Lecture 19: Stability and Equilibrium

Start Recording!

Reminders

- Office Hours tomorrow with Adrien (11-12AM)
- No Talks this Friday. (Non-working day)
- Last lecture on Smooth Games is Today
- Two Last lectures will be on empirical game theory, self-play and other interesting things.

Talk on StarCraft II by Wojciech M. Czarnecki

On Friday 16th (3rd author on the paper)

References for this lecture:

1. Daskalakis, Constantinos, and Ioannis Panageas. "The limit points of (optimistic) gradient descent in min-max optimization." arXiv preprint arXiv:1807.03907 (2018).
2. Mazumdar, Eric, Lillian J. Ratliff, and S. Shankar Sastry. "On gradient-based learning in continuous games." SIAM Journal on Mathematics of Data Science 2.1 (2020): 103-131. (Arxiv in 2018)
3. Berard, Hugo, et al. "A closer look at the optimization landscapes of generative adversarial networks." ICLR (2020).

Today: Stability of gradient based methods using spectral Analysis

Local Nash Equilibrium

Two player games (everything generalizes to more than 2).

Nash Equilibria:

$$\begin{cases} \theta^* \in \arg\min_{\theta \in \Theta} L_1(\theta, \phi) \\ \phi^* \in \arg\min_{\phi \in \Phi} L_2(\theta, \phi) \end{cases}$$

Two player games (everything generalizes to more than 2).

Local Nash Equilibria:

$$\begin{cases} \theta^* \in \arg\min_{\theta \in B(\theta^*, \delta)} L_1(\theta, \phi) \\ \phi^* \in \arg\min_{\phi \in B(\phi^*, \delta)} L_2(\theta, \phi) \end{cases}$$

Local Neighborhoods

Variational Inequality Perspective

We only 'care' about the gradient-based updates, i.e., the vector field:

$$F(\theta, \phi) = \begin{pmatrix} \nabla_\theta L_1(\theta, \phi) \\ \nabla_\phi L_2(\theta, \phi) \end{pmatrix}$$

Previous plots. We represented the joint space (θ_t, ϕ_t)

More compact formalism:

$$\omega = (\theta, \phi)$$

Variational Inequality Perspective

Goal: Find a stationary (fixed) point of the vector field:

$$F(\omega^*) = 0$$

In zero sum game: Equivalent to find a point with 0 gradient for each player

If the game is convex concave: equivalent to find a Nash!

Beyond Convex-concave:
Only Necessary (First Order) Conditions!!!
What about Sufficient (Second Order) Condition?

Gradient ~~Descent~~ Method

Update rule:

$$\omega_{t+1} = \omega_t - \eta F(\omega_t)$$

Stability of a fixed point given by the spectrum of:

$$\nabla F(\omega^*)$$

Stability of Gradient Based Method

Let ω^* be a stationary point

<u>Property (from last time):</u>

When $\Re(\lambda) > 0\,,\ \lambda \in \nabla F(\omega^*)$

The gradient method (locally) converges to ω^*

Motivates

↓

<u>Definition:</u> A stationary point ω^* is said to be differentially locally stable only if

$$\Re(\lambda) > 0\,,\ \lambda \in \nabla F(\omega^*)$$

Stability of Gradient Based Method

Let ω* be a stationary point

Property

When ℜ

The gradi

Question (Miranda): Is there a difference between "stable points" and "limit points" ? or do they refer to the same thing?

A: same thing!

