IFT 6756 - Lecture 19
(Stability and Equilibrium)

This version of the notes has not yet been thoroughly checked. Please report any bugs to the scribes or instructor.

Scribes **Instructor:** Gauthier Gidel
Winter 2021: [Bharath Govindaraju, Nehal Pandey]

1 Summary

In the previous lecture we covered spectral analysis tool to analyze the local convergence of games. We saw that for convergence we need $\Re(\lambda) > 0, \forall\lambda \in \nabla F(\omega^*)$ where λ is the eigen value of the Jacobian of the vector field F at the stationary point ω^* . In this lecture we will study the stability of gradient based methods using spectral analysis tools developed in the last lecture. Here we study two player games but the idea generalizes to more than 2 players. Goal is to connect the three notions of equilibrium: Local Nash that jointly minimizes the two Nashes$^{2.1}$, Differentiable Nash which is a stronger notion of local Nash and Locally Stable Stationary Points(LSSP) [1] which exhibits a dynamic equilibrium at a point which may not be a local Nash and which looks like a saddle point. We will look at these in more detail in the following sections.

2 Notions of Equilibrium

2.1 Local Nash Equilibrium(LNE)

Nash equilibrium : $\begin{cases} \theta^* \in \underset{\theta\in\Theta}{argmin} L_1(\theta, \phi^*) \\ \phi^* \in \underset{\phi\in\Phi}{argmin} L_2(\theta^*, \phi) \end{cases}$

We have to find global minima for θ when ϕ^* is fixed w.r.t L_1 and global minima for ϕ when θ^* is fixed w.r.t L_2.

In local Nash, instead of having $\theta and \phi$ anywhere, i.e unbounded, we have them bounded in a local neighbourhood of θ^* and ϕ^* respectively. Here, we seek to jointly minimize the two Nashes.
It is stated in [1] that as the GAN's objectives are highly non-convex we cannot typically expect better than a local Nash Equilibrium.

Local Nash: $\begin{cases} \theta^* \in \underset{\theta\in B(\theta^*,\delta)}{argmin} L_1(\theta, \phi^*) \\ \phi^* \in \underset{\phi\in B(\phi^*,\delta)}{argmin} L_2(\theta^*, \phi) \end{cases}$

Where $B(\cdots, \delta)$ refers to a local neighbourhood and Nash is only locally true in this neighbourhood.

2.2 Differential Nash Equilibrium(DNE)

From [1] : For being a local Nash it is sufficient to be a differential Nash equilibrium.
A point, $\omega^* := (\theta^*,\phi^*)$ is a differential Nash equilibrium iff
$\|\nabla_\theta L_1(\omega^*)\| = \|\nabla_\phi L_2(\omega^*)\| = 0, \nabla^2_\theta L_1(\omega^*) \succ 0$ and $\nabla^2_\phi L_2(\omega^*) \succ 0$.

We see that this requires the Hessians to be Positive definite. Whereas for Local Nash, it is enough if Hessians are Positive Semidefinite. Differential Nash is a stronger notion of local Nash. However, being a Differential Nash Equilibrium is not necessary to be a Local Nash Equilibrium.

For a Local Nash Equilibrium defined in 2.1,
Necessary Stationary conditions: $\nabla_\theta L_1(\theta^*, \phi^*) = \nabla_\phi L_2(\theta^*, \phi^*) = 0$
Sufficient 2nd Order conditions: $\nabla_\theta^2 L_1(\theta^*, \phi^*) \succ 0$ and $\nabla_\phi^2 L_2(\theta^*, \phi^*) \succ 0$
From a Nash equilibrium perspective, we look at the individual losses, we care about minimizing the loss for each player and are interested in Hessians. For gradient dynamics we consider the Jacobian of the vector field in the next section.

2.3 Locally Stable Stationary Points(LSSP)

In order to take the interaction between the players into account which is missed in the LNE we jointly consider the parameters of both the players, θ and ϕ as a joint state $\omega := (\theta, \phi)$ which we had seen in the Variational Inequality Perspective. Here, we only care about the gradient updates, i.e the vector field which is a concatenation of each player's gradient. We are interested in the points we are attracted to when we are following the vector field F.

