Many Paths to Equilibrium: GANS Do Not Need To Decrease a Divergence at Every Step

[1]Aniruddha Anand Damle - 1222585013
[1]Prakriti Biswas - 1222851266
[1]Varun Rahul Lale - 1219565061

[1]School of Electrical, Computing and Energy Engineering
Ira A. Fulton Schools of Engineering
Arizona State University
March 02, 2022

Motivation

- Problem: Several variants of the GAN training process have been proposed. Different variants of GANs have been interpreted as approximately minimizing different divergences or distances between p_{data} and p_{model}. However, it has been difficult to understand whether the improvements are caused by a change in the underlying divergence or the learning dynamics.

- Importance: The paper aims to find an optimal approach to train GANs to minimize divergence between the training distribution and the model distribution.

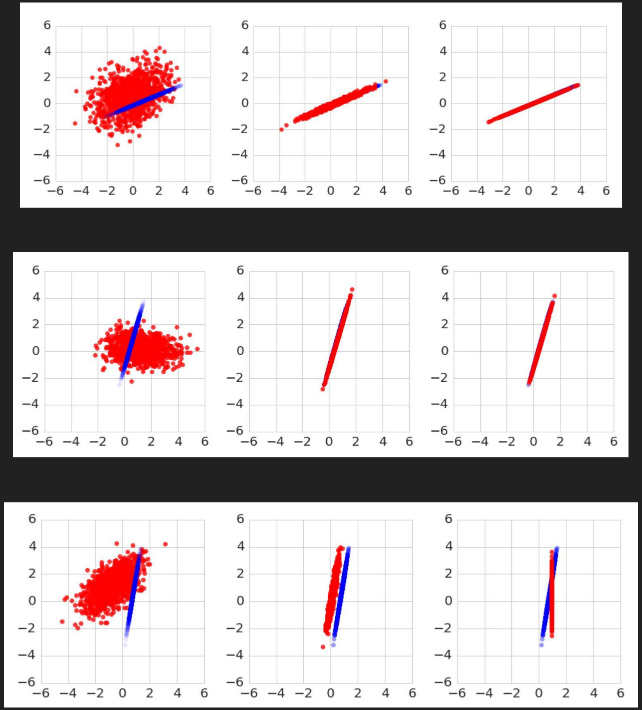

Fig.1: (a) NS-GAN (b) GAN-GP (c) DRAGAN_NS training at 0, 5000 and 10000 steps

Problem Statement

- Discriminator cost function for Minimax and Non-saturating training:
    - $J^{(D)}(D,G) = - E_{x\sim pdata}[\log D(x)] - E_{z\sim pz}[\log(1 - D(G(z)))]$
- Generator cost function:
    - Minimax: $J^{(G)}(G) = E_{z\sim pz}\log[1 - D(G(z))]$
    - Non-saturating: $J^{(G)}(G) = - E_{z\sim pz}\log[D(G(z))]$
- Wasserstein Gan:
    - Discriminator cost function: $W^{(D)}(D,G) = E_{x\sim pdata}[D(x)] - E_{z\sim pz}[D(G(z))]$
    - Generator cost function: $W^{(G)} = - W^{(D)}(D,G)$
- Deep Regret Analytic GAN (DRAGAN):
    - Discriminator cost function: $J^{\sim(D)}(D,G) = - E_{x\sim pdata}[\log D(x)] - E_{z\sim pz}[\log(1 - D(G(z)))] + \lambda E_{\dot{X}\sim px}[(||\nabla_{\dot{X}} D(\dot{X})||_2 - 1)^2]$
    - Generator cost function: $J^{(G)}(G) = - E_{z\sim pz}\log[D(G(z))]$

Here, the variables are defined as follows:
J = cost function
D = discriminator network
G = generator network
Pdata = true distribution of data
Pmodel = probability distribution of outcome of model data
Pz = probability distribution of noise

Literature Review

- Non-Saturating GANs, MINIMAX GANs, WGAN and DRAGAN-NS [1]
- Bridging the Gap Between f-GANs and Wasserstein GANs [2]
- Semi-Amortized Generative Modeling by Exploring Energy of the Discriminator [3]

References:
[1] W. Fedus, M. Rosca, B. Lakshminarayanan, A. M. Dai, S. Mohamed, and I. Goodfellow, “Many Paths to Equilibrium: GANs Do Not Need to Decrease a Divergence At Every Step,” Oct. 2017, doi: 10.48550/arXiv.1710.08446.
[2] J. Song and S. Ermon, “Bridging the Gap Between f-GANs and Wasserstein GANs,” in International Conference on Machine Learning, Nov. 2020, pp. 9078–9087. Accessed: Mar. 04, 2022. [Online]. Available: https://proceedings.mlr.press/v119/song20a.html
[3] Y. Song, Q. Ye, M. Xu, and T.-Y. Liu, “Discriminator Contrastive Divergence: Semi-Amortized Generative Modeling by Exploring Energy of the Discriminator,” Apr. 2020, doi: 10.48550/arXiv.2004.01704.

Our Plan

We plan to model and formulate the problem in the given paper. Our plan is to:

1. Understand GANs and its variants
2. Compare different variants using mathematical formulation
3. Testing gradient penalties and divergence metrics for different GANs
4. Evaluating performance metrics for different data sets using techniques such as Color MNIST, CelebA and CIFAR 10.

Project Proposal

MANY PATHS TO EQUILIBRIUM: GANS DO NOT NEED TO DECREASE A DIVERGENCE AT EVERY STEP

Prakriti Biswas, Aniruddha Anand Damle, Varun Lale

Anh-Duong Nguyen

Mar 4, 2022

The authors of the paper[1] conduct several experiments to assess whether the improvements associated with new GAN methods are due to the reasons cited in their design motivation. They perform a comprehensive study of GANs on simplified, synthetic tasks for which the true p_{data} is known and the relevant distances are straightforward to calculate, to assess the performance of proposed models against baseline methods. The authors compare the newer methods of GANs with the one originally introduced by "Goodfellow (et. al. 2014)"[2] in terms of the improvements made by the new GANs. They also evaluate GANs using several independent evaluation measures on real data to better understand new approaches.

A generative adversarial network is a class of machine learning frameworks designed by Ian Goodfellow and his colleagues in June 2014. Given a training set, this technique learns to generate new data with the same statistics as the training set. Generative adversarial networks (GANs) do not minimize a single training criterion. Unlike other generative models, the data distribution is learned via a game between a generator (the generative model) and a discriminator (a teacher providing training signal) that each minimizes their own cost. GANs are designed to reach a Nash equilibrium at which each player cannot reduce their cost without changing the other players' parameters.

The gradient penalty is an idea that enforces the constraint such that the gradient of the discriminator's output has a unit norm with respect to the input. The models are compared in such a way that we keep gradient penalty the same and take different evaluations for underlying adversarial losses and vice versa. The original paper focuses on demonstrating that gradient penalties designed in the divergence minimization framework improve Wasserstein GANs, or justified from a game theory perspective, improve minimax GANs. The authors have also demonstrated experiments to improve the non-saturating GAN on both synthetic and real data. We may observe that non-saturating GANs can fit problems that cannot be fit by Jensen-Shannon divergence minimization. The authors analyze how the gradient penalties in Wasserstein GANs and DRAGANs are effective outside the defining scope and assess if the exact form of gradient penalty matters.

In conclusion, the authors provide empirical counterexamples to the view of GAN training as divergence minimization. Specifically, they demonstrate that GANs can learn distributions in situations where the divergence minimization point of view predicts they would fail. They also show that gradient penalties motivated from the divergence minimization perspective are equally helpful when applied in other contexts in which the divergence minimization perspective does not predict they would be helpful. This contributes to a growing body of evidence that GAN training may be more usefully viewed as approaching Nash equilibria via trajectories that do not necessarily minimize a specific divergence at each step.

References:

[1] W. Fedus, M. Rosca, B. Lakshminarayanan, A. M. Dai, S. Mohamed, and I. Goodfellow, "Many Paths to Equilibrium: GANs Do Not Need to Decrease a Divergence At Every Step," Oct. 2017, doi: 10.48550/arXiv.1710.08446.

[2] I. J. Goodfellow *et al.*, "Generative Adversarial Networks," Jun. 10, 2014. Accessed: Mar. 04, 2022. [Online]. Available: http://arxiv.org/abs/1406.2661