Supplementary Materials

Appendix A.

A.1 Proof of Theorem 1

Theorem 1. *Let* $p_{\mathrm{d}}(x)$ *and* $p_{\mathrm{g}}(x)$ *be the density functions for the data and model distributions,* $\mathbb{P}_{\mathrm{d}}$ *and* $\mathbb{P}_{\mathrm{g}}$*, respectively. Consider* $\mathcal{L}^{aw}(D, p_{\mathrm{g}}) = w_r \mathbb{E}_{x\sim p_d}\big[\log D(x)\big] + w_f \mathbb{E}_{x\sim p_g}\big[\log(1-D(x))\big]$ *with fixed* $w_r, w_f > 0$*.*

1. *Given a fixed* $p_{\mathrm{g}}(x)$*,* $\mathcal{L}^{aw}(D, p_{\mathrm{g}})$ *is maximized by* $D^*(x) = \frac{w_r p_{\mathrm{d}}(x)}{w_r p_{\mathrm{d}}(x) + w_f p_{\mathrm{g}}(x)}$ *for* $x \in \mathrm{supp}(p_{\mathrm{d}}) \cup \mathrm{supp}(p_{\mathrm{g}})$*.*

2. $\min_{p_{\mathrm{g}}} \max_D \mathcal{L}^{aw}(D, p_{\mathrm{g}}) = w_r \log \frac{w_r}{w_r + w_f} + w_f \log \frac{w_f}{w_r + w_f}$ *with the minimum attained by* $p_{\mathrm{g}}(x) = p_{\mathrm{d}}(x)$*.*

Proof.

1. First, the function $f(t) = a \log t + b \log(1-t)$ has its maximum in $[0, 1]$ at $t = \frac{a}{a+b}$. Given a fixed $p_{\mathrm{g}}(x)$, $w_r > 0$ and $w_f > 0$.

$$\mathcal{L}^{aw}(D, p_{\mathrm{g}}) = w_r \mathbb{E}_{x\sim p_{\mathrm{d}}}\left[\log\left(D(x)\right)\right] + w_f \mathbb{E}_{x\sim p_{\mathrm{g}}}\left[\log\left(1 - D(x)\right)\right] \tag{8}$$

$$= \int_x w_r p_{\mathrm{d}}(x) \log\left(D(x)\right) + w_f p_{\mathrm{g}}(x) \log\left(1 - D(x)\right) dx \tag{9}$$

$$\leq \int_x w_r p_{\mathrm{d}}(x) \log\left(D^*(x)\right) + w_f p_{\mathrm{g}}(x) \log\left(1 - D^*(x)\right) dx \tag{10}$$

$$= w_r \mathbb{E}_{x\sim p_{\mathrm{d}}}\left[\log\left(\frac{w_r p_{\mathrm{d}}(x)}{w_r p_{\mathrm{d}}(x) + w_f p_{\mathrm{g}}(x)}\right)\right] + w_f \mathbb{E}_{x\sim p_{\mathrm{g}}}\left[\log\left(\frac{w_f p_{\mathrm{g}}(x)}{w_r p_{\mathrm{d}}(x) + w_f p_{\mathrm{g}}(x)}\right)\right]. \tag{11}$$

where the equality holds if $D(x) = D^*(x)$. Therefore, $\mathcal{L}^{aw}(D, p_{\mathrm{g}})$ is maximum when $D = D^*$.

