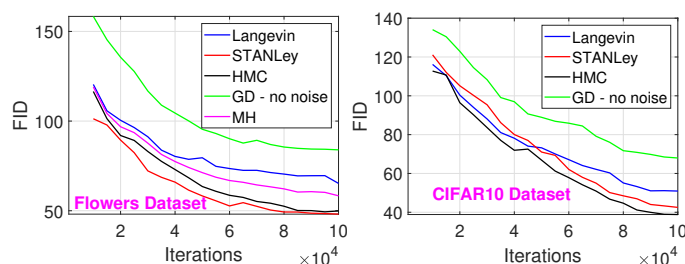

Figure 1. (FID values per method against 100k iterations elapsed). Left: Oxford Flowers dataset. Right: CIFAR-10 dataset.

More baselines: Regarding our current numerical setup, which was to generate synthetic samples from the Flowers dataset and CIFAR-10, we add below additional baselines including the Langevin with a Metropolis-Hastings (MH) rejection step, the HMC sampler and a plain Gradient Descent (GD) without noise update of the samples. Figure 1 shows the FID curves for those new baselines vs. theS-TANLey and plain Langevin sampler. Additional synthetic samples after 100k EBM iterations for those new baselines can also be found Figure 2.
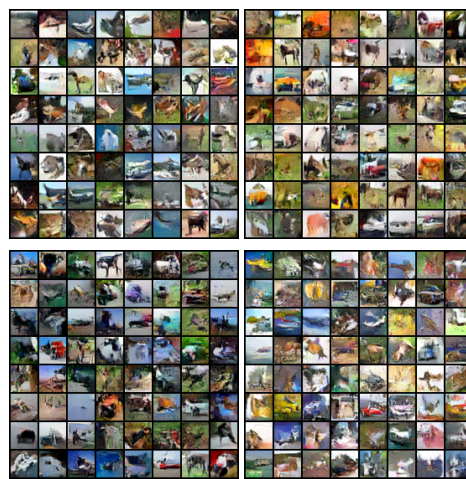

Figure 2. (CIFAR Dataset, from left to right) 1: STANLey 2: MH. 3: HMC 4: GD without noise. After 100k iterations.