AniLA: Anisotropic Langevin Dynamics for training Energy-Based Models

Belhal Karimi, Jianwen Xie, Ping Li

Cognitive Computing Lab
Baidu Research
10900 NE 8th St. Bellevue, WA 98004, USA

Abstract

1 Introduction

Given a stream of input noted x, the energy-based model (EBM) is a Gibbs distribution defined as:

$$p_\theta(x) = \frac{1}{Z(\theta)} \exp(f_\theta(x)) \tag{1}$$

2 MCMC based EBM

Energy Based Models: Energy based models are a class of generative models that leverages the power of Gibbs potential and high dimensional sampling techniques to produce high quality synthetic image samples.

MCMC procedures:

Focus on Langevin Diffusion:

3 ANILA sampler based EBM

3.1 Curvature informed MCMC

We introduce a new sampler based on the Langevin updates presented above.

Algorithm 1 ANILA FOR ENERGY-BASED MODEL

1: **Input**: Total number of iterations T, number of MCMC transitions K and of samples M learning rate η, initial values θ_0, $\{z_0^m\}_{m=1}^M$ and n observations $\{x_i\}_{i=1}^n$.

2: **for** $t = 1$ to T **do**

3: Compute the anisotropic stepsize as follows:

$$\gamma_t = \frac{b}{\max(b, |\nabla f_{\theta_t}(x)|} \tag{2}$$

4: Draw m samples $\{z_t^m\}_{m=1}^M$ from the objective potential (1) via Langevin diffusion:

$$z_t^m = z_t^m + \gamma_t/2\nabla f_{\theta_t}(x) + \sqrt{\gamma}\mathsf{B}_t \tag{3}$$

where B_t is the brownian motion, drawn from a Normal distribution.

5: Samples m positive observations $\{x_i\}_{i=1}^m$ from the empirical data distribution

6: Compute the gradient of the empirical log-EBM (1) as follows:

$$\nabla \sum_{i=1}^m \log p_{\theta_t}(x_i) = \mathbb{E}_{p_{\text{data}}}\left[\nabla_\theta f_{\theta_t}(x)\right] - \mathbb{E}_{p_\theta}\left[\nabla_{\theta_t} f_\theta(z_t^m)\right] \approx \frac{1}{m}\sum_{i=1}^m \nabla_\theta f_{\theta_t}(x_i) - \frac{1}{m}\sum_{i=1}^m \nabla_\theta f_{\theta_t}(z_t^m) \tag{4}$$

7: Update the vector of global parameters of the EBM:

$$\theta_{t+1} = \theta_{t+1} + \eta\nabla \sum_{i=1}^m \log p_{\theta_t}(x_i) \tag{5}$$

8: **end for**

9: **Output:** Generated samples $\{z_T^m\}_{m=1}^M$

3.2 Geometric ergodicity of ANILA sampler

We will present in this subsection, a convergence result for the Markov Chain constructed using Line 3-4.

4 Numerical Experiments

5 Conclusion