Supplementary Material for Sphere GAN

Sung Woo Park and Junseok Kwon
School of Computer Science and Engineering, Chung-Ang University, Seoul, Korea
pswkiki@gmail.com jskwon@cau.ac.kr

1. Mathematical Proofs

To prove propositions in the paper, we first define notations, as follows.

Definition 1. *(Wesserstein spaces). Let $P(X)$ be the set of all Borel probability measures on a metric space X and (X, d) be the Polish metric space. The r-Wesserstein space where $1 \leq r < \infty$ is defined as*

$$P_r(X) = \left\{ \mathbb{P} \in P(X) \middle| \int_X d^r(x_0, x) d\mathbb{P}(x) < \infty, x_0 \in X \right\}. \tag{1}$$

The Wesserstein spaces generalize the space of probability measures that have finite central moments and it can be shown that the r-Wesserstein distance is a true metric on $P_r(X)$ [3, 2]. In the paper, we use the metric space $(X, d) = (\mathbb{S}^n, d_s)$ that has the Riemannian structure. As $(\mathbb{S}^n, d_s)$ is the Polish space[1] and $d_s(p_0, p)$ is a non-negative bounded function for any $p \in \mathbb{S}^n$, we can say that $P_s(\mathbb{S}^n) \subseteq P_1(\mathbb{S}^n)$ for all $1 \leq s$ and we have $P_r(\mathbb{S}^n) = P(\mathbb{S}^n)$ for all $r \in [1, \infty)$ [3]. Let $(\mathbb{P}_k)_{k \in N}$ be a sequence of probability measures in $P(X)$ and $\mathbb{Q}$ be another measure in $P(X)$. For sure, $(\mathbb{P}_k)_{k \in N} \subset P_r(\mathbb{S}^n)$. Let $r \in \Psi = \{1, 2, \dots\}$ be a set of moment order for r.

Proposition 1. *As $\mathbb{P}_k$ weakly converges to $\mathbb{Q}$,*

- $\gamma_{\mathbb{S}^n} \to 0$
- $\mathbf{m}^{\mathbf{r}}_{\mathbb{P}_\mathbf{k}} \to \mathbf{m}^{\mathbf{r}}_{\mathbb{Q}}$ *for all* $\mathbf{r}$

Proof. By the Definition (6.8, [2]), $\mathbb{P}_k \to \mathbb{Q}$ if and only if $\mathbf{m}^r_{\mathbb{P}_k} \to \mathbf{m}^{\mathbf{r}}_{\mathbb{Q}}$. This proves the second statement. Let $g^r_k = |\mathbf{m}^r_{\mathbb{P}_k} - \mathbf{m}^{\mathbf{r}}_{\mathbb{Q}}|$ and $g^s_k = |\mathbf{m}^s_{\mathbb{P}_k} - \mathbf{m}^{\mathbf{s}}_{\mathbb{Q}}|$ for some $r \neq s \in \Psi$. Then, the weak convergence means $g^r_k, g^s_k \to 0$ as $k \to \infty$. In other words, there exists some $\epsilon > 0$ such that $|g^r_k| < \frac{\epsilon}{2}, |g^s_k| < \frac{\epsilon}{2}$. However $g^r_k + g^s_r \leq |g^r_k + g^s_k| \leq |g^r_k| + |g^s_k| < \epsilon$, which implies that $\sum_{r,s} |m_{\mathbb{P}_k} - m_{\mathbb{Q}}| \to 0$. If we generalize this for all elements in Ψ, then $\gamma_{\mathbb{S}^n} \to 0$. This proves the first statement. □

[1] As $(\mathbb{S}^n, d_s)$ is compact metric space, it is complete and separable.

As a consequence of Proposition 1 and Theorem (6.9, [2]), the convergence in the r-Wesserstein space is equivalent to the convergence of the central moments of the order r. So, here after, we assume that $\mathbb{P}_k$ is weakly converge to $\mathbb{Q}$. Then followings are true.

