Weekly Report KARIMI-2021-09-17

My work this week has mainly been towards

1. Compressed EBM for memory efficiency
2. Spars AMS code for Imagenet
3. FEDLAMB arxiv paper and Bernoulli journal EM paper formatting and submission

1 Memory Efficient EBM Training

Definition 1 (Top-k). *For $x \in \mathbb{R}^d$, denote $\mathcal{S}$ as the size-k set of $i \in [d]$ with largest k magnitude $|x_i|$. The **Top-k** compressor is defined as $\mathcal{C}(x)_i = x_i$, if $i \in \mathcal{S}$; $\mathcal{C}(x)_i = 0$ otherwise.*

Definition 2 (Block-Sign). *For $x \in \mathbb{R}^d$, define M blocks indexed by $\mathcal{B}_i$, $i = 1, ..., M$, with $d_i := |\mathcal{B}_i|$. The **Block-Sign** compressor is defined as $\mathcal{C}(x) = [sign(x_{\mathcal{B}_1})\frac{\|x_{\mathcal{B}_1}\|_1}{d_1}, ..., sign(x_{\mathcal{B}_M})\frac{\|x_{\mathcal{B}_M}\|_1}{d_M}]$.*

Algorithm 1 EFF-EBM

1: **Input**: Number of iterations T, MCMC transitions K and of samples M, global learning rate $\{\eta_t\}_{t>0}$, MCMC stepsizes $\gamma_{k k>0}$, initial value θ_0, MCMC initialization $\{z_0^m\}_{m=1}^M$ and observations $\{x_i\}_{i=1}^n$.
2: **for** $t = 1$ to T **do**
3: Draw M samples $\{z_t^m\}_{m=1}^M$ from the objective potential via Langevin diffusion:
4: **for** $k = 1$ to K **do**
5: Use black box compression operators to compress the gradient with respect to z:

$$\tilde{g}_{k-1}^m = \mathcal{C}(\nabla_z f_{\theta_t}(z_{k-1}^m))$$

where $\mathcal{C}$ is either Sign, Topk, or maybe a simple count sketch operator (need to see in practice what makes sense).
6: Construct the Markov Chain as follows:

$$z_k^m = z_{k-1}^m + \frac{\gamma_k}{2}\tilde{g}_{k-1}^m + \sqrt{\gamma_k}\mathsf{B}_k \ , \tag{1}$$

where B_t denotes the Brownian motion (Gaussian noise).
7: **end for**
8: Assign $\{z_t^m\}_{m=1}^M \leftarrow \{z_K^m\}_{m=1}^M$.
9: Sample m positive observations $\{x_i\}_{i=1}^m$ from the empirical data distribution.
10: Compute the gradient of the empirical log-EBM:

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

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

$$\theta_{t+1} = \theta_t + \eta_t \nabla \log p(\theta_t) \ .$$

12: **end for**
13: **Output:** Vector of fitted parameters θ_{T+1}.

---

2 Spars AMS Imagenet and 1-bit Adam