Appendix for “Adversarial Symmetric Variational Autoencoder”

Yunchen Pu, Weiyao Wang, Ricardo Henao, Liqun Chen, Zhe Gan, Chunyuan Li and Lawrence Carin
Department of Electrical and Computer Engineering, Duke University
{yp42, ww109, r.henao, lc267, zg27, cl319, lcarin}@duke.edu

A Proof

Proof of Corollary 1.1 We start from a simple observation \( p_\theta(x) = \int_z p_\theta(x,z) dz = \int_z p(z) p_\theta(x|z) dz \). The second term in (5) of main paper can be rewritten as

\[ \mathbb{E}_{x \sim p_\theta(x|z'), z' \sim p(z), z \sim p(z)}[\log(1 - \sigma(f_{\psi_1}(x, z)))] , \tag{1} \]

\[ = \int_x \int_{z'} \int_z p_\theta(x|z') p(z') p(z) \log(1 - \sigma(f_{\psi_1}(x, z))) dx dz dz' \tag{2} \]

\[ = \int_x \int_z \left\{ \int_{z'} p_\theta(x|z') p(z') dz' \right\} p(z) \log(1 - \sigma(f_{\psi_1}(x, z))) dx dz \tag{3} \]

\[ = \int_x \int_z p_\theta(x)p(z) \log(1 - \sigma(f_{\psi_1}(x, z))) dx dz \tag{4} \]

Therefore, the objective function \( \mathcal{L}_{A1}(\psi_1) \) in (5) can be expressed as

\[ \int_x \int_z q(x) q_\psi(z|x) \log[\sigma(f_{\psi_1}(x, z))] dz dx + \int_x \int_z p_\theta(x)p(z) \log(1 - \sigma(f_{\psi_1}(x, z))) dx dz \]

\[ = \int_x \int_z \left\{ q_\psi(x, z) \log[\sigma(f_{\psi_1}(x, z))] + p_\theta(x)p(z) \log(1 - \sigma(f_{\psi_1}(x, z))) \right\} dx dz \tag{5} \]

This integral of (5) is maximal as a function of \( f(x, z) \) if and only if the integrand is maximal for every \( (x, z) \). Note that the problem \( \max_{x, a} \log x + b \log(1-x) \) achieves maximum at \( x = a/(a+b) \) and \( \sigma(x) = 1/(1 + e^{-x}) \). Hence, we have the optimal function of \( f_{\psi_1} \) at

\[ \sigma(f_{\psi_1}^*) = \frac{q_\psi(x, z)}{q_\psi(x, z) + p_\theta(x)p(z)} \quad f_{\psi_1}^* = \log q_\psi(x, z) + \log p_\theta(x)p(z) \tag{6} \]

Similarly, we have \( f_{\psi_2}(x, z) = \log p_\theta(x, z) - \log q_\psi(z) q(x) \)

Proof of Proposition 1 If \( \{\theta^*, \phi^*, \psi_1^*, \psi_2^*\} \) achieves an equilibrium of (12) of main paper. The Corollary 1.1 indicates that \( f_{\psi_1}^* = \log q_\psi(x, z) + \log p_\theta(x)p(z) \) and \( f_{\psi_2}(x, z) = \log p_\theta(x, z) - \log q_\psi(z) q(x) \).

Note that

\[ \mathcal{L}_{V A E}(\theta, \phi) = \mathbb{E}_{q(x)} \log p_\theta(x) - \text{KL}(q_\phi(x, z) \parallel p_\theta(x, z)) \tag{7} \]

\[ = \mathbb{E}_{q(x)} \log q(x) - \text{KL}(q_\phi(x, z) \parallel p_\theta(x, z)) - \text{KL}(q_\phi(x) \parallel p_\theta(x)) \tag{8} \]

and

\[ \mathcal{L}_{V A E}(\theta, \phi) = \mathbb{E}_{p(z)} \log q_\phi(z) - \text{KL}(p_\theta(x, z) \parallel q_\phi(x, z)) \tag{9} \]

\[ = \mathbb{E}_{p(z)} \log p(z) - \text{KL}(p_\theta(x, z) \parallel q_\phi(x, z)) - \text{KL}(p(z) \parallel q_\phi(z)) \tag{10} \]
where \( \mathbb{E}_{p(x)} \log p(z) \) and \( \mathbb{E}_{q(x)} \log q(x) \) can be considered as constant. Therefore, maximize \( \mathcal{L}_{VAEex} \) is equivalent to minimize
\[
\text{KL}(p_\theta(x, z) || q_\phi(x, z)) + \text{KL}(q_\phi(x, z) || p_\theta(x, z)) + \text{KL}(p_\theta(z) || q_\phi(z)) + \text{KL}(q_\phi(x) || p_\theta(x))
\]
The minimum of first two terms is achieved if and only if \( p_\theta(x, z) = q_\phi(x, z) \) while the minimums of last two terms are achieved at \( p_\theta(x) = q(x) \) and \( p(z) = q_\phi(z) \), respectively. Note that the joint match \( p_\theta(x, z) = q_\phi(x, z) \) is achieved, the marginals also matches which indicates the optimal \((\theta^*, \phi^*)\) is achieved if and only if \( p_\theta(x, z) = q_\phi^*(x, z) \).

## B Model Architecture

The model architectures are shown as following. For \( f_{\psi_1}(x, z) \) and \( f_{\psi_2}(x, z) \), we use the same architecture but the parameters are not shared.

---

**Figure 1:** Model architecture for MNIST

**Figure 2:** Model architecture for CIFAR

---
Figure 3: Encoder and decoder for ImageNet

Figure 4: Discriminator for ImageNet
Figure 5: Generated samples trained on CIFAR-10.
Figure 6: Generated samples trained on ImageNet.