Proposition 2. *As $\gamma_{\mathbb{S}^n}$ converges to 0,*

$$\sum_{r \in \Psi} W^r_{\mathbb{S}^n}(\mathbb{P}_k, \mathbb{Q}) \to 0. \tag{2}$$

Proof. As we mentioned before, $P_r(\mathbb{S}^n) = P(\mathbb{S}^n)$ and $\mathbb{Q}, \mathbb{P}_k \in \cup_{r \in \Psi} P_r(\mathbb{S}^n)$. By Theorem (6.9, [2]), $\mathbb{P}_k$ converges weakly to $\mathbb{Q}$ in $\cup_{r \in \Psi} P_r(\mathbb{S}^n)$ if and only if $\sum_{r \in \Psi} W_r(\mathbb{P}_k, \mathbb{Q}) \to 0$. □

Lemma 1. *The following distance function is differentiable and bounded.*

$$d_s(\mathbf{\Pi}^{-1}(p), \mathbf{\Pi}^{-1}(q)) = \arccos\left(\frac{||p||^2||q||^2 - ||p||^2 - ||q||^2 + 4p \cdot q + 1}{(||p||^2 + 1)(||q||^2 + 1)} \right). \tag{3}$$

Proof. The composition of two differentiable functions is differentiable and $\arccos$ is bounded because the inner product of two points in the image of inverse of stereographic projection is bounded. □

Lemma 2. $\mathbb{E}_{x \sim \mathcal{P}}\left[||\nabla_x d^r_s(\mathbf{N}, D(x))||_2 \right] < \infty$ *for all r.*

Proof. We assume that $\nabla_x D(x) < \infty$ and $y = D(x) \cdot \mathbf{N}$.

$$\nabla_x d^r_s(\mathbf{N}, D(x)) = \nabla_x \left[D(x) \cdot \mathbf{N} \right] \left[\frac{1}{-\sqrt{1 - y^2}} \sum_{r \in \Psi} (r-1) \arccos^{(r-1)}(y) \right] \leq \left| [\nabla_x D(x)] \cdot \mathbf{N} \right| \left| \frac{1}{-\sqrt{1 - y^2}} \sum_{r \in \Psi} (r-1) \arccos^{(r-1)}(y) \right| < \infty. \tag{4}$$

The first and second lines are induced by the chain rule and the following fact $\nabla_x [D(x) \cdot \mathbf{N}] = [\nabla_x D(x)] \cdot \mathbf{N}$. The last line is induced by $y = D(x) \cdot \mathbf{N} \neq 1$, because the north pole is excluded in the image of inverse of stereographic projection. □

Table 1. Generator used in CIFAR-10.

$z \in \mathbb{R}^{128} \sim \mathcal{N}(0, I)$
Dense (N: $4 \times 4 \times 256$)
ResBlock up (N:256, K:3, S:1, P:same, BN, ReLU)
ResBlock up (N:256, K:3, S:1, P:same, BN, ReLU)
ResBlock up (N:256, K:3, S:1, P:same, BN, ReLU)
Conv (N:3, K:3, S:1, BN, Tanh)

Table 2. Discriminator used in CIFAR-10.

Image $x \in \mathbb{R}^{32\times32\times3}$
ResBlock down (N:128, K:3, S:1, P:same, LN, LReLU)
ResBlock down (N:128, K:3, S:1, P:same, LN, LReLU)
ResBlock (N:128, K:3, S:1, P:same, LN, LReLU)
ResBlock (N:128, K:3, S:1, P:same, LN, LReLU)
Geometric Block

Table 3. Generator used in STL-10.

$z \in \mathbb{R}^{128} \sim \mathcal{N}(0, I)$
Dense (N: $3 \times 3 \times 512$)
ResBlock up (N:256, K:3, S:1, P:same, BN, ReLU)
ResBlock up (N:256, K:3, S:1, P:same, BN, ReLU)
ResBlock up (N:256, K:3, S:1, P:same, BN, ReLU)
ResBlock up (N:256, K:3, S:1, P:same, BN, ReLU)
Conv (N:3, K:3, S:1, BN, Tanh)

2. Network Architectures

In experiments for unsupervised image generation tasks using ResNet, we used similar structures which are introduced by [1]. In the last residual block of the discriminator network, we attached the geometric block for geometric transformation. We used the kernel of 3×3 size with stride 1 and same padding for all convolutional layers. For the generator, we used batch normalization with ReLU and layer normalization with Leaky ReLU of slope 0.2 in the discriminator network. Each residual block contained two convolutional layers. All latent codes were sampled from the 128 dimensional normal distribution with zero mean and unit variance.