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

Goal is to find a stationary point ω^* of the vector field where

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

At this point we are not moving anymore when following the vector field. We care for such a point because in any game it's equivalent to finding a point with zero gradient for each player. If the game is convex-concave it is equivalent finding a Nash. We will later see the connection between LSSP and Nash while remembering that we are considering only local Nash.

Beyond convexity-concavity what are the sufficient conditions to converge to such a point ω^* ?
Necessary condition is $F(\omega^*) = 0$.

2.3.1 Gradient Method

Standard gradient update rule : $\omega_{t+1} = \omega_t - \eta F(\omega_t)$
Stability of the fixed point is given by the spectrum of the matrix $\nabla F(\omega*)$, the Jacobian of the vector field.
It is important to note that in minimization we look at the Hessian. If the Hessian is Positive Semidefinite we are at a local minima. In the case of games we look at the Jacobian of the vector field to see what are the sufficient conditions to get a stable point.

2.3.2 Stability of Gradient Based Method

From the results in the previous lecture we have seen that when the eigen values of the Jacobian have positive real part the gradient method will locally converge to ω^* i.e, when $\Re(\lambda) > 0, \lambda \in \nabla F(\omega^*)$ where λ is a eigen value of the Jacobian it is a sufficient condition to be attracted to $\omega*$ which motivates the following definition.

A stationary point ω^* is said to be differentially locally stable only if $\Re(\lambda) > 0, \lambda \in \nabla F(\omega^*)$

We will define all the points that are locally stable, all the points that are attracted by the gradient dynamics using the above definition. In the last lecture we saw that the above condition was sufficient for local convergence. Here, we are saying that such points are the locally stable ones as well.

Q: Is there a difference between "stable" and "limit" points ?
Gauthier: Same. If we start from the equilibrium, perturb the point by a small amount, we get back to the original point by following the gradient. It's a limit point if we start in a small enough neighbourhood. Notion of neighbourhood is important.

3 Local Nash and LSSP

Consider the Jacobian of the vector field, where $\omega* = (\theta^*, \phi^*)$ which is the matrix of interest to characterize stability

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

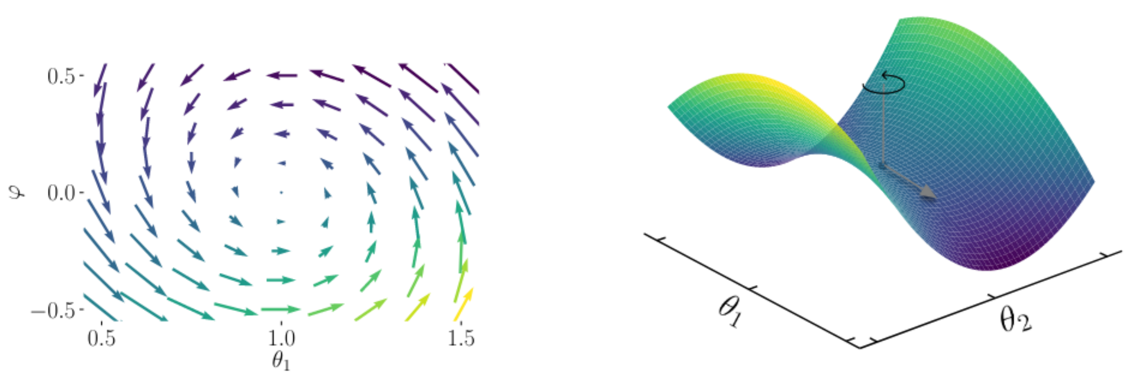
Here, the anti-diagonal terms, $\nabla_\phi \nabla_\theta L_1(\omega^*)$ and $\nabla_\theta \nabla_\phi L_2(\omega^*)$ capture the interaction between the parameters θ and ϕ of the two players. They are non-zero as the gradient of each of L_1 and L_2 are dependent on θ and ϕ . When the second player is changing the ϕ parameter it will change the loss of the first player and vice-versa. For stability w.r.t gradient dynamics we care about these interaction terms.
Diagonal terms have the Hessian of L_1 and L_2

For Differential Nash equilibrium in 2.2 we see that only the Hessians matter which should be positive definite and interaction is of no concern. We are only interested in minimizing individual losses.
For Locally Differentially Stable Stationary points as seen in 2.3.2 interaction matters as it depends on the eigen values, λ of the Jacobian of the vector field which is influenced by the anti-diagonal interaction terms .