2. If $p_{\mathrm{g}}(x) = p_{\mathrm{d}}(x)$, then $D^*(x) = \frac{w_r}{w_r + w_f}$ and

$$\max_D \mathcal{L}^{aw}(D, p_{\mathrm{g}}) = w_r \mathbb{E}_{x\sim p_{\mathrm{d}}}\left[\log\left(\frac{w_r}{w_r + w_f}\right)\right] + w_f \mathbb{E}_{x\sim p_{\mathrm{g}}}\left[\log\left(\frac{w_f}{w_r + w_f}\right)\right] \tag{12}$$

$$= w_r \log\left(\frac{w_r}{w_r + w_f}\right) + w_f \log\left(\frac{w_f}{w_r + w_f}\right). \tag{13}$$

On the other hand,

$$\max_D \mathcal{L}^{aw}(D, p_{\mathrm{g}}) = w_r \mathbb{E}_{x\sim p_{\mathrm{d}}}\left[\log\left(\frac{w_r p_{\mathrm{d}}(x)}{w_r p_{\mathrm{d}}(x) + w_f p_{\mathrm{g}}(x)}\right)\right] + w_f \mathbb{E}_{x\sim p_{\mathrm{g}}} \log\left[\left(\frac{w_f p_{\mathrm{g}}(x)}{w_r p_{\mathrm{d}}(x) + w_f p_{\mathrm{g}}(x)}\right)\right] \tag{14}$$

$$\begin{aligned} &= w_r \log\left(\frac{w_r}{w_r + w_f}\right) + w_f \log\left(\frac{w_f}{w_r + w_f}\right) \\ &\quad + w_r KL\left(p_{\mathrm{d}} \middle| \frac{w_r p_{\mathrm{d}} + w_f p_{\mathrm{g}}}{w_r + w_f}\right) + w_f KL\left(p_{\mathrm{g}} \middle| \frac{w_r p_{\mathrm{d}} + w_f p_{\mathrm{g}}}{w_r + w_f}\right) \end{aligned} \tag{15}$$

$$\geq w_r \log\left(\frac{w_r}{w_r + w_f}\right) + w_f \log\left(\frac{w_f}{w_r + w_f}\right), \tag{16}$$

where KL is the Kullback-Leibler divergence and equality holds when $p_{\mathrm{d}} = \frac{w_r p_{\mathrm{d}} + w_f p_{\mathrm{g}}}{w_r + w_f}$ and $p_{\mathrm{g}} = \frac{w_r p_{\mathrm{d}} + w_f p_{\mathrm{g}}}{w_r + w_f}$. Thus we have shown that

$$\min_{p_{\mathrm{g}}} \max_D \mathcal{L}^{aw}(D, p_{\mathrm{g}}) = w_r \log\left(\frac{w_r}{w_r + w_f}\right) + w_f \log\left(\frac{w_f}{w_r + w_f}\right). \tag{17}$$

and minimum is attained when $p_{\mathrm{g}} = p_{\mathrm{d}}$.

$\square$

A.2 Proof of Theorem 2

Theorem 2. *Consider $\mathcal{L}_D^{aw}$ in (2) and the gradient $\nabla \mathcal{L}_D^{aw}$.*

1. *If $w_r = \frac{1}{\|\nabla \mathcal{L}_r\|_2}$ and $w_f = \frac{1}{\|\nabla \mathcal{L}_f\|_2}$, then $\nabla \mathcal{L}_D^{aw}$ is the angle bisector of $\nabla \mathcal{L}_r$ and $\nabla \mathcal{L}_f$, i.e.*

$$\angle_2 \left(\nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_r\right) = \angle_2 \left(\nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_f\right) = \angle_2 \left(\nabla \mathcal{L}_r, \nabla \mathcal{L}_f\right)/2. \tag{18}$$

2. *If $w_r = \frac{1}{\|\nabla \mathcal{L}_r\|_2}$ and $w_f = -\frac{\langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \rangle_2}{\|\nabla \mathcal{L}_f\|_2^2 \cdot \|\nabla \mathcal{L}_r\|_2}$, then*

$$\angle_2 \left(\nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_f\right) = 90^\circ, \ \angle_2 \left(\nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_r\right) \leq 90^\circ. \tag{19}$$

3. *If $w_r = -\frac{\langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \rangle_2}{\|\nabla \mathcal{L}_r\|_2^2 \cdot \|\nabla \mathcal{L}_f\|_2}$ and $w_f = \frac{1}{\|\nabla \mathcal{L}_f\|_2}$, then*

$$\angle_2 \left(\mathcal{L}_D^{aw}, \nabla \mathcal{L}_r\right) = 90^\circ, \ \angle_2 \left(\nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_f\right) \leq 90^\circ. \tag{20}$$

Proof.