The detailed architectures of our sphere GAN are as follows, where N, K, S, P, BN, LN, and LReLU denote the number of channels, kernel size, stride, padding, batch normalization, layer normalization, and Leaky ReLU, respectively.

In Algorithm 1, we provide Tensorflow-like pseudo code for the inverse of stereographic projection. Both input and output are tensors. Note that source code is available at https://github.com/pswkiki/SphereGAN.

Table 4. Discriminator used in STL-10.

Image $x \in \mathbb{R}^{48\times48\times3}$
ResBlock down (N:256, K:3, S:1, P:same, LN, LReLU)
ResBlock down (N:256, K:3, S:1, P:same, LN, LReLU)
ResBlock down (N:256, K:3, S:1, P:same, LN, LReLU)
ResBlock down (N:256, K:3, S:1, P:same, LN, LReLU)
ResBlock (N:256, K:3, S:1, P:same, LN, LReLU)
Geometric Block

Table 5. Geometric Block

LReLU
Average mean pooling
Dense (N: 1024)
Inverse of stereographic projection

Algorithm 1 Inverse of Stereographic projection

Input: Input tensor $u \in \mathbb{R}^n$.
Output: Output tensor $v \in \mathbb{S}^n \subset \mathbb{R}^{n+1}$.

$u \leftarrow \text{Divide}\big(\text{Transpose}(2u), \text{Pow}(\text{Norm}(u),2)+1\big)$.
$v \leftarrow \text{Divide}\big(\text{Pow}(\text{Norm}(u),2)-1, \text{Pow}(\text{Norm}(u),2)+1\big)$.
$v \leftarrow \text{Transpose}\ \big(\text{Concat}(u, v)\big)$.
return v.

3. Additional Results

Fig.1 shows the qualitative results of sphere GAN for CIFAR-10. In Figs.2 and 4, both IS and FID converged smoothly after 700K iterations. As shown in Figs.3 and 5, $\gamma_{\mathbb{S}^n}$ on CIFAR-10, STL-10 gradually reduced.

References

[1] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville. Improved training of wasserstein GANs. In *NIPS*, 2017.

[2] C. Villani. *Optimal Transport: Old and New*. Springer Berlin Heidelberg, 2008.

[3] D. Virosztek. On wasserstein isometries of probability measures on unit spheres. 02 2018.

