Research on the stability of generative adversarial network

Maxim Zubkov, Andrey Filatov
Moscow Institute of Physics and Technology (MIPT)

Introduction

Generative Adversarial Networks (GAN) — is a type of generative model able to generate complex data such as pictures, sound and video, but such models often exhibit unstable behavior during training. The purpose of our study is to review and apply theoretical results proposed in papers [1, 2] to stabilize GAN learning in a task of generating a multimodal distribution.

Aim

Despite the great success of GANs in various ML tasks, training GAN is quite problematic. As shown in the article [1], GAN does not generally converge. To solve this problem, the article [2] considers a theorem which formulates a sufficient condition (D1-D3, G1-G2) of GAN convergence, some of them are well studied as shown in table 1. According to the theorem, the discriminator must be trained till convergence, and in our research, we studied the influence of the number of discriminator's training iterations (n_{cr}) on GAN convergence and performance.

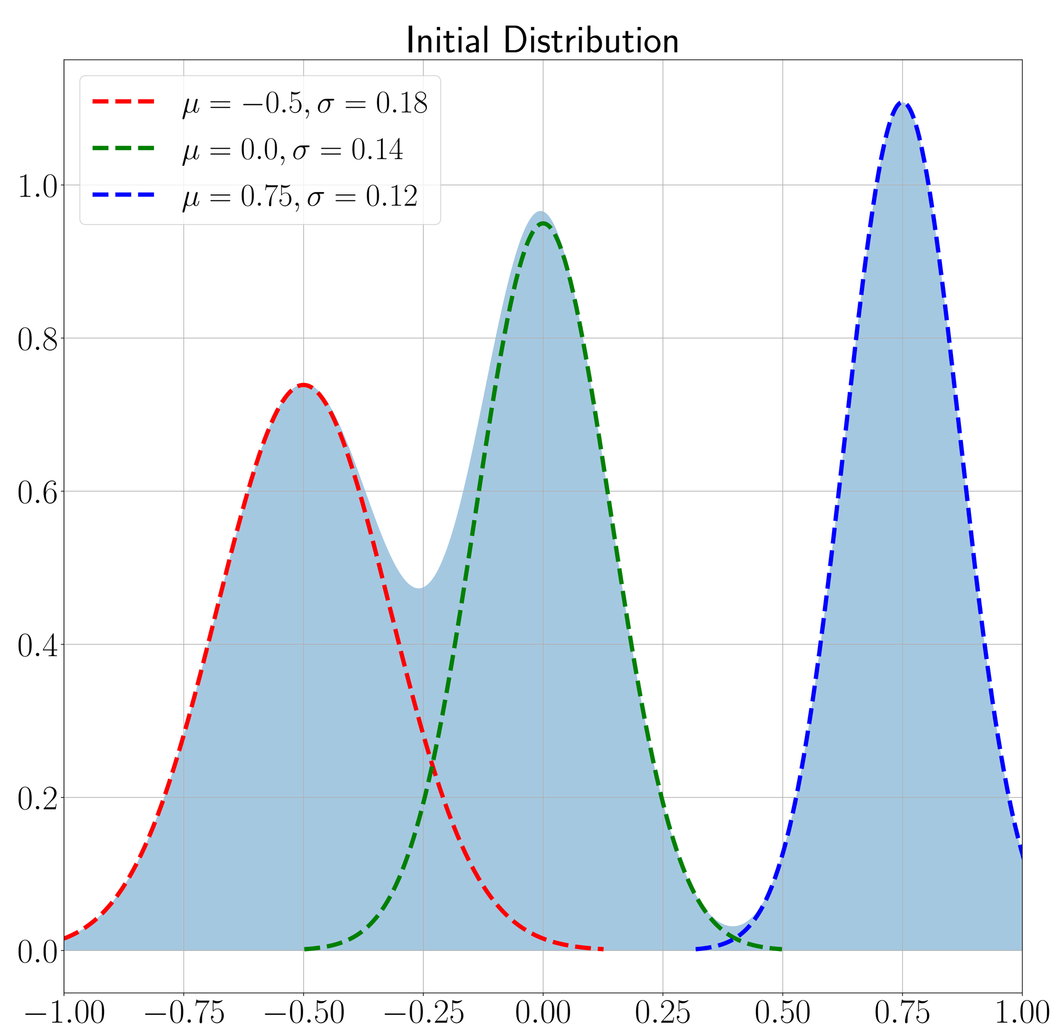


Figure 1: Linear combination of normal distributions

Condition	GAN techniques
(D0)	Train discriminator till convergence
(D1)	Spectral normalization [4] Lipschitz regularization [5]
(D2)	Smooth activation function, spectral normalization [4]
(D3)	Adversarial attack [6] WGAN-GP [3]
(G1)	Not studied yet
(G2)	Not studied yet