1. If $w_r = \frac{1}{\|\nabla \mathcal{L}_r\|_2}$ and $w_f = \frac{1}{\|\nabla \mathcal{L}_f\|_2}$, then

$$\mathcal{L}_D^{aw} = \frac{1}{\|\nabla \mathcal{L}_r\|_2} \mathcal{L}_r + \frac{1}{\|\nabla \mathcal{L}_f\|_2} \mathcal{L}_f. \tag{21}$$

Using the definition of Euclidean inner product,

$$\begin{aligned}
\cos\left(\angle_2 \left(\nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_r\right)\right) &= \frac{\langle \nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_r \rangle_2}{\|\nabla \mathcal{L}_D^{aw}\|_2 \|\nabla \mathcal{L}_r\|_2} && (22)\\
&= \frac{\frac{1}{\|\nabla \mathcal{L}_r\|_2} \langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_r \rangle_2 + \frac{1}{\|\nabla \mathcal{L}_f\|_2} \langle \nabla \mathcal{L}_f, \nabla \mathcal{L}_r \rangle_2}{\|\nabla \mathcal{L}_D^{aw}\|_2 \|\nabla \mathcal{L}_r\|_2} && (23)\\
&= \frac{1}{\|\nabla \mathcal{L}_D^{aw}\|_2} + \frac{\langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \rangle_2}{\|\nabla \mathcal{L}_D^{aw}\|_2 \|\nabla \mathcal{L}_r\|_2 \|\nabla \mathcal{L}_f\|_2} && (24)
\end{aligned}$$

$$\begin{aligned}
\cos\left(\angle_2 \left(\nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_f\right)\right) &= \frac{\langle \nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_f \rangle_2}{\|\nabla \mathcal{L}_D^{aw}\|_2 \|\nabla \mathcal{L}_f\|_2} && (25)\\
&= \frac{\frac{1}{\|\nabla \mathcal{L}_r\|_2} \langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \rangle_2 + \frac{1}{\|\nabla \mathcal{L}_f\|_2} \langle \nabla \mathcal{L}_f, \nabla \mathcal{L}_f \rangle_2}{\|\nabla \mathcal{L}_D^{aw}\|_2 \|\nabla \mathcal{L}_f\|_2} && (26)\\
&= \frac{1}{\|\nabla \mathcal{L}_D^{aw}\|_2} + \frac{\langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \rangle_2}{\|\nabla \mathcal{L}_D^{aw}\|_2 \|\nabla \mathcal{L}_r\|_2 \|\nabla \mathcal{L}_f\|_2} && (27)
\end{aligned}$$

We can rewrite $\|\nabla \mathcal{L}_D^{aw}\|_2$ in term of $\angle_2 \left(\nabla \mathcal{L}_r, \nabla \mathcal{L}_f\right)$, that is

$$\begin{aligned}
\|\nabla \mathcal{L}_D^{aw}\|_2^2 &= \langle \nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_D^{aw} \rangle_2 \\
&= \left\langle \frac{1}{\|\nabla \mathcal{L}_r\|_2} \nabla \mathcal{L}_r + \frac{1}{\|\nabla \mathcal{L}_f\|_2} \nabla \mathcal{L}_f, \frac{1}{\|\nabla \mathcal{L}_r\|_2} \nabla \mathcal{L}_r + \frac{1}{\|\nabla \mathcal{L}_f\|_2} \nabla \mathcal{L}_f \right\rangle_2 \\
&= \frac{\langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_r \rangle_2}{\|\nabla \mathcal{L}_r\|_2^2} + \frac{\langle \nabla \mathcal{L}_f, \nabla \mathcal{L}_f \rangle_2}{\|\nabla \mathcal{L}_f\|_2^2} + \frac{2 \langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \rangle_2}{\|\nabla \mathcal{L}_r\|_2 \|\nabla \mathcal{L}_f\|_2} \\
&= 2\left(1 + \cos\left(\angle_2 \left(\nabla \mathcal{L}_r, \nabla \mathcal{L}_f\right)\right)\right). && (28)
\end{aligned}$$

