Manifold Hamiltonian Dynamics for Variational Auto-Encoders

Yuanzhao Zhang, Yichuan Zhang

Department of Engineering, University of Cambridge, Cambridge, U.K.

UNIVERSITY OF CAMBRIDGE

Objectives

- Combine variational inference with MCMC methods
- Use Hamiltonian Monte Carlo first to improve vanilla variational autoencoder
- Use Manifold Hamiltonian Monte Carlo for faster convergence and better performance
- Apply to MNIST for marginal likelihood improvement
- Generate new images using trained decoder

Introduction

In variational auto-encoders, we approximate the posterior $p(z|x)$ with a family of distribution $q_\lambda(z|x)$ where λ includes the mean and variance of a Gaussian distribution for each data point $\lambda_{x_i} = (\mu_{x_i}, \sigma^2_{x_i})$. We then use KL-divergence to measure how well our approximation models the true posterior [1]:

$$KL(q_\lambda(z|x)||p(z|x)) = E_q[\log q_\lambda(z|x)] - E_q[\log p(x,z)] + \log p(x) \quad (1)$$

This is difficult to compute directly thus we try to maximum the evidence lower bound(ELBO) instead.

$$\mathcal{L} = E_{q(z|x)}[\log p(x,z) - \log q_\lambda(z|x)] \quad (2)$$

The maximization of ELBO can be addressed in multiple approaches, here in this project we are trying to address that in a way that combine the best of variational inference and MCMC methods [2]. In particular, we are interested in Hamiltonian Monte Carlo on Riemannian manifolds (MHMC), which is a state-of-the-art method for approximate Bayesian inference.

