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

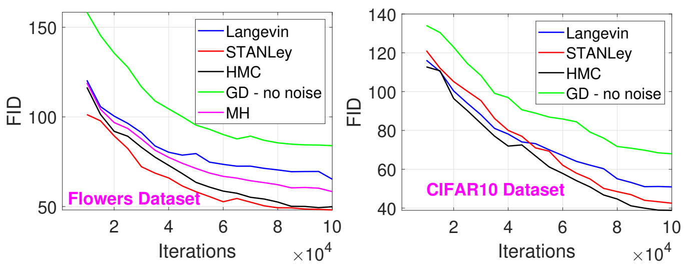

Figure 1. (FID values per method against 100k iterations elapsed). Left: Oxford Flowers dataset. Right: CIFAR-10 dataset.

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 R1, R2 and R3 for conciseness.

– Notations (R1/R2): 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 (R1/R2): Longer training procedure is doable, yet we want to stress of the fundamental aspect of our contribution. 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.

– Additional numerical experiments (R1/R2/R3):

More baselines:

– Originality of our contributions (R1/R2/R3):

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