3.1 Intuition

Q:What is the intuition of a stable fixed point compared to the intiution of Nash equilibrum?
Gauthier:To be at a local Nash we have to be at a local minima in each loss landscape when we fix the other parameter. As seen earlier a sufficient but not a necessary condition to be at local minima is for the Hessain of the Loss to be Positive definite.

For LSSP consider the following visualization from an example in [1]



(a) 2D projection of the vector field.

(b) Landscape of the generator loss.

Figure 1: Vector field(left) Loss landscape (right)

In the above figure, descent direction for Generator is controlled by the Discriminator parameter φ plotted in 1(a). When the generator follows the descent direction the discriminator updates it's parameter, rotating the saddle and causes the saddle point to remain stable. This phenomenon is called *dynamic stability*

We look at two perspectives - Joint vector field and Loss landscape.
Loss landscape is obtained by fixing the parameter at the equilibrium for one player and looking at the loss of the other player. This doesn't capture whether the point is stable for the former. Joint vector field captures the dynamics of the interaction between players.
Vector field plot in 1a is rotating around equilibrium and converging. If we look at the loss landscape there maybe a saddle. Hessian of the loss can tell us whether we are at local minima/maxima or at a saddle point. If the Hessian is

not Positive semiDefinite we may be at a saddle.

It's possible for a saddle point to be stable due to interaction between the players. Loss landscape of a player depends on the parameter of the other player. So, while the player2 parameter is changing, player1 loss landscape can change and turn the saddle point, i.e saddle point moves. If the the saddle turns at the right speed it can be stable. This is an idea from Physics seen in Pauli's trap.

Dynamic Equilibrium

A static saddle point is not stable. However, in the game, as seen in the previous paragraph, saddle is not static. Due the the dynamics of the game, loss landscape can move at the same time we are moving in it due to the interaction with the other player. For a landscape that is fixed, a saddle point may look unstable but it can become stable as landscape can change in a manner to make the saddle point stable. If the landscape rotates too quickly saddle point may become unstable.

Condition in 2.2 where $\Re(\lambda) > 0$ ensures that interactions are well balanced to make the point stable. Better intuition is obtained from the vector field to understand the dynamic equilibrium.

We can be at a saddle point for one of the loss function but we are not at Nash equilibrium as a saddle point is not a local minima.

Q: What leads to rotation of saddle point?

Gauthier: As the loss landscape of each player depends on the parameters of the other player, updating the parameters of one player causes the loss landscape of the other player to change. There is a stable equilibrium as, while one player is rotating the other player is converging to a saddle. While the latter converges it causes the former's landscape to rotate. This kind of interaction make the process stable.

Exercise: Find a 2 player zero-sum game that has a Nash but not Differential Nash equilibrium.

Gauthier:

Example 1:

$$L(\theta, \phi) = \theta.\phi$$

Example 2:

$$L(\theta, \phi) = \theta^2 - \theta.\phi$$

Example 3:

$$L(\theta, \phi) = \theta^2 - \theta.\phi - \phi^2$$

All 3 examples have a Nash at (0,0). To check for Differential Nash consider the Hessian which needs to be positive definite.

Example 1:

$F(\theta, \phi) = \begin{pmatrix} \nabla_\theta L(\theta, \phi) \\ -\nabla_\phi L(\theta, \phi) \end{pmatrix} = \begin{pmatrix} \phi \\ -\theta \end{pmatrix}$

$\nabla F(\theta, \phi) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, a constant

Diagonal elements of the Jacobian correspond to Hessians which are zer, thus are not Positive definite and thus not a Differential Nash.

Example 2:

$F(\theta, \phi) = \begin{pmatrix} \nabla_\theta L(\theta, \phi) \\ -\nabla_\phi L(\theta, \phi) \end{pmatrix} = \begin{pmatrix} 2\theta - \phi \\ \theta \end{pmatrix}$

$\nabla F(\theta, \phi) = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$, One of the Hessian is zero, thus not a Differential nash.