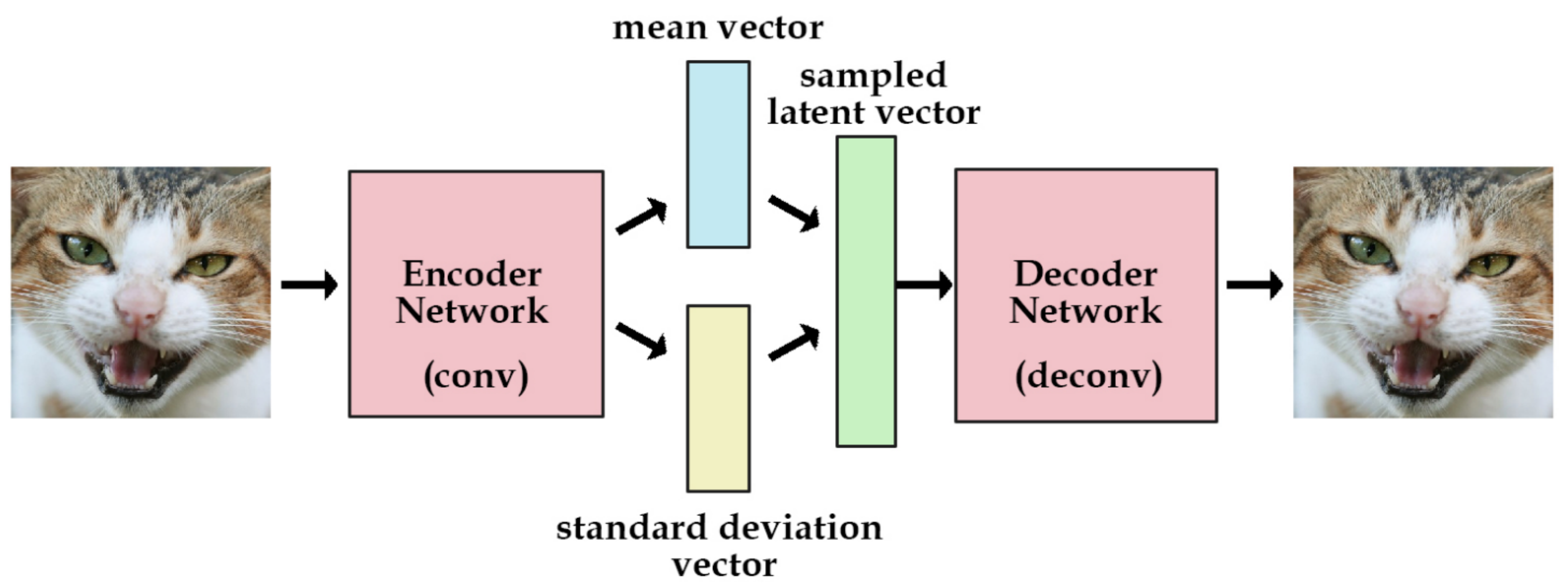


Figure 1: Structure of vanilla variational autoencoder

Methodology

(1) MCMC and auxiliary variables

In variational inference we have the evidence lower bound:

$$\mathcal{L} = E_{q_\theta(z|x)}[\log p(x,z) - \log q_\theta(z|x)] \quad (3)$$

If we integrate the auxiliary random variables into the lower bound we'll have:

$$\begin{aligned}\mathcal{L}_{aux} &= E_{q(y,z_T|x)}[\log[p(x,z_T)r(y|x,z_T)] - \log q(y,z_T|x)] \\ &= \mathcal{L} - E_{q(z_T|x)}\{KL[q(y|z_T,x)||r(y|z_T,x)]\} \\ &=\leq \mathcal{L} \leq \log[p(x)]\end{aligned}$$

where $r(y|x,z_T)$ is an auxiliary inference distribution. This auxiliary will again need to be approximated by a Markov structure:

$$r(y|x,z_T) = r(z_0,...z_{t-1}|x,z_T) = \prod_{t=1}^{T} r_t(z_{t-1}|x,z_t) \quad (4)$$

so that the variational lower bound can be rewritten as:

$$\log p(x) \geq E_q[\log p(x,z_T)/q(z_0|x) + \sum_{t=1}^{T} \log[r_t(z_{t-1}|x,z_t)/q_t(z_t|x,z_{t-1})]] \quad (5)$$

By specifying q_t and r_t in some flexible parametric form we can then optimize the lower bound to get a good approximation to the true posterior.

(2) Optimizing the lower bound

For most q_t and r_t, the lower bound in equation 5 cannot be computed analytically but we can approximate it by sampling from q_t. This involves the calculation of ratio:

$$\alpha_t = \frac{p(x,z_t)r_t(z_{t-1}|x,z_t)}{p(x,z_{t-1})q_t(z_t|x,z_{t-1})} \quad (6)$$

which can be used to update the lower bound: $L = L + \log[\alpha_t]$. The next step is to obtain the gradient of the lower bound with respect to parameter θ (parameters in q and r) using some reparameterization trick. Once we obtain that, we can use it in a stochastic gradient-based optimization algorithm for fitting $p(z|x)$ and get the final optimized variational parameters θ.

(3) Hamiltonian variational inference

Hamiltonian dynamics is a very effective way of exploring the posterior distribution $p(z|x)$. because the dynamics is guided by the gradient of the exact log posterior and random walks are suppressed by momentum variable v. What's more we have a nice property of HMC which is:

$$\begin{aligned}q(v_t, z_t|z_{t-1}, x) &= q(v_t, z_t, z_{t-1}|x)/q(z_{t-1}|x) \\ &= q(v'_t, z_{t-1}|x)/q(z_{t-1}|x) = q(v'_t|z_{t-1}, x)\end{aligned} \quad (7)$$

similarly we have $r(v'_t, z_{t-1}|z_t, x) = r(v_t|z_t, x)$. We can utilize this property to approximate the log marginal likelihood lower bound by calculating a ratio:

$$\alpha_t = \frac{p(x,z_t)r_t(v_t|x,z_t)}{p(x,z_{t-1})q_t(v'_t|x,z_{t-1})} \quad (8)$$

and update the lower bound: $L = L + log[\alpha_t]$. We fit the variational approximation to the true posterior by maximizing the lower bound with respect to q, r and the parameters of the Hamiltonian dynamics.