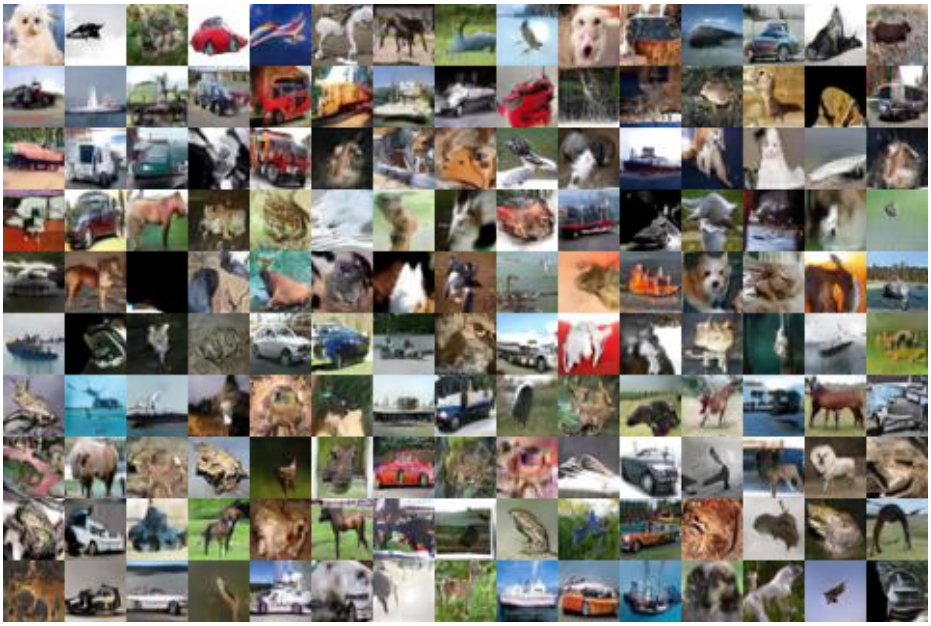

Figure 1. **Qualitative results** of sphere GAN for CIFAR-10 dataset

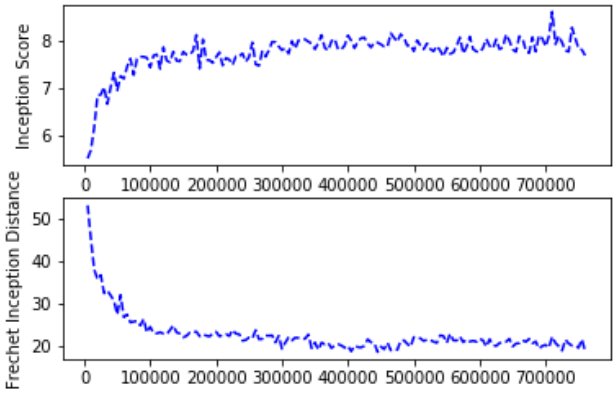

Figure 2. Learning curves of IS and FID for CIFAR-10 using the network architectures in Tables 1 and 2.

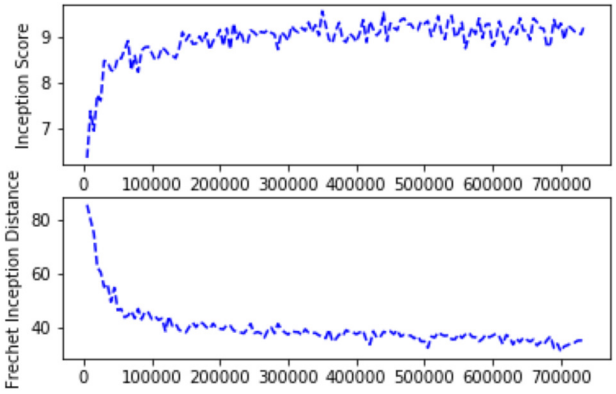

Figure 4. Learning curves of IS and FID for STL-10 using the network architectures in Tables 3 and 4.

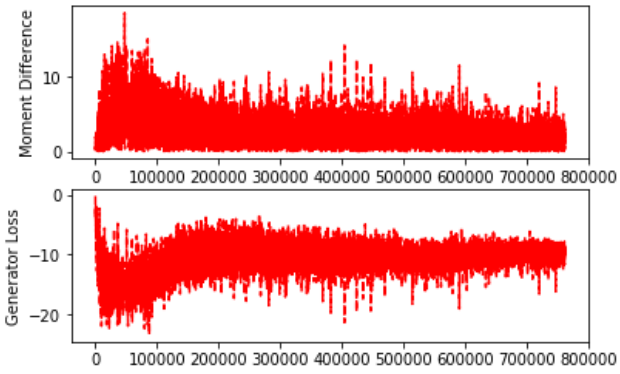

Figure 3. Learning curves of $\gamma_{\mathbb{S}^n}$ and the generator loss for CIFAR-10 using the network architectures in Tables 1 and 2.

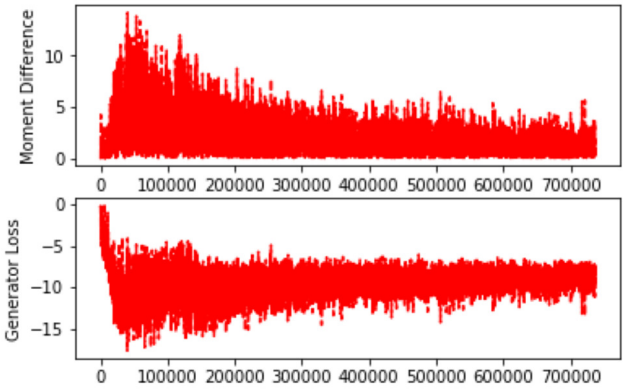

Figure 5. Learning curves of $\gamma_{\mathbb{S}^n}$ and the generator loss for STL-10 using the network architectures in Tables 3 and 4.