Table 1: The table matches the theoretical results obtained in the article [2] with the known methods of machine learning

Code: `https://github.com/maximzubkov/opt-project`.
Email: zubkov.md@phystech.edu, filatov.av@phystech.edu

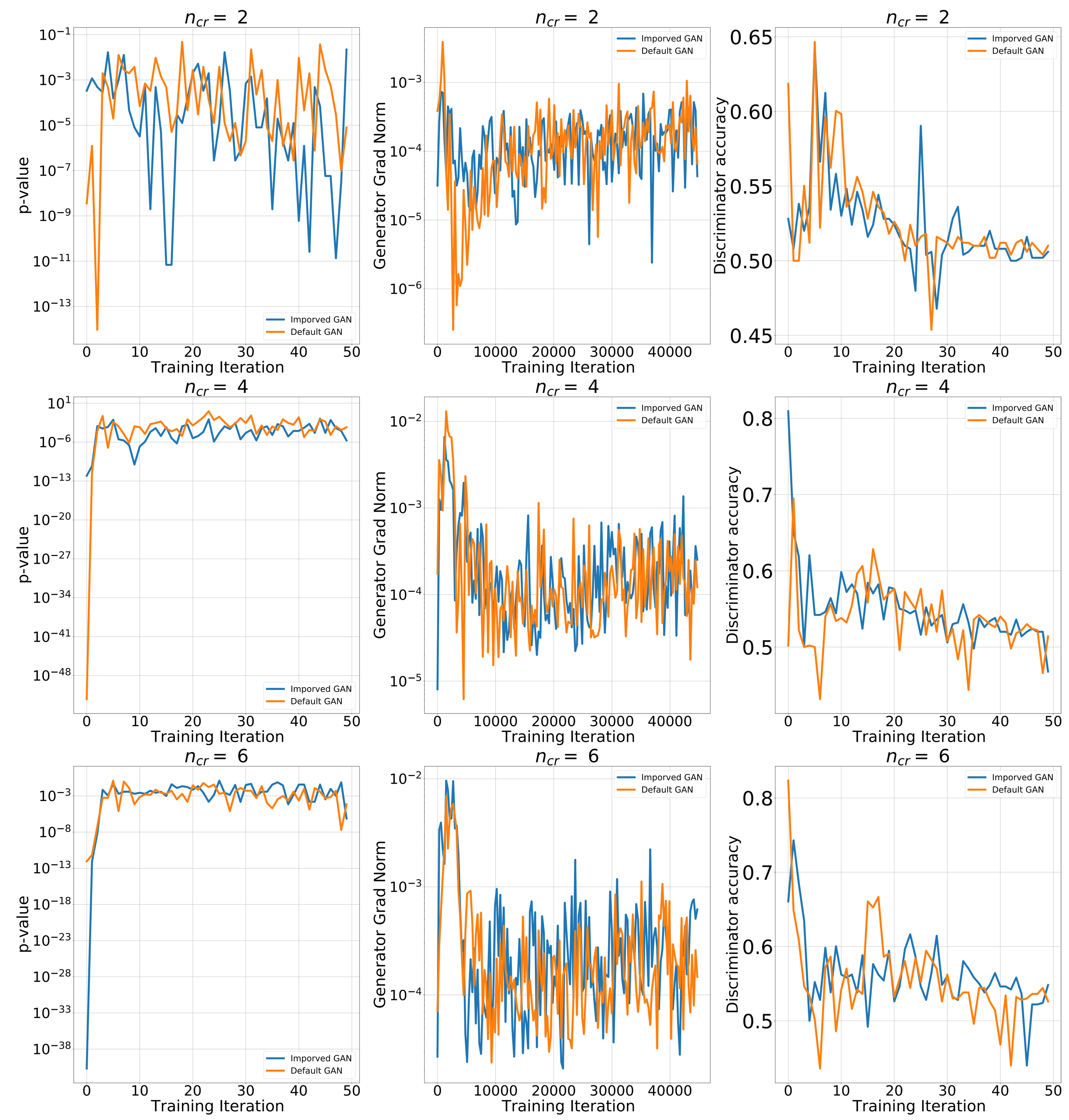
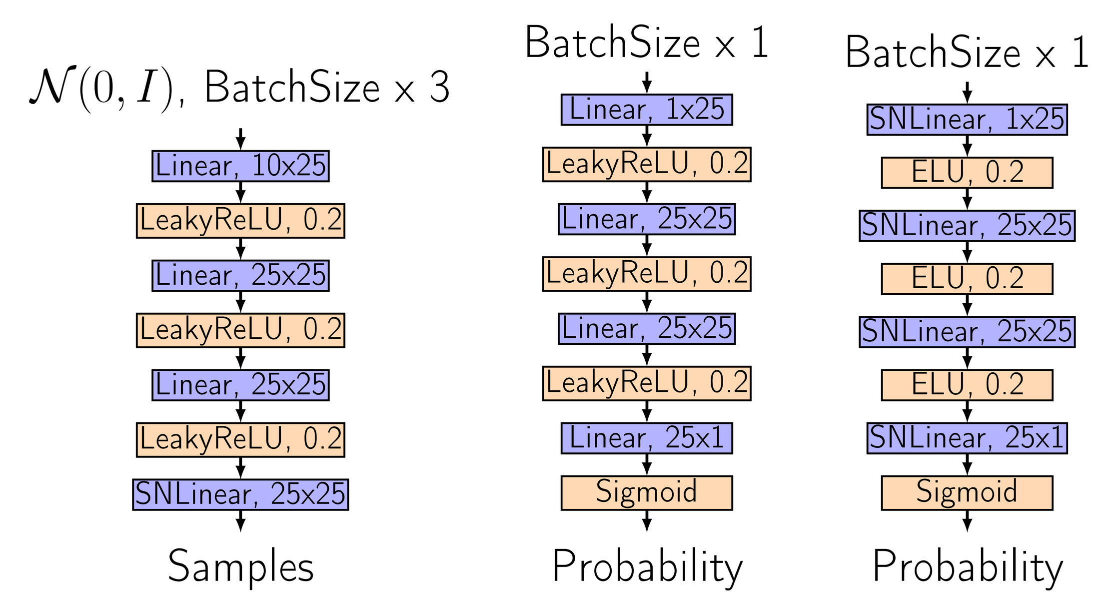