(4) Riemannian Manifold Hamiltonian Monte Carlo

A Riemannian Manifold Hamiltonian Monte Carlo (RMHMC) sampler can resolve the shortcomings of existing Monte Carlo algorithms when sampling from target densities that may be high dimensional and exhibit strong correlations. It also converges faster [3]. The major difference of it comparing with HMC is in $q(v'_t|z_{t-1}, x)$.

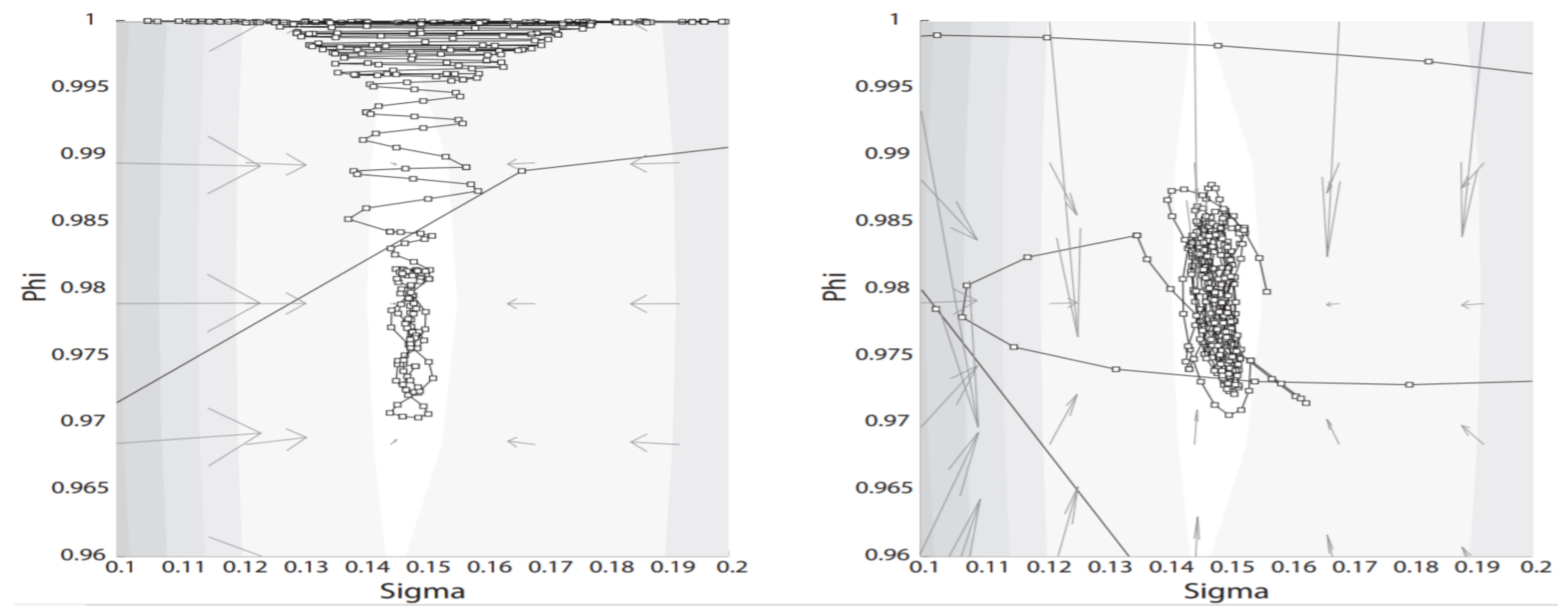
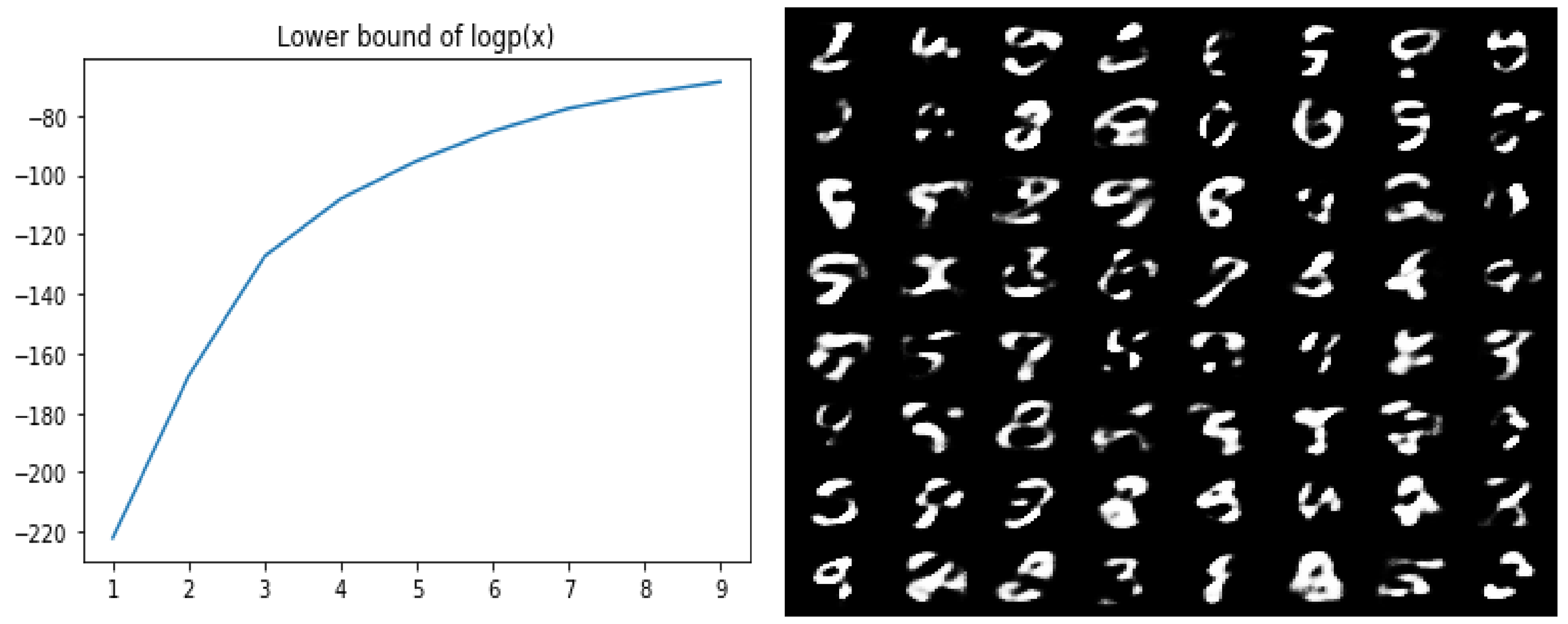


Figure 2: Comparison between RMHMC and HMC sampler

Results and Discussion



(a) Preliminary lower bound using HMC (b) Generated digits

Result is derived using Hamiltonian variational inference. A great increase in the lower bound can be seen. Each value is averaged over one epoch.

Future Work

Modify the current HMC sampler to implement MHMC sampler for a faster convergence and better lower bound on MNIST dataset. Then generate new images using other dataset like CIFAR-10.

References

[1] D. P. Kingma and M. Welling, "Auto-encoding variational bayes.," *CoRR*, vol. abs/1312.6114, 2013.

[2] T. Salimans, D. Kingma, and M. Welling, "Markov chain monte carlo and variational inference: Bridging the gap," in *Proceedings of the 32nd International Conference on Machine Learning* (F. Bach and D. Blei, eds.), vol. 37 of *Proceedings of Machine Learning Research*, (Lille, France), pp. 1218–1226, PMLR, 07–09 Jul 2015.

[3] M. Girolami, B. Calderhead, and S. A. Chin, "Riemann manifold langevin and hamiltonian monte carlo methods," *J. of the Royal Statistical Society, Series B (Methodological*.