

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 address common concerns shared by the reviewers, noted R2, R3 and R4.

– **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 our STANley update. While we want to insist on the theoretical aspect 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.

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. the STANley and plain Langevin sampler. Additional synthetic samples after 100k EBM iterations for those new baselines can also be found Figure 2.

Image inpainting:

– **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 resolu-

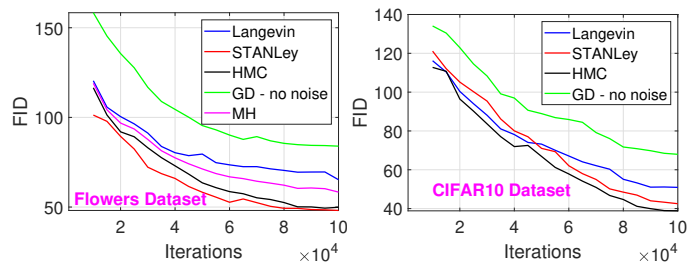


Figure 1. (FID values per method against 100k iterations elapsed). Left: Oxford Flowers dataset. Right: CIFAR-10 dataset.

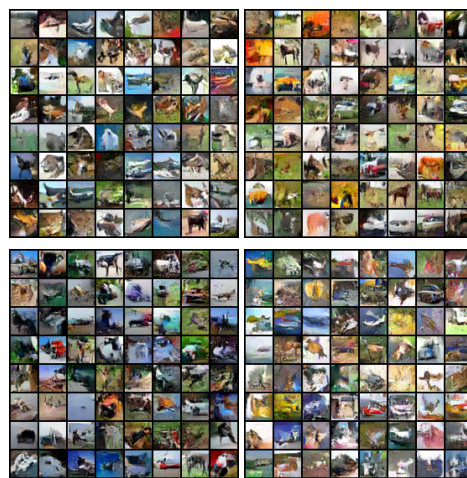


Figure 2. (CIFAR Dataset). 1: STANley 2: MH. 3: HMC 4: GD without noise. After 100k iterations.

tion 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 in-

108	formed stepsize.	162
109		163
110		164
111		165
112		166
113		167
114		168
115		169
116		170
117		171
118		172
119		173
120		174
121		175
122		176
123		177
124		178
125		179
126		180
127		181
128		182
129		183
130		184
131		185
132		186
133		187
134		188
135		189
136		190
137		191
138		192
139		193
140		194
141		195
142		196
143		197
144		198
145		199
146		200
147		201
148		202
149		203
150		204
151		205
152		206
153		207
154		208
155		209
156		210
157		211
158		212
159		213
160		214
161		215