Notice that (24) can be rewritten using $\angle_2 \left(\nabla \mathcal{L}_r, \nabla \mathcal{L}_f\right)$ as

$$\cos\left(\angle_2 \left(\nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_r\right)\right) = \frac{1}{\|\nabla \mathcal{L}_D^{aw}\|_2} + \frac{\langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \rangle_2}{\|\nabla \mathcal{L}_D^{aw}\|_2 \|\nabla \mathcal{L}_r\|_2 \|\nabla \mathcal{L}_f\|_2} \tag{29}$$

$$= \frac{1}{\left\|\nabla \mathcal{L}_D^{aw}\right\|_2} \left(1 + \cos\left(\angle_2\left(\nabla \mathcal{L}_r, \nabla \mathcal{L}_f\right)\right)\right) \tag{30}$$

$$= \sqrt{\frac{1 + \cos\left(\angle_2\left(\nabla \mathcal{L}_r, \nabla \mathcal{L}_f\right)\right)}{2}} \tag{31}$$

$$= \cos\left(\angle_2\left(\nabla \mathcal{L}_r, \nabla \mathcal{L}_f\right)/2\right). \tag{32}$$

Thus, $\angle_2\left(\nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_r\right) = \angle_2\left(\nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_f\right) = \angle_2\left(\nabla \mathcal{L}_r, \nabla \mathcal{L}_f\right)/2$.

2. If $w_r = \frac{1}{\|\nabla \mathcal{L}_r\|_2}$ and $w_f = -\frac{\langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \rangle_2}{\|\nabla \mathcal{L}_f\|_2^2 \|\nabla \mathcal{L}_r\|_2}$ then

$$\mathcal{L}_D^{aw} = \frac{1}{\left\|\nabla \mathcal{L}_r\right\|_2} \mathcal{L}_r - \frac{\left\langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \right\rangle_2}{\left\|\nabla \mathcal{L}_f\right\|_2^2 \left\|\nabla \mathcal{L}_r\right\|_2} \mathcal{L}_f. \tag{33}$$

Using this aw-loss function, we have

$$\left\langle \nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_f \right\rangle_2 = \left\langle \frac{1}{\left\|\nabla \mathcal{L}_r\right\|_2} \nabla \mathcal{L}_r - \frac{\left\langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \right\rangle_2}{\left\|\nabla \mathcal{L}_f\right\|_2^2 \left\|\nabla \mathcal{L}_r\right\|_2} \nabla \mathcal{L}_f, \nabla \mathcal{L}_f \right\rangle_2 \tag{34}$$

$$= \frac{\left\langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \right\rangle_2}{\left\|\nabla \mathcal{L}_r\right\|_2} - \frac{\left\langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \right\rangle_2 \left\langle \nabla \mathcal{L}_f, \nabla \mathcal{L}_f \right\rangle_2}{\left\|\nabla \mathcal{L}_f\right\|_2^2 \left\|\nabla \mathcal{L}_r\right\|_2} \tag{35}$$

$$= \frac{\left\langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \right\rangle_2}{\left\|\nabla \mathcal{L}_r\right\|_2} - \frac{\left\langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \right\rangle_2}{\left\|\nabla \mathcal{L}_r\right\|_2} = 0, \tag{36}$$