Motivates

Definition: A stationary point ω* is said to be differentially locally stable only if

$$\Re(\lambda) > 0\,,\ \lambda \in \nabla F(\omega^*)$$

What about Nash Equilibrium???

$$\begin{cases} \theta^* \in \arg \min\limits_{\theta \in B(\theta^*, \delta)} L_1(\theta, \phi) \\ \phi^* \in \arg \min\limits_{\phi \in B(\phi^*, \delta)} L_2(\theta, \phi) \end{cases}$$

Necessary Stationary conditions: $\nabla_\theta L_1(\theta^*, \phi^*) = 0 \quad \text{and} \quad \nabla_\phi L_2(\theta^*, \phi^*) = 0$

Sufficient 2nd order conditions: $\nabla^2_\theta L_1(\theta^*, \phi^*) \succ 0 \quad \text{and} \quad \nabla^2_\phi L_2(\theta^*, \phi^*) \succ 0$

Sufficient Condition For a Local Nash

Assume ω^* is a stationary point:

$$\nabla F(\omega^*) = \begin{pmatrix} \nabla_\theta^2 L_1(\theta^*, \phi^*) & \nabla_\phi \nabla_\theta L_1(\theta^*, \phi^*) \\ \nabla_\theta \nabla_\phi L_2(\theta^*, \phi^*) & \nabla_\phi^2 L_2(\theta^*, \phi^*) \end{pmatrix}$$

<u>Definition:</u> Differentiable Nash Equilibrium

$$\nabla_\theta^2 L_1(\theta^*, \phi^*) \succ 0 \quad \text{and} \quad \nabla_\phi^2 L_2(\theta^*, \phi^*) \succ 0$$

Conclusion

Interaction term

$$\nabla F(\omega^*) = \begin{pmatrix} \nabla^2_\theta L_1(\theta^*, \phi^*) & \nabla_\phi \nabla_\theta L_1(\theta^*, \phi^*) \\ \nabla_\theta \nabla_\phi L_2(\theta^*, \phi^*) & \nabla^2_\phi L_2(\theta^*, \phi^*) \end{pmatrix}$$

Differentiable Nash Equilibrium

$$\nabla^2_\theta L_1(\theta^*, \phi^*) \succ 0$$
$$\nabla^2_\phi L_2(\theta^*, \phi^*) \succ 0$$

No interaction!

Locally differentially stable stationary point

$$\Re(\lambda) > 0\,,\ \lambda \in \nabla F(\omega^*)$$

Interaction Matters!

Conclusion

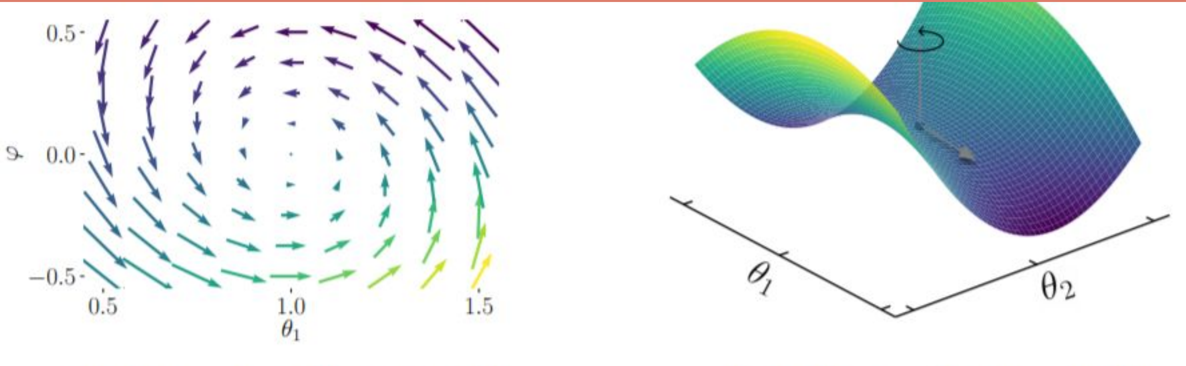
Interaction term

$\nabla F(\omega^*)$ ... $, \phi^*)$ ... $\phi^*)$

Question (Amit): What is the intuition of a stable fixed point compared to the intuition of a Nash Eq?

A: Rotating saddle may become stable!

