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

Belhal Karimi   Jianwen Xie   Ping Li

1   Rebuttal

We thank the four reviewers for their valuable feedback. Our-point-by-point responses go as follows:

— Reviewer 1:Thanks for reading through our paper and your remarks.

— Reviewer 2: Thanks for your valuable suggestions and support.

— Reviewer 3: Thanks for your valuable feedback.

* Assumption H1

— Reviewer 5: Thanks for your valuable comments.

* **FID plots:** We would like to invite the reviewer to re-assess this remark on the FID curves. Indeed

* **Running times:** We provide the running times of our method and the baselines in Table 1 on Cifar-10 and Celeb-A datasets. We would like to stress on the similar computational complexity between the vanilla Langevin and our method STANLEY since our newly introduced stepsize uses the already computed gradient vector. On the contrary, the HMC method has recourse to both the gradient and the Hessian of the target distribution, resulting in longer computation time as reported on Table 1.

Table 1: Runtime in seconds for training our EBM during 1 epoch.

	Vanilla Langevin	HMC	GD	**STANLEY**
Cifar-10 Dataset	2	2	2	2
Celeb-A dataset	2	2	2	2

We run each of the method, including ours, on a single TitanXx8 GPU for our experiments.

* **Reproducibility:** As the reviewer as evaluated the reproducibility of our work as fair, we would like to highlight that the proofs of our results can be found in the supplementary material, the data is open source (CIFAR, Flowers and celeb-A) and the code can be requested.

* **Conclusion:** Except the FID curves concern, which was out of context, we addressed the main concern of Reviewer 5. For that reason we would like the reviewer to consider increasing its score on our contribution.

— Reviewer 6: Thanks for your valuable comments.

* **Originality of our contribution:**

* **EBMs vs GANs:**