and

$$\left\langle \nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_r \right\rangle_2 = \left\langle \frac{1}{\left\|\nabla \mathcal{L}_r\right\|_2} \nabla \mathcal{L}_r - \frac{\left\langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \right\rangle_2}{\left\|\nabla \mathcal{L}_f\right\|_2^2 \left\|\nabla \mathcal{L}_r\right\|_2} \nabla \mathcal{L}_f, \nabla \mathcal{L}_r \right\rangle_2 \tag{37}$$

$$= \frac{\left\|\nabla \mathcal{L}_r\right\|_2^2}{\left\|\nabla \mathcal{L}_r\right\|_2} - \frac{\left\langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \right\rangle_2^2}{\left\|\nabla \mathcal{L}_f\right\|_2^2 \left\|\nabla \mathcal{L}_r\right\|_2} \tag{38}$$

$$\geq \left\|\nabla \mathcal{L}_r\right\|_2 - \frac{\left\|\nabla \mathcal{L}_f\right\|_2^2 \left\|\nabla \mathcal{L}_r\right\|_2^2}{\left\|\nabla \mathcal{L}_f\right\|_2^2 \left\|\nabla \mathcal{L}_r\right\|_2} \tag{39}$$

$$= \left\|\nabla \mathcal{L}_r\right\|_2 - \left\|\nabla \mathcal{L}_r\right\|_2 = 0. \tag{40}$$

Thus, $\angle_2\left(\nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_f\right) = 90^\circ$ and $\angle_2\left(\nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_r\right) \leq 90^\circ$.

3. If $w_r = -\frac{\langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \rangle_2}{\|\nabla \mathcal{L}_r\|_2^2 \cdot \|\nabla \mathcal{L}_f\|_2}$ and $w_f = \frac{1}{\|\nabla \mathcal{L}_f\|_2}$ then

$$\mathcal{L}_D^{aw} = -\frac{\left\langle \nabla \mathcal{L}_r, \nabla \mathcal{L}_f \right\rangle_2}{\left\|\nabla \mathcal{L}_r\right\|_2^2 \cdot \left\|\nabla \mathcal{L}_f\right\|_2} \mathcal{L}_r + \frac{1}{\left\|\nabla \mathcal{L}_f\right\|_2} \mathcal{L}_f, \tag{41}$$

and similar argument as above proves that $\angle_2\left(\nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_r\right) = 90^\circ$ and $\angle_2\left(\nabla \mathcal{L}_D^{aw}, \nabla \mathcal{L}_f\right) \leq 90^\circ$.

□

Appendix B.

