Training WGANs with Peer Instruction

Taif Ghiwaa **Supervisor: James Foulds**
Department of Information Systems, UMBC
{taifg1, jfoulds}@umbc.edu

Generative adversarial networks (GANs) have shown superb results for generating synthetic data via a competitive game between two networks (Goodfellow et al., 2014). In these models, a generator network maps random noise to produce synthetic data, and a discriminator network discriminates whether the data fed to it is real or synthetic. Regardless of their success, GANs suffer from instability and difficulties in training. In (Arjovsky et al., 2017; Gulrajani et al., 2017), an alternative model, the WGAN, was proposed to improve the stability by using the Wasserstein distance and enforcing a Lipschitz constraint through weight clipping in the former, and a gradient penalty in the latter.

One specific property for WGANs, which improves stability over GANs, is that the critic should be trained to convergence, to compute the gradient for the generator. Based on this perspective, we propose an alternative method for training WGAN models, where multiple WGANs are trained in parallel and critique each other's examples. Therefore, WGANs search for the best critic among others, to compute the Wasserstein distance and its gradient. So, when one critic is far away from the optimality, another critic from another WGAN can help it to learn better. Thus, the procedure computes the generator's gradient more accurately, allowing the method to converge faster than the standard WGAN.

The proposed method can be thought of as analogous to the real-life setting of students in an active learning classroom, which is a collaborative learning environment involving interactions between two or more students (Freeman et al., 2014). Our WGAN learning algorithm resembles how humans learn via the pedagogical technique of *peer instruction* in a *flipped classroom*: each WGAN is a "student," trying to understand the latent concept behind the data by studying on their own before class (i.e. training the discriminators), then helping each other to answer questions in class (i.e. training the generators). During class, each student (WGAN) shares and helps each other answer multi-choice questions (i.e. whether data is real or generated). Peer instruction has proved effective in improving student learning outcomes, by considering a diversity of individual background and experience, and using this diversity to benefit the whole classroom (Freeman et al., 2014). Our preliminary results indicate that this intuition extends to WGANs.

We conducted two experiments to demonstrate the improvement in both the speed of convergence and the quality of generated samples. While we focus on image datasets here, natural language processing (NLP) applications will be considered in our future work, e.g. generating sentences with specific characteristics (Subramanian et al., 2017), dialog generation (Li et al., 2017), and translation (Yang et al., 2018). We trained the proposed model on a toy dataset and MNIST digits. Our model achieves faster convergence speed, and also sharper and more coherent images sampled from a generator when comparing it with the standard WGAN (see Figures in the Appendix). In addition, it saves time by accelerating the convergence, due to providing a stronger critic to the generator. We are currently working on a parallel implementation, and large-scale experimental results.

References

Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN. *arXiv preprint arXiv:1701.07875v3.*

Freeman, S., Eddy, S. L., Mcdonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410-8415.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2014). Generative adversarial nets. *In Advances in Neural Information Processing Systems*, 2672-2680

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). Improved training of Wasserstein GANs. *arXiv preprint arXiv:1704.00028v3.*

Li, J., Monroe, W., Shi, T., Jean, S., Ritter, A., & Jurafsky, D. (2017). Adversarial Learning for Neural Dialogue Generation. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing.

Rajeswar, S., Subramanian, S., Dutil, F., Pal, C., & Courville, A. (n.d.). Adversarial Generation of Natural Language. 2017. *arXiv preprint arXiv:1705.10929v1.*

Yang, Z., Chen, W., Wang, F., & Xu, B. (2018). Improving neural machine translation with conditional sequence generative adversarial nets. *arXiv preprint arXiv:1703.04887v4.*

Appendix

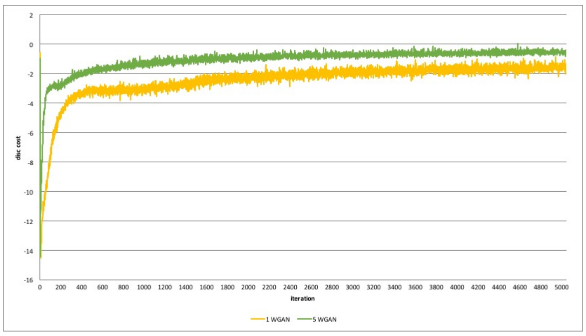

Figure 1: The line graph illustrates the disc cost for 5K iterations for two WGANs frameworks. Our model (green line) shows the faster converge comparing with standard WGAN (yellow line).

Figure 2: 2D generated samples for running MNIST dataset. The left side shows samples from standard WGAN, and the right side shows samples from our model. Our model generates better samples comparing it with standard WGAN for a selected iteration.