For more see [Berard et al. 2020]



(a) 2D projection of the vector field.

(b) Landscape of the generator loss.

Differentiable ... stationary

$\nabla^2_\theta L$ ...

$\nabla^2_\phi L$ ...

... $F(\omega^*)$

No intera...

Differentiable Equilibrium

Exercice: Find a (2 player 0-sum) game that has a Nash equilibrium but **no Differentiable Nash Equilibrium!**

Hints in a skipped slide.

Question: (Amit) Is differentiable Nash Eq the same as Nash Eq?
A: No, But close

Differentiable Equilibrium

Exercice: Find a (2 player 0-sum) game that has a Nash equilibrium but **no Differentiable Nash Equilibrium!**

Example 1: $$L(\theta, \phi) = \theta \cdot \phi$$

Example 2: $$L(\theta, \phi) = \theta^2 - \theta \cdot \phi$$

Example 3: $$L(\theta, \phi) = \theta^2 - \theta \cdot \phi - \phi^2$$

Differentiable Equilibrium

Exercice: Find a (2 player 0-sum) game that has a Nash equilibrium but **no Differentiable Equilibrium!**

Conclusion:
weaker notion of Equilibria.
Easier do deal with (only related to eigenvalues)

Zero-Sum Case: $L_1 = -L_2$

$$\nabla^2_\theta L_1(\omega^*) = S_1$$

$$\nabla_\phi \nabla_\theta L_1(\theta^*, \phi^*) = -\nabla_\theta \nabla_\phi L_2(\theta^*, \phi^*)^\top = B$$

$$\nabla^2_\phi L_2(\theta^*, \phi^*) = S_2$$

$$\nabla F(\omega^*) = \begin{pmatrix} S_1 & B \\ -B^\top & S_2 \end{pmatrix}$$

Zero-Sum Case: $L_1 = -L_2$

$\nabla^2_\theta L$ …

$\nabla_\phi \nabla$ … $= B$

$\nabla^2_\phi L$ …

Exercice: What is

$$\nabla F(\omega^*) = \begin{pmatrix} \nabla^2_\theta L_1(\theta^*, \phi^*) & \nabla_\phi \nabla_\theta L_1(\theta^*, \phi^*) \\ \nabla_\theta \nabla_\phi L_2(\theta^*, \phi^*) & \nabla^2_\phi L_2(\theta^*, \phi^*) \end{pmatrix}$$

For the bilinear game: $\min_\theta \max_\phi \theta^\top B \phi$

$$\nabla F(\omega^*) = \begin{pmatrix} S_1 & B \\ -B^\top & S_2 \end{pmatrix}$$

Does interaction Matter?

$$\nabla F(\omega^*) = \begin{pmatrix} S_1 & B \\ -B^\top & S_2 \end{pmatrix}$$

Differentiable Nash Equilibrium

$$S_1 \succ 0$$

$$S_2 \succ 0$$

No interaction!

⇒

Locally differentially stable stationary point

$$\Re(\lambda) > 0\,,\ \lambda \in \nabla F(\omega^*)$$

Interaction Matters!

Does interaction Matter?

$$\nabla F(\omega^*) = \begin{pmatrix} S_1 & B \\ -B^\top & S_2 \end{pmatrix}$$

Differentiable

S_1

S_2

No interaction!

ry point

$\nabla F(\omega^*)$

Interaction Matters!