Example 3:

$F(\theta, \phi) = \begin{pmatrix} \nabla_\theta L(\theta, \phi) \\ -\nabla_\phi L(\theta, \phi) \end{pmatrix} = \begin{pmatrix} 2\theta - \phi \\ \theta + 2\phi \end{pmatrix}$

$\nabla F(\theta, \phi) = \begin{pmatrix} 2 & -1 \\ 2 & 1 \end{pmatrix}$, both the Hessians are positive thus this has a Differential Nash equilibrium.

3.2 Zero-Sum Case

For a Zero-Sum case we will see how the Differentiable Nash and LSSP are related.
Zero-Sum :

$$
\begin{aligned}
L_1 &= -L_2 \\
\nabla_\theta^2 L_1(\omega^*) &= S_1 \\
\nabla_\theta \nabla_\phi L_1(\omega^*) = -\nabla_\theta \nabla_\phi L_2(\omega^*)^T &= B \\
\nabla_\phi^2 L_2(\omega^*) &= S_2
\end{aligned}
$$

S_1, S_2 are symmetric matricies, we want them to be positive definite.

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

Rotation is due to anti-symmetric interaction. Complexity of games in terms of optimization emerge due to anti-symmetric part.
Exercise:What is the Jacobian of the vector field for the bilinear game $\underset{\theta}{min} \underset{\phi}{max} \theta^T B\phi$
$F(\theta, \phi) = (\begin{pmatrix} B\phi \\ -B^T\theta \end{pmatrix}$
$\nabla F(\theta, \phi) = \begin{pmatrix} 0 & B \\ -B^T & 0 \end{pmatrix}$
Symmetric part is 0, thus we not at Differential Nash.

In general games, symmetric and anti-symmetrics are not separated. Anti-diagonal elements may not be anti-symmetric and may count in symmetric part.
In zero-sum game case we have separation between symmetric and anti-symmetric part.

3.2.1 Does interaction matter?

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

Differential Nash Equilibrium		Locally differentially stable stationary point
$S_1 \succ 0$ $S_2 \succ 0$	$\Rightarrow$	$\Re(\lambda) > 0, \forall \lambda \in \nabla F(\omega^*)$
No interaction Static Equilibrium		Interaction Matters Dynamic Equilibrum

1. In zero sum game anti-diagonal elements are related as we have seen above, $B \;\; B^T$
2. In non-zero sum game which we will see later anti-diagonal elements may not be related, they can be arbitary and there can be no link between 2 losses

In the case of zero-sum game we can leverage the link between the two losses to show that Differential Nash is a Locally Differentially Stable Point. This is desirable as we can use gradients to find Nash. Here, all Nash equilibrium are stable for gradient dynamics.
Proof:

$$S = \begin{pmatrix} S_1 & B \\ -B^T & S_2 \end{pmatrix};$$
$$S_1 \succ 0, S_2 \succ 0$$
$$Thus, \Re(\lambda) > 0, \forall \lambda \in \nabla F(\omega^*)$$

Let $\begin{pmatrix} \jmath_1 \\ \jmath_2 \end{pmatrix} \in \mathbb{C}^n$ be an eigen vector

$$\begin{pmatrix} S_1 & B \\ -B^T & S_2 \end{pmatrix} \begin{pmatrix} \jmath_1 \\ \jmath_2 \end{pmatrix} = \lambda \begin{pmatrix} \jmath_1 \\ \jmath_2 \end{pmatrix}$$
$$S_1 \jmath_1 + B \jmath_2 = \lambda \jmath_1$$
$$-B^T \jmath_1 + S_2 \jmath_2 = \lambda \jmath_2$$

For a symmetric matrix S, $x^T S x > 0 \;\; \forall x \in \mathbb{R}^n$
Let
$\jmath^* = \begin{pmatrix} \bar{\jmath}_1 \\ \bar{\jmath}_2 \end{pmatrix}^T$ Thus, $\jmath^* S \jmath = \lambda \jmath^* \jmath = \lambda \|\jmath\|^2$
$\lambda \|\jmath\|^2 = \jmath^* \begin{pmatrix} S_1 & B \\ -B^T & S_2 \end{pmatrix} \jmath = \bar{\jmath}_1 S_1 \jmath_1 + \bar{\jmath}_2 S_2 \jmath_2 + \underbrace{\bar{\jmath}_1 B \jmath_2 - \bar{\jmath}_2 B^T \jmath_1}_{=0} = \bar{\jmath}_1 S_1 \jmath_1 + \bar{\jmath}_2 S_2 \jmath_2$
As we only care about $\Re(\lambda \|J\|^2)$, $\Re(\bar{\jmath}_1 B \jmath_2) = \Re(\overline{(\bar{\jmath}_1 B \jmath_2)}) = \Re(\bar{\jmath}_2 B \jmath_1)$