B.1 Ablation study of α_1 and α_2 parameters

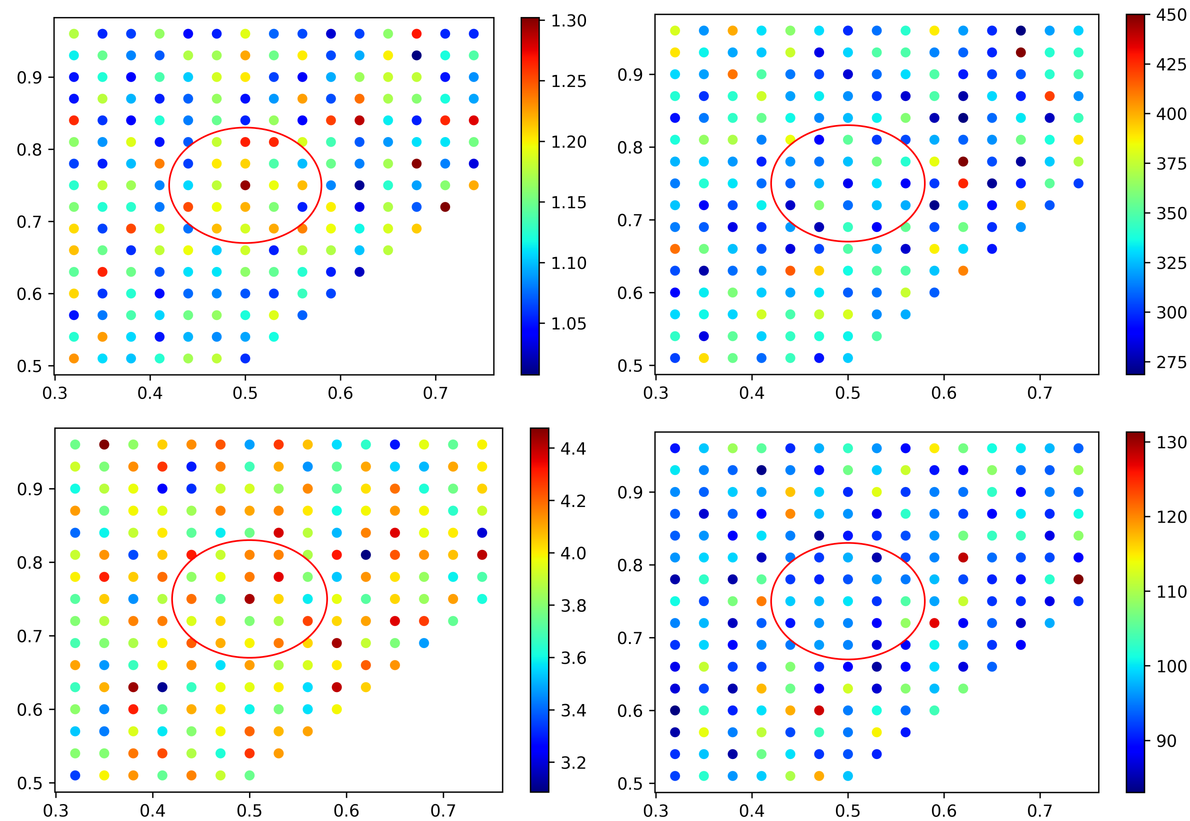

Figure 7: Top-left: IS grid after the first epoch; Bottom-left: IS grid after the fifth epoch; Top-right: FID grid after the first epoch; Bottom-right: FID grid after the fifth epoch.

We have considered the choice of α_1 and α_2 by experimenting with aw-AutoGAN models on the CIFAR-10 dataset. We recorded IS and FID scores after the first and the fifth epochs where the models were trained with the parameters in the grids shown in Figure 7 with all other settings fixed. We present the IS and FID scores in heat map plots in Figure 7. The results show that neighbors around the point $(0.5, 0.75)$ lead to good scores; the point $(0.5, 0.75)$ itself produces one of the highest IS and one of the lowest FID. In particular, the performance is not too sensitive to the selections.

B.2 Dataset and Implementation Details

We test our aw-method on the following datasets:

- The *CIFAR-10* dataset [25] consists of 60,000 color images with 50,000 for training and 10,000 for testing. All images have resolution 32×32 pixels and are divided equally into 10 classes, with 6,000 images per class. No data augmentation;

- The *STL-10* is a dataset proposed in [10] and designed for image recognition and unsupervised learning. STL-10 consists of 100,000 unlabeled images with 96×96 pixels and split into 10 classes. All images are resized to 48×48 pixels, without any other data augmentation;

- The *CIFAR-100* from [25] is a dataset similar to CIFAR-10 that consists of 60,000 color 32×32 pixel images that are divided into 100 classes. No data augmentation.

We follow the original implementations of SN-GAN, AutoGAN and BigGAN-PyTorch [4] that use the following hyper-parameters:

- *Generator:* learning rate: 0.0002; batch size: 128 (SN-GAN, AutoGAN) and 50 (BigGAN); optimizer: Adam optimizer with $\beta_1 = 0$ and $\beta_2 = 0.999$ [24]; loss: hinge [28, 43]; spectral normalization: False; learning rate decay: linear; # of training epochs: 320 (SN-GAN), 300 (AutoGAN) and 1000 (BigGAN);

- *Discriminator:* learning rate: 0.0002; batch size: 64 (SN-GAN, AutoGAN) and 50 (BigGAN); optimizer: Adam optimizer with $\beta_1 = 0$ and $\beta_2 = 0.999$; loss: hinge; spectral normalization: True; learning rate decay: linear; training iterations ratio: 3 (SN-GAN) and 2 (AutoGAN, BigGAN).