<u>Exercice:</u>

1. Try to prove this Implication!
2. Try to Find an example of Stationary point that is **not** a Differentiable Nash.

Conclusion: Zero-Sum Game

$$\nabla F(\omega^*) = \begin{pmatrix} S_1 & B \\ -B^\top & S_2 \end{pmatrix}$$

Differentiable Nash Equilibrium

$$S_1 \succ 0$$

$$S_2 \succ 0$$

Locally differentially stable stationary point

$$\Re(\lambda) > 0\,,\ \lambda \in \nabla F(\omega^*)$$

Non-Zero Sum Games

Just Because A is not $-B^T$

$$\nabla F(\omega^*) = \begin{pmatrix} S_1 & \boxed{B} \\ \boxed{A} & S_2 \end{pmatrix}$$

Dif
Equ

$S_1 \succ 0$

$S_2 \succ 0$

oint

$\Re(\lambda) > 0 \,,\ \lambda \in \nabla F(\omega^*)$

<u>Exercice:</u>

1. Try to Find an example of Differentiable Nash that is **not** a locally stable Stationary Point.

What about ExtraGradient and Optimistic Methods?

Previous Stability: GD was the reference

ExtraGradient:

$$\omega_{t+1} = \omega_t - \eta \underbrace{F(\omega_t - \eta F(\omega_t))}_{F_\eta(\omega_t)}$$

Q: Can we get, More details on this

A: See Jamboard and [Azizian et al. 2020]

Same thing

Same conditions but on $\nabla F_\eta(\omega^*)$

Stability of ExtraGradient:

The locally stable stationary points of EG are:

$$\Re(\lambda) > 0\,,\ \forall \lambda \in \nabla F_\eta(\omega^*)$$

Last thing to do:

$$\nabla F_\eta(\omega) = (I_d - \eta \nabla F(\omega))\nabla F(\omega - \eta F(\omega))$$
$$\nabla F_\eta(\omega^*) = (I_d - \eta \nabla F(\omega^*))\nabla F(\omega^*)$$

Stability of ExtraGradient:

The locally stable stationary points of EG are:

<u>Proposition:</u> A stationary point ω^* is a locally stable points of EG iif

$$\underbrace{\Re(\lambda) - \eta\Re(\lambda)^2}_{\text{Positive for small enough eta}} + \underbrace{\eta\Im(\lambda)^2}_{\text{Positive even when the real part is 0!!!}} > 0 \,, \; \forall \lambda \in \nabla F(\omega^*)$$

Positive for small enough eta

Positive even when the real part is 0!!!

Question: (Jonathan) Why is this set larger????
(Elio) What about the result in Daskalakis et al.
Does the case of imaginary eigenvalues apply at all in this case?

EG is More Stable

Locally Differentially Stable Points of EG

(for small enough eta)

$$\Re(\lambda) - \eta\Re(\lambda)^2 + \eta\Im(\lambda)^2 > 0\,,\ \forall\lambda \in \nabla F(\omega^*)$$

