1 Q1

There was a question on how do one determine when to stop iterating the original GAN algorithm. We show here that the GAN algorithm does not converge in general using an example. Therefore the study of number of iterations required for GAN is formally ill defined.

Define the data space to be a bounded subset on the real line, $[\alpha, \beta] \subset \mathbb{R}$. Let the real data set consist of only one data point $x = (\alpha + \beta)/2$. The noise data set also consist of only one point $z \neq 0, z \in \mathbb{R}$. Define the generator with only one parameter $\theta \in \mathbb{R}$ to generate the fake data using $\hat{x} = \theta z$. The cost function is defined as,

$$C(w, b, \theta) = -\frac{1}{1 + \exp(-wx - b)} + \frac{1}{1 + \exp(-w\theta z - b)} \tag{1}$$

The parameters are constrained to $w \in [-1, 1]$, $b \in \mathbb{R}$, $\theta z \in [\alpha, \beta]$.

1. Initialise θ_0 such that $\theta_0 z \neq x$.
2. Holding θ constant, solve w, b for the global minimum of $C(w, b, \theta)$
3. Holding w, b constant, solve θ for the global maximum of $C(w, b, \theta)$
4. Iterate steps (2) & (3)

Show that the above iterative steps does not converge. That is w, b, θ does not tend to a fix point.

2 Q2

Fig. 1 shows the algorithm for WGAN, read the original paper M. Arjovsky *et al*, Wasserstein GAN arXiv:1701.07875v2, and discuss the short comings of this algorithm. Focus your discussion on the correctness of this algorithm.

Algorithm 1 WGAN, our proposed algorithm. All experiments in the paper used the default values $\alpha = 0.00005$, $c = 0.01$, $m = 64$, $n_{\text{critic}} = 5$.

Require: : α, the learning rate. c, the clipping parameter. m, the batch size. n_{critic}, the number of iterations of the critic per generator iteration.

Require: : w_0, initial critic parameters. θ_0, initial generator's parameters.

while θ has not converged **do**

  for $t = 0, ..., n_{\text{critic}}$ **do**

    Sample $\{x^{(i)}\}_{i=1}^m \sim \mathbb{P}_r$ a batch from the real data.

    Sample $\{z^{(i)}\}_{i=1}^m \sim p(z)$ a batch of prior samples.

    $g_w \leftarrow \nabla_w \left[\frac{1}{m}\sum_{i=1}^m f_w(x^{(i)}) - \frac{1}{m}\sum_{i=1}^m f_w(g_\theta(z^{(i)}))\right]$

    $w \leftarrow w + \alpha \cdot \text{RMSProp}(w, g_w)$

    $w \leftarrow \text{clip}(w, -c, c)$

  end for

  Sample $\{z^{(i)}\}_{i=1}^m \sim p(z)$ a batch of prior samples.

  $g_\theta \leftarrow -\nabla_\theta \frac{1}{m}\sum_{i=1}^m f_w(g_\theta(z^{(i)}))$

  $\theta \leftarrow \theta - \alpha \cdot \text{RMSProp}(\theta, g_\theta)$

end while

Figure 1: WGAN algorithm