Figure 2: Comparison of p-value, gradient norm of generator, and discriminator accuracy for default GAN and improved GAN at different value n_{cr}

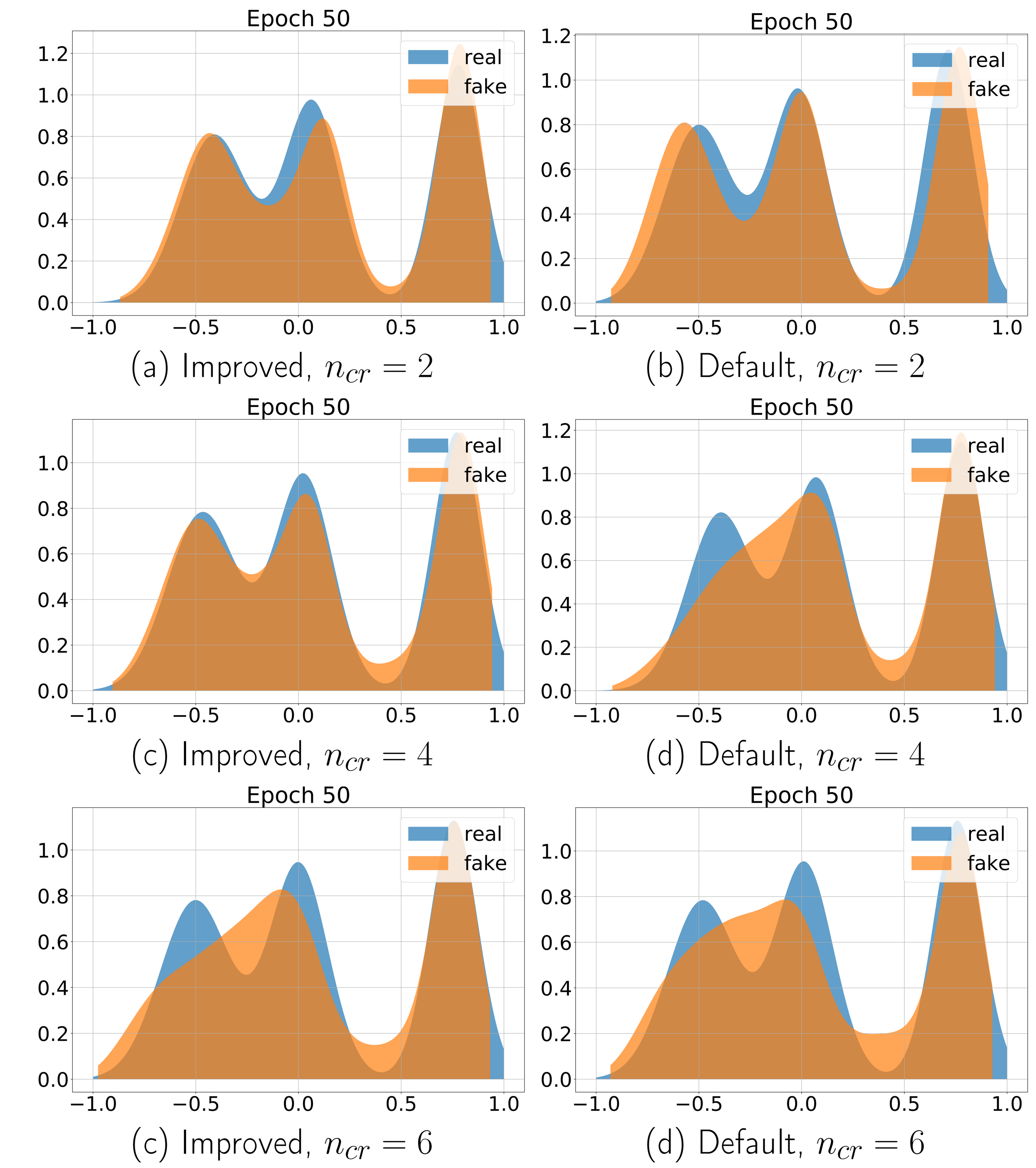
Experiments

1. **Dataset**: a multimodal distribution. Let $\xi \sim \mathcal{N}(\mu_1, \sigma_1), \eta \sim \mathcal{N}(\mu_2, \sigma_2), \zeta \sim \mathcal{N}(\mu_3, \sigma_3)$. The task is to learn generator to sample from real distribution $\xi + \eta + \zeta \sim \mu_0$. Distribution and its parameters are shown in the figure 1
2. **Loss function**: according to [2] WGAN loss was chosen as the loss function
$$J(\mu_0, \mu_\theta) = \max_{\Phi_w} \mathcal{J}(\mu_\theta, \Phi_w) = \max_{\|\Phi_w\|_{Lip} \leq 1} \mathbb{E}_{x \sim \mu_0}[\Phi_w(x)] - \mathbb{E}_{x \sim \mu_\theta}[\Phi_w(x)],$$
where Φ_w is a discriminator, $\|\Phi_w\|_{Lip} \leq 1$ implies that a function is Lipschitz with constant 1, μ_θ is the distribution of samples produced by generator and μ_0 is real data distribution.
3. **Evaluation**: to measure quality of generator p-value of Kolmogorov–Smirnov test was chosen.
4. **Architecture**: shown on the following plot:



(a) Generator (b) Discriminator default (c) Discriminator improved

Figure 3: Neural networks architecture



(a) Improved, $n_{cr} = 2$ (b) Default, $n_{cr} = 2$
(c) Improved, $n_{cr} = 4$ (d) Default, $n_{cr} = 4$
(c) Improved, $n_{cr} = 6$ (d) Default, $n_{cr} = 6$

Figure 4: Comparison of distributions for default and improved GAN with respect to different values of n_{cr}

Conclusion

The following conclusions were drawn from the results obtained:

1. The investigated parameter n_{cr} has a strong influence on learning default GAN and improved GAN. Spectral normalization limits the norm of the gradient, and in its absence forces gradients to explode, which leads to overfitting of the discriminator. As a result, there is a gradient saturation, and stopping the learning process. In those experiments where spectral normalization is involved, the overfitting of the discriminator is controlled, as a result, GAN generates a more plausible distribution.
2. In our study, we found that an increase of n_{cr} results in a reduction in time for one learning epoch, although on the other hand, more epochs are required to train the generator.

References

[1] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017.

[2] C. Chu, K. Minami, and K. Fukumizu. Smoothness and stability in gans. *arXiv preprint arXiv:2002.04185*, 2020.

[3] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville. Improved training of wasserstein gans. In *Advances in neural information processing systems*, pages 5767–5777, 2017.

[4] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida. Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*, 2018.

[5] Y. Yoshida and T. Miyato. Spectral norm regularization for improving the generalizability of deep learning. *arXiv preprint arXiv:1705.10941*, 2017.

[6] B. Zhou and P. Krähenbühl. Don't let your discriminator be fooled. 2018.