Locally Differentially Stable Stationary Point

$$\Re(\lambda) > 0\,,\ \lambda \in \nabla F(\omega^*)$$

EG is More Stable

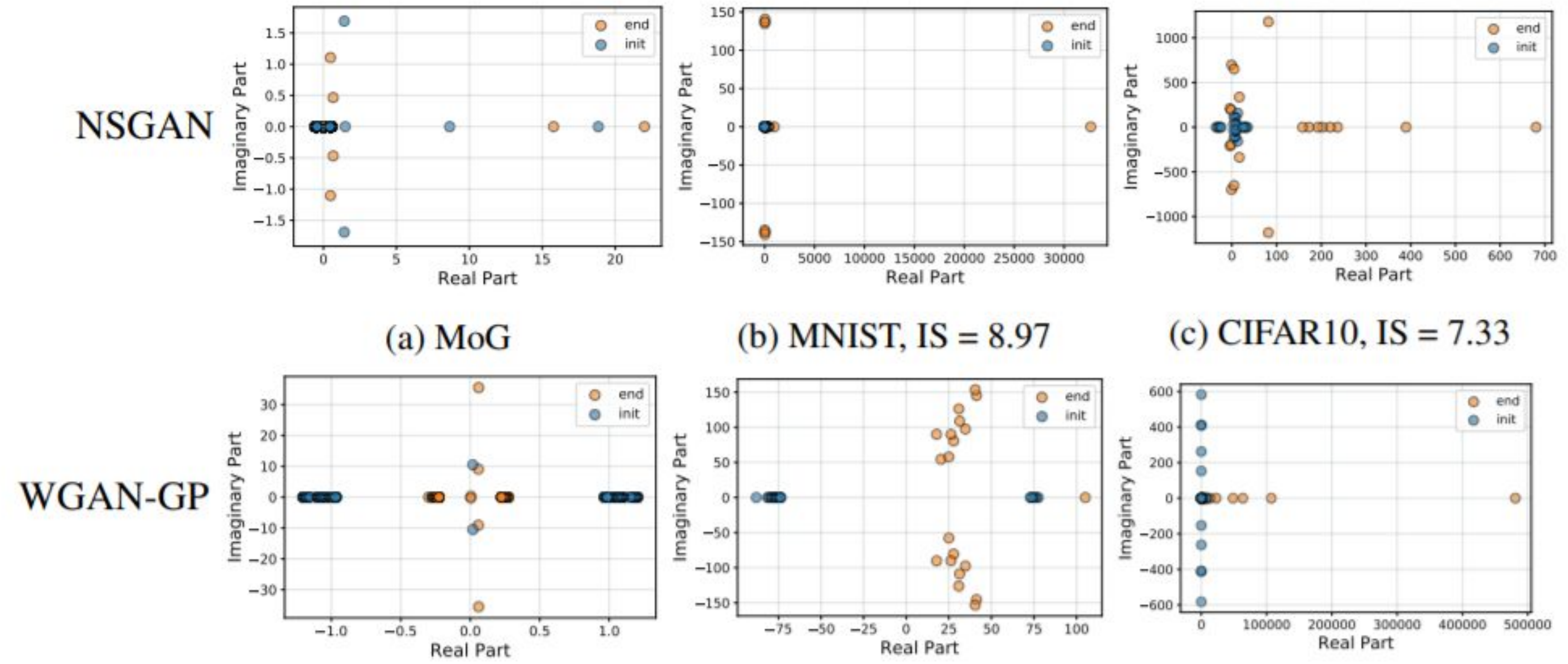
<u>Exercice:</u>

Show that the vector field of the Bilinear game is Stable for EG but not for the Gradient method.

Locally Differentially Stable Stationary Point

$$\Re(\lambda) > 0\,,\ \lambda \in \nabla F(\omega^*)$$

What about GANs in practice?