For S $\succ 0$, $\Re(\jmath^* S \jmath) > 0 \; \forall \jmath \in \mathbb{C}^n$
proof: $\Re((x - \imath y) S_1 (x + \imath y)) = \underbrace{x^T S_1 x}_{\in \mathbb{R}^n} + \underbrace{y_T S_1 y}_{\in \mathbb{R}^n} > 0$

Thus, $\Re(\lambda \|\jmath\|_2^2) = x^T S_1 x + y^T S_1 y + x^T S_2 x + y^T S_2 y > 0$
Thus, for a zero-sum game any Differential Nash is a Locally Differentially Stable Point.

Set of Differential Nash Equilibrium points $\subseteq$ Locally Differentially Stable Stationary points.

What matters for gradient is $\Re(\lambda)$, real part of eigen values. We want to know if gradient will converge to Differential Nash Equilibrium. In zero-sum case as $B = -B^T$ it is true. Ideally in zero-sum case we would like to converge only to Differential Nash to certain extent but we converge to a lot of other points including Differential Nash equilibrium point.

In the next section for non-zero sum case we will see that some Differential Nash are locally stable and some are not locally stable as anti-diagonal elements are not equal.

3.3 Non-Zero Sum Case

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

In non-zero sum games, some differentiable nash equilibrium are locally stable but some of them are not locally stable. This is because A is not equal to $-B^T$

Exercise: Try to find an example of differentiable Nash that is not locally stable stationary point.
Answer: Trick here is to take $S_1 = 1$ and $S_2 = 1$ we play with A and B in order to have eigen value that has a negative real part.

Question: What if we want to converge to only differentiable nash equilibrium?
Answer: Look at the second order information and check if the hessian is definite positive.

4 Extra gradient(EG) and optimistic methods

In the previous lecture we have seen that extragradient has the following update:

$$w_{t+1} = w_t - \eta F(w_t - \eta F(w_t))$$

Question: Can we get more details on the above equation.
Soln. By definition of extra gradient:

$$w_{t+1} = w_t - \eta F(w_t - \eta F(w_t))$$

Let $G(w_t) = F(w_t - \eta F(w_t))$
Now we can rewrite EG as:

$$w_{t+1} = w_t - \eta G(w_t)$$

[From previous lecture]
Conclusion: It looks like Gradient Method stability condition:

$$Re(\lambda) > 0; \forall \lambda \in Sp(\nabla G(w^*))$$

Now we just need to study $\nabla F_\eta(w^*)$ instead of $\nabla F(w_t - \eta F(w_t))$

5 Stability of Extra Gradient

The locally stable stationary points of E.G are:

$$\Re(\lambda) > 0; \forall \lambda \in \nabla F_\eta(w^*)$$

Things to do:

$$\nabla F_\eta(w) = (I_d - \eta \nabla F(w)) \nabla F(w - \eta F(w))$$

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

Preposition: A stationary point w^* is a locally stable point of E.G if and only if: $\Re(\lambda) - \eta \Re(\lambda)^2 + \eta \Im(\lambda)^2 > 0$; $\forall \lambda \in \nabla F(w^*)$
Proof: By definition, stable points of EG satisfy

$$\Re(\lambda) > 0; \forall \lambda \in \nabla F_\eta(w^*)$$

$$\nabla F_\eta(w^*) = \nabla F(w^*) - \eta \nabla F(w^*)^2$$

Theorem:

$$\forall A : (A + A^2) = (\lambda + \eta \lambda^2 | \lambda \in A)$$

Now we get:

$$\nabla F_\eta(w^*) = (\lambda + \eta \lambda^2 | \lambda \in F(w^*))$$

Stable for EG if:

$$\Re(\lambda + \eta \lambda^2) > 0; \forall \lambda \in \nabla F(w^*)$$

$$\Re(\lambda + \eta \lambda