Experiments based on SN-GAN and AutoGAN models are performed on a single NVIDIA® QUADRO® P5000 GPU running Python 3.6.9 with PyTorch v1.1.0 for AutoGAN based models and Chainner v4.5.0 for SN-GAN based models. Experiments based on BigGAN-Pytorch [4] model are performed on two NVIDIA® Tesla® V100 GPU running Python 3.6.12 with PyTorch v1.4.0.

B.3 Aw-method with non-normalized gradients

In section 3, we introduced Algorithm 1 which was developed using Theorem 2 where the weights w_r and w_f include normalization of the gradients $\nabla\mathcal{L}_r$ and $\nabla\mathcal{L}_f$. This is not necessary and we can consider using non-normalized gradients directly where one of the weights is chosen as 1. The corresponding results are stated as the following theorem.

Theorem 3. *Consider $\mathcal{L}_D^{aw}$ in (2) and the gradient $\nabla\mathcal{L}_D^{aw}$.*

1. *If $w_r = 1$ and $w_f = -\frac{\langle\nabla\mathcal{L}_r, \nabla\mathcal{L}_f\rangle_2}{\|\nabla\mathcal{L}_f\|_2^2}$, then*

$$\angle_2\left(\nabla\mathcal{L}_D^{aw}, \nabla\mathcal{L}_f\right) = 90^\circ,\ \angle_2\left(\nabla\mathcal{L}_D^{aw}, \nabla\mathcal{L}_r\right) \leq 90^\circ. \tag{42}$$

2. *If $w_r = -\frac{\langle\nabla\mathcal{L}_r, \nabla\mathcal{L}_f\rangle_2}{\|\nabla\mathcal{L}_r\|_2^2}$ and $w_f = 1$, then*

$$\angle_2\left(\mathcal{L}_D^{aw}, \nabla\mathcal{L}_r\right) = 90^\circ,\ \angle_2\left(\nabla\mathcal{L}_D^{aw}, \nabla\mathcal{L}_f\right) \leq 90^\circ. \tag{43}$$

Proof. Identical to the proof of Theorem 2. □

Similar to section 3, we have developed Algorithm 2 using Theorem 3. The key difference between Algorithms 1 and 2 is normalization of the gradients; the rest of the algorithm is unchanged including the values for $\alpha_1 = 0.5$, $\alpha_2 = 0.75$, $\varepsilon = 0.05$ and $\delta = 0.05$.

Similar to Section 4, we tested Algorithm 2 on unconditional and conditional image generating tasks, with results provided in Tables 5 and 6, respectively. The implementation details are the same as in Section 4.

For the unconditional GAN, we tested our non-normalized aw-method on SN-GAN and AutoGAN baselines on three datasets: CIFAR-10, STL-10, and CIFAR-100. Our non-normalized methods also significantly improve SN-GAN baseline, in particular achieving the highest IS for the STL-10 dataset among comparisons in Table 2, and aw-AutoGAN non-normalized models improve the baseline AutoGAN models as well on all the datasets.

For the conditional GAN, we tested our non-normalized aw-method on SN-GAN and BigGAN models as baselines on CIFAR-10 and CIFAR-100 datasets. Both non-normalized aw-SN-GAN and aw-BigGAN significantly improve baseline models for both datasets.

On average, normalized weights achieve better results than non-normalized ones. We advocate the normalized version (Algorithm 1), but both produce quite competitive results and should be considered in implementations.

Algorithm 2: Adaptive weighted discriminator method w/o normalization for one step of discriminator training.