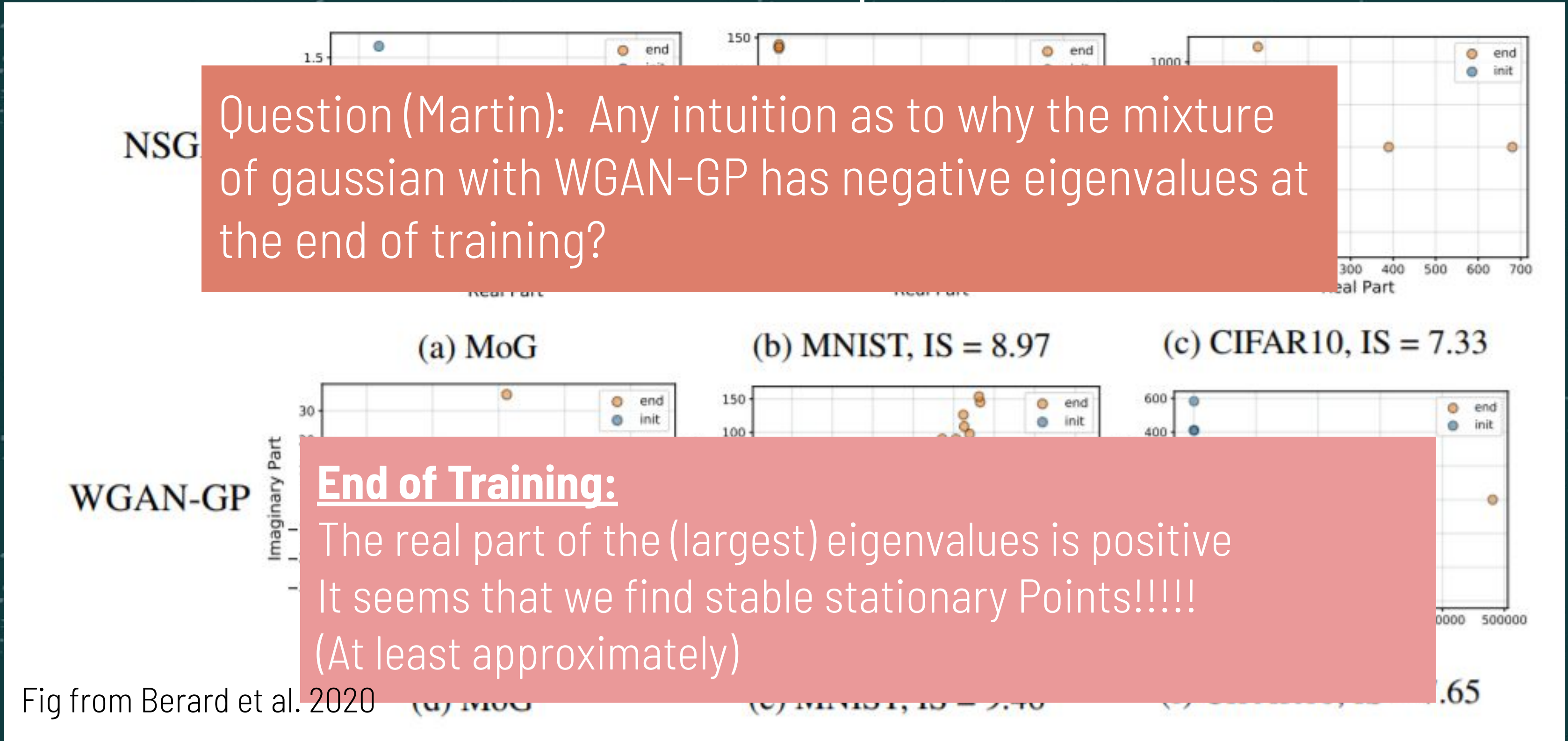


(a) MoG (b) MNIST, IS = 8.97 (c) CIFAR10, IS = 7.33

(d) MoG (e) MNIST, IS = 9.46 (f) CIFAR10, IS = 7.65

Fig from Berard et al. 2020

What about GANs in practice?

