Lecture 13: From GAN Training to Optimization of Differentiable Games

Start Recording!

Reminders

- Office Hours tomorrow with Ahmed and myself (11-12) [register here]
- Deadline pushed to this Saturday (11:59 PM AoE)
- Friday: 2h of Office hours with me
- We Are Done with GANs.
- Now: Game optimization (still motivated by GANs)

(I'm supposed to be a world expert on that topic)

More theory in the following Weeks

Extended abstract

- Submit it here: [link]
- Friday: no lecture -> Two hours of office hours instead. [register here] (to ask questions and get feedback before the deadline)
- Deadline postponed to Saturday 13th (11:59PM AoE)
- Penalty: 1 point (out of ten) per 24h late.
- Overleaf contains the latest guidelines.
- You should roughly indicate **the computational resources you will have access** to (cluster, personal GPU, collab,...)

References to read for this lecture:

1. Gidel, Gauthier, et al. "A variational inequality perspective on generative adversarial networks." ICLR 2019
2. Sion, Maurice. "On general minimax theorems." *Pacific Journal of mathematics* 8.1 (1958): 171-176

GANs Objectives:

Minimax GAN:

$$\min_\theta \max_\phi \mathbb{E}_{x\sim p_d}[\log \sigma(F_\phi(x))] + \mathbb{E}_{z\sim p_z}[\log(1-\sigma(F_\phi(G_\theta(z)))]$$

Wasserstein GAN:

$$\min_\theta \max_\phi \mathbb{E}_{x\sim p_d}[F_\phi(x)] - \mathbb{E}_{z\sim p_z}[F_\phi(G_\theta(z))]$$

GANs Objectives:

<u>Overall:</u>

$$\min_{\theta} \max_{\phi} \mathcal{L}(\theta, \phi)$$

<u>Goal:</u> solve this using gradient based methods.

<u>First Idea:</u> Gradient Descent-Ascent

$$\begin{cases} \theta_{t+1} = \theta_t - \eta \nabla_{\theta} \mathcal{L}(\theta_t, \phi_t) \\ \phi_{t+1} = \phi_t + \eta \nabla_{\phi} \mathcal{L}(\theta_t, \phi_t) \end{cases}$$

Smooth Games

Problem:

$$\min_{\theta} \max_{\phi} \mathcal{L}(\theta, \phi)$$

Nash Equilibrium: (θ^*, ϕ^*)

$$\mathcal{L}(\theta^*, \phi) \leq \mathcal{L}(\theta^*, \phi^*) \leq \mathcal{L}(\theta, \phi^*) \quad \forall(\theta, \phi)$$

Smooth Games

Standard Assumption: The payoff is convex-concave and differentiable

$$\mathcal{L}(\theta, \phi)$$

Consequence: We are at a Nash if and only if:

$$\nabla \mathcal{L}(\theta^*, \phi^*) = 0$$

Nash Equilibria always Exist for convex-concave payoffs

Theorem: [Sion 1958] If the payoff is convex-concave and U and V are convex and compact sets then,

$$\min_{\theta \in U} \max_{\phi \in V} \mathcal{L}(\theta, \phi) = \max_{\phi \in V} \min_{\theta \in U} \mathcal{L}(\theta, \phi)$$

Application: Prove Nash theorem for zero-sum two player games (Theorem 1.8 Lecture 2)

Let us start Simple

WGAN with **linear** discriminator and generator [Mescederer et al., 2018] (d=1)

$$\min_{\theta} \max_{\phi} \mathbb{E}_{x \sim p_d}[F_\phi(x)] - \mathbb{E}_{z \sim p_z}[F_\phi(G_\theta(z))]$$

$$\min_{\theta} \max_{\phi} \phi \cdot (\mathbb{E}_{x \sim p_d}[x] - \theta)$$

Task: Match the means!

Let us start Simple

For gradient based method, we will show we can equivalently study:

$$\min_{\theta} \max_{\phi} \phi \cdot \theta$$

Gradient Descent step

Gradient Ascent step

Exercice: Show that for that objective Gradient Descent Ascent updates are:

$$\begin{cases} \theta_{t+1} = \theta_t - \eta\phi_t \\ \phi_{t+1} = \phi_t + \eta\theta_t \end{cases}$$

Let us start Simple

For gradient based method, we will show we can equivalently study:

$$\min_{\theta} \max_{\phi} \phi \cdot \theta$$

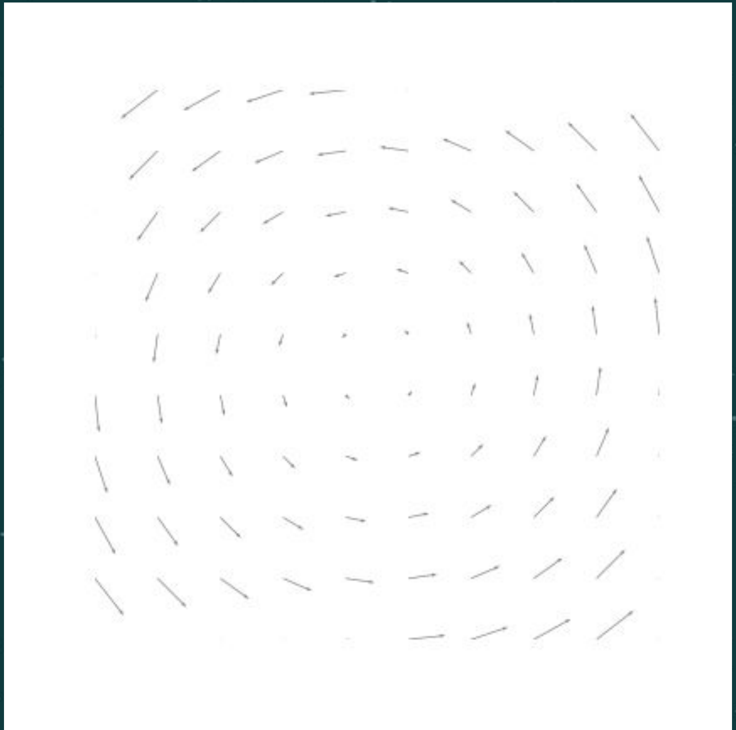
Gradient Descent step

Gradient Ascent step

Goal: Find the Nash Equilibrium of this Game.
Exercice: Prove that the NAsh Equilibrium of this game is (0,0)

First: some quick experiments

$$\begin{cases} \theta_{t+1} = \theta_t - \eta\phi_t \\ \phi_{t+1} = \phi_t + \eta\theta_t \end{cases}$$



What can we prove?

Seems like the iterates diverges:

$$\theta_t^2 + \phi_t^2 \geq \rho^t(\theta_0^2 + \phi_0^2) \quad \text{where} \quad \rho > 1$$

Exercice: Find \rho and prove it.

How do we implement Gradient Descent-Ascent in Practice?

```
theta = theta_0
phi = phi_0
For t= 1,...,N_ITER:
    theta = theta - eta * grad_theta (theta,phi)
    phi   = phi + eta * grad_phi(theta,phi)

Return (theta,phi)
```

Theta at step t+!1 !!!

$$\begin{cases} \theta_{t+1} = \theta_t - \eta \nabla_\theta \mathcal{L}(\theta_t, \phi_t) \\ \phi_{t+1} = \phi_t + \eta \nabla_\phi \mathcal{L}(\theta_t, \phi_t) \end{cases}$$

How do we implement Gradient Descent-Ascent in Practice?

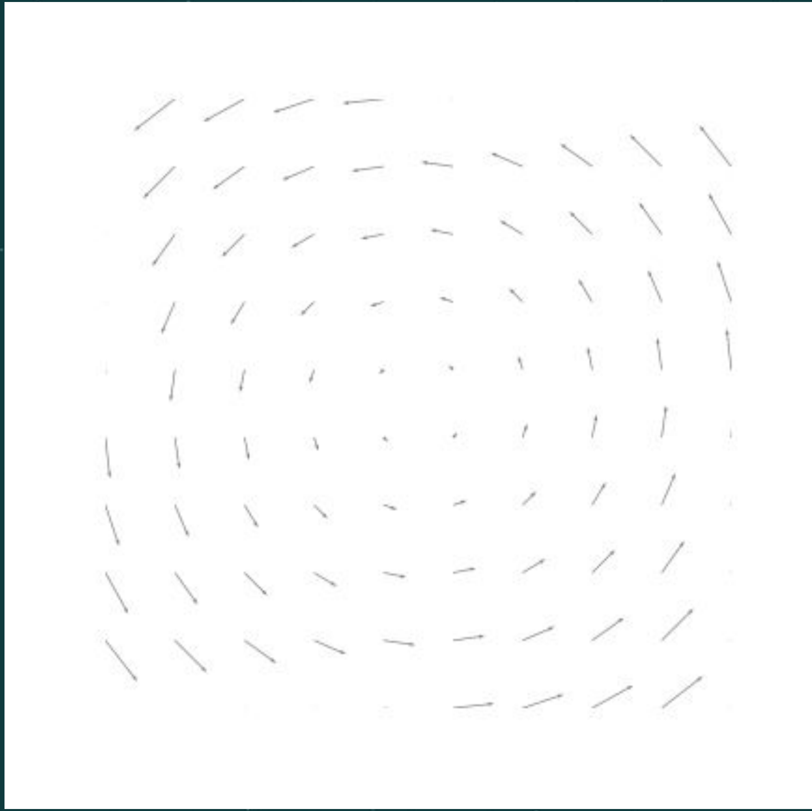
```
theta = theta_0
phi = phi_0
For t= 1,...,N_ITER:
        theta = theta - eta * grad_theta (theta,phi)
        phi   = phi + eta * grad_phi(theta,phi)

Return (theta,phi)
```

Theta at step t+!1 !!!

$$\begin{cases} \theta_{t+1} = \theta_t - \eta \nabla_\theta \mathcal{L}(\theta_t, \phi_t) \\ \phi_{t+1} = \phi_t + \eta \nabla_\phi \mathcal{L}(\theta_{t+1}, \phi_t) \end{cases}$$