Given: $\mathbb{P}_{\mathrm{d}}$ and $\mathbb{P}_{\mathrm{g}}$ - data and model distributions;
Given: $\alpha_1 = 0.5$, $\alpha_2 = 0.75$, $\varepsilon = 0.05$, $\delta = 0.05$;
Sample: $x_1, \ldots, x_n \sim \mathbb{P}_{\mathrm{d}}$ and $y_1, \ldots, y_n \sim \mathbb{P}_{\mathrm{g}}$;
Compute: $\nabla\mathcal{L}_r$, $\nabla\mathcal{L}_f$, $s_r = \frac{1}{n}\sum_{i=1}^{n}\sigma(D(x_i))$, $s_f = \frac{1}{n}\sum_{j=1}^{n}\sigma(D(y_j))$;
if $s_r < s_f - \delta$ *or* $s_r < \alpha_1$ **then**
  if $\angle_2\left(\nabla\mathcal{L}_r, \nabla\mathcal{L}_f\right) > 90°$ **then**
    $w_r = 1 + \varepsilon; w_f = -\frac{\langle\nabla\mathcal{L}_r, \nabla\mathcal{L}_f\rangle_2}{\|\nabla\mathcal{L}_f\|_2^2} + \varepsilon;$
  else
    $w_r = 1 + \varepsilon; w_f = \varepsilon;$
  end
else if $s_r > s_f - \delta$ *and* $s_r > \alpha_2$ **then**
  if $\angle_2\left(\nabla\mathcal{L}_r, \nabla\mathcal{L}_f\right) > 90°$ **then**
    $w_r = -\frac{\langle\nabla\mathcal{L}_r, \nabla\mathcal{L}_f\rangle_2}{\|\nabla\mathcal{L}_r\|_2^2} + \varepsilon; w_f = 1 + \varepsilon;$
  else
    $w_r = \varepsilon; w_f = 1 + \varepsilon;$
  end
else
  $w_r = 1 + \varepsilon; w_f = 1 + \varepsilon;$
end

	CIFAR-10		STL-10		CIFAR-100	
Method	IS ↑	FID ↓	IS ↑	FID ↓	IS ↑	FID ↓
SN-GAN [34]	8.22±.05	21.7	9.10±.04	40.10	8.18±.12*	22.40*
aw-SN-GAN (Ours)	8.53±.11	12.32	9.53±.10	36.41	8.31±.02	19.08
aw-SN-GAN (non-norm.; Ours)	8.43±.07	12.65	9.61±.12	34.72	8.30±.11	19.48
AutoGAN [14]	8.55±.10	12.42	9.16±.12	31.01	8.54±.10*	19.98*
aw-AutoGAN (Ours)	9.01±.03	11.82	9.41±.09	26.32	8.90±.06	19.00
aw-AutoGAN (non-norm.; Ours)	8.98±.06	13.17	9.59±.14	27.97	8.72±.05	19.89

Table 5: Unconditional GAN: CIFAR-10, STL-10, and CIFAR-100 scores for the normalized (Algorithm 1) and non-normalized (Algorithm 2) versions of aw-method; * - results from our test.

	CIFAR-10		CIFAR-100	
Method	IS ↑	FID ↓	IS ↑	FID ↓
SN-GAN (cond.) [34]	8.60±.08	17.5	9.30†	15.6†
aw-SN-GAN (cond.; Ours)	9.03±.11	8.11	9.48±.13	14.42
aw-SN-GAN (non-norm.; cond.; Ours)	9.00±.12	8.03	9.44±.16	14.00
BigGAN [5]	9.22	14.73	10.99±.14*	11.73*
aw-BigGAN (cond.; Ours)	9.52±.10	7.03	11.22±.17	10.23
aw-BigGAN (non-norm.; cond.; Ours)	9.50±.07	6.89	11.26±.20	10.25

Table 6: Conditional GAN: CIFAR-10 and CIFAR-100 scores for the normalized (Algorithm 1) and non-normalized (Algorithm 2) versions of aw-method; † - quoted from [41]; * - results from our tests based on [4].