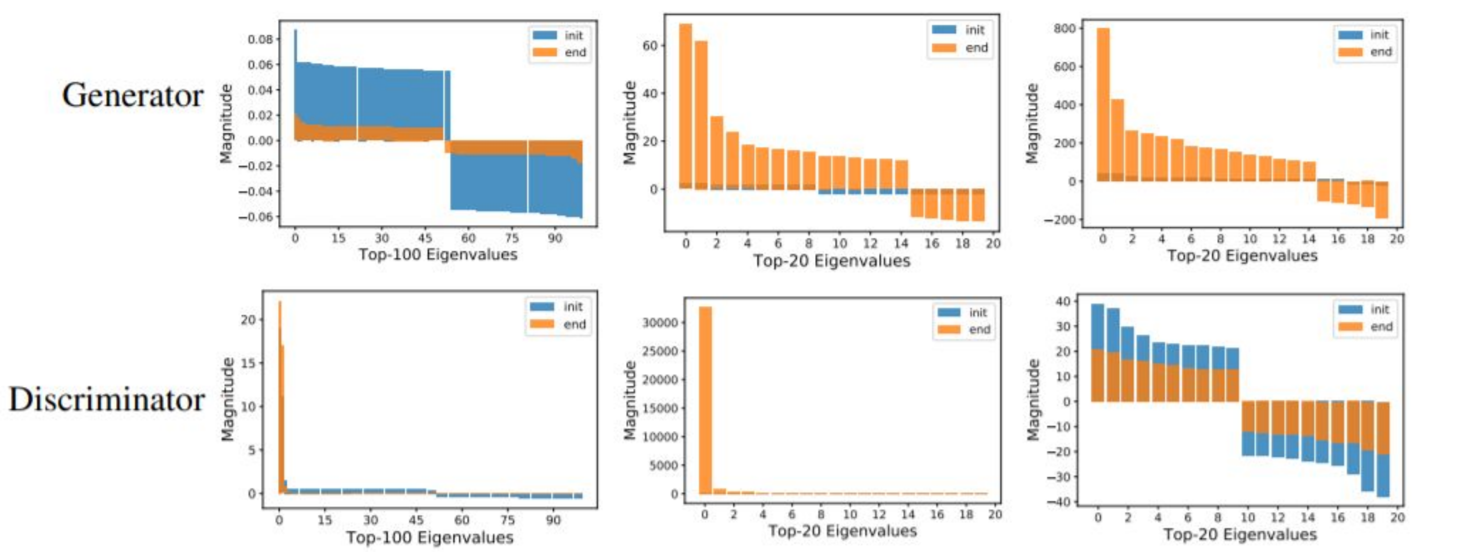


Question (Martin): Any intuition as to why the mixture of gaussian with WGAN-GP has negative eigenvalues at the end of training?

End of Training:
The real part of the (largest) eigenvalues is positive
It seems that we find stable stationary Points!!!!!
(At least approximately)

Fig from Berard et al. 2020

What about GANs in practice? $\nabla F(\omega^*) = \begin{pmatrix} S_1 & B \\ A & S_2 \end{pmatrix}$



: 7.33

Figure 5: NS… tude) in
descending or… um but a
saddle point (… s). Thus
the training al…

End of Training:

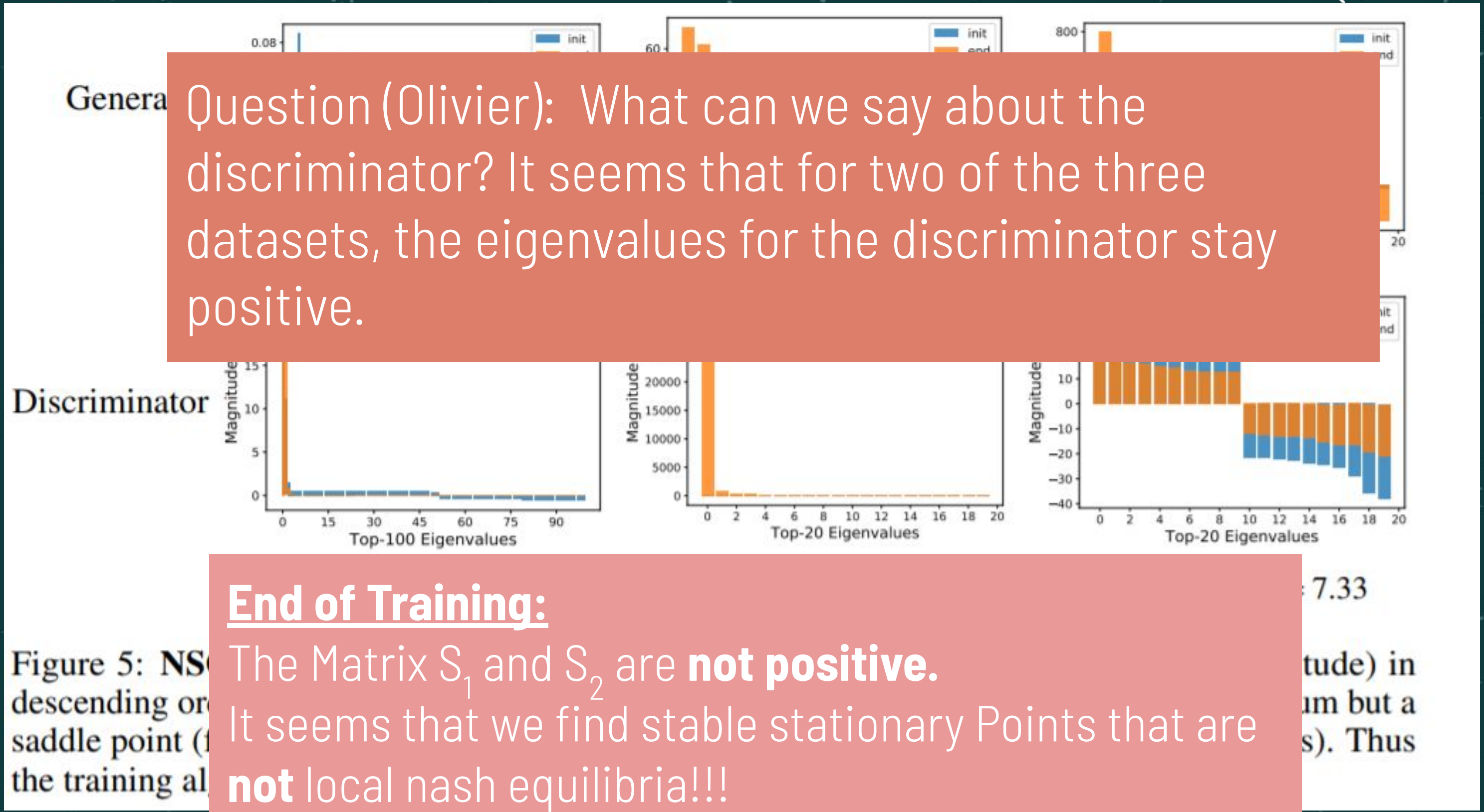
The Matrix S_1 and S_2 are **not positive.**

It seems that we find stable stationary Points that are **not** local nash equilibria!!!

Fig from Bera

What about GANs in practice?

$$\nabla F(\omega^*) = \begin{pmatrix} S_1 & B \\ A & S_2 \end{pmatrix}$$



Question (Olivier): What can we say about the discriminator? It seems that for two of the three datasets, the eigenvalues for the discriminator stay positive.

End of Training:

The Matrix S_1 and S_2 are **not positive.**

It seems that we find stable stationary Points that are **not** local nash equilibria!!!

Fig from Berard et al. 2020

Conclusion

- Can analyze stability using $\nabla F(\omega^*)$

- Can use the block decomposition :

$$\nabla F(\omega^*) = \begin{pmatrix} \nabla^2_\theta L_1(\theta^*, \phi^*) & \nabla_\phi \nabla_\theta L_1(\theta^*, \phi^*) \\ \nabla_\theta \nabla_\phi L_2(\theta^*, \phi^*) & \nabla^2_\phi L_2(\theta^*, \phi^*) \end{pmatrix}$$

  To define Differentiable Nash Equilibrium

- Slightly weaker notion (Sufficient second order conditions) of stability/Equilibrium.

- The optimization method change the stability conditions
- (for instance EG stabilizes the bilinear game)

Conclusion

- Can analyze stability using $\nabla F(\omega^*)$
- Can u

To de

- Slight
stabi

- The optimization method change the stability conditions
- (for instance EG stabilizes the bilinear game)

Question (Justine): Since in practice we do not reach the Nash Equilibrium, but the models still work well, I was wondering if it is worth trying to reach the Nash Equilibrium anyways?

Answer:

- We do not know
- Methods to only reach Nash equilibrium:
  - Adolphs et al. (2018); Mazumdar et al. (2019)
  - Use **Second order information**.