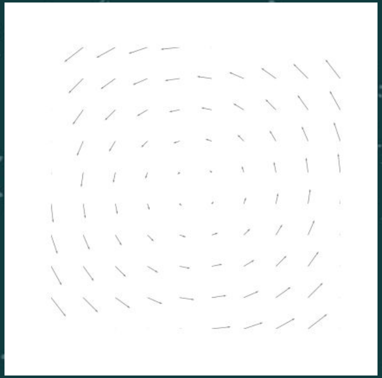
What about this One?

$$\begin{cases} \theta_{t+1} = \theta_t - \eta\phi_t \\ \phi_{t+1} = \phi_t + \eta\theta_{t+1} \end{cases}$$



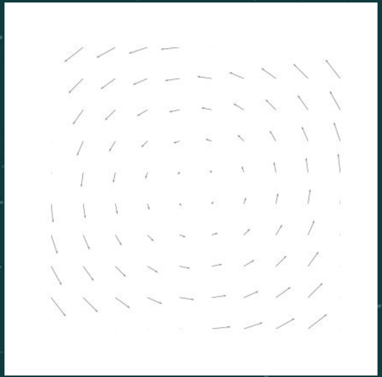
Summary

Simultaneous Gradient Descent-Ascent (Sim-GDA):



$$\begin{cases} \theta_{t+1} = \theta_t - \eta\phi_t \\ \phi_{t+1} = \phi_t + \eta\theta_t \end{cases}$$

Alternated Gradient Descent-Ascent (Alt-GDA)



t+1 here????

$$\begin{cases} \theta_{t+1} = \theta_t - \eta\phi_t \\ \phi_{t+1} = \phi_t + \eta\theta_{t+1} \end{cases}$$

Next Idea: Proximal Point Method

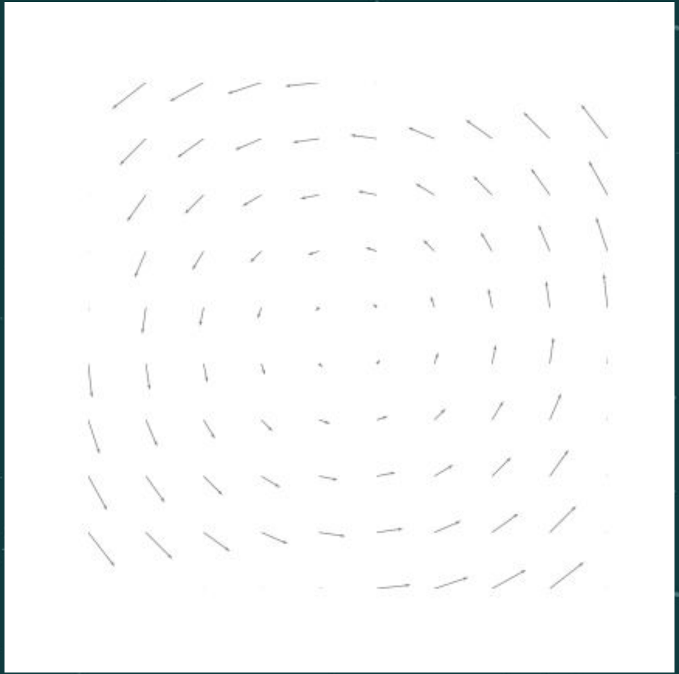
$$\begin{cases} \theta_{t+1} = \theta_t - \eta \phi_{t+1} \\ \phi_{t+1} = \phi_t + \eta \theta_{t+1} \end{cases}$$

Exercice: Show that

$$\theta_t^2 + \phi_t^2 \leq \rho^t (\theta_0^2 + \phi_0^2) \quad \text{where} \quad 0 < \rho < 1$$

Proximal Point Method

$$\begin{cases} \theta_{t+1} = \theta_t - \eta\phi_{t+1} \\ \phi_{t+1} = \phi_t + \eta\theta_{t+1} \end{cases}$$



Generalization of Proximal Point Method:

$$\begin{cases} \theta_{t+1} = \theta_t - \eta \nabla_\theta \mathcal{L}(\theta_{t+1}, \phi_{t+1}) \\ \phi_{t+1} = \phi_t + \eta \nabla_\phi \mathcal{L}(\theta_{t+1}, \phi_{t+1}) \end{cases}$$

Implicit Update: we need to solve a non-linear System

Conclusion: Not practical

Conclusion:

- Standard Gradient Methods **Fail to converge** on a simple 2D example
- Proximal point methods **Does Converge** (but not practical-> implicit)
- We need to find something explicit and that converge.
- See next lecture!

Useful Links and refs:

- Mescheder, Lars, Andreas Geiger, and Sebastian Nowozin. "Which training methods for GANs do actually converge?." *International conference on machine learning*. PMLR, 2018.
- Mescheder, Lars, Sebastian Nowozin, and Andreas Geiger. "The numerics of gans." NeurIPS (2017).
- Azizian, Waïss, et al. "A tight and unified analysis of gradient-based methods for a whole spectrum of differentiable games." *AISTATS*, 2020.
- Sion, Maurice. "On general minimax theorems." *Pacific Journal of mathematics* 8.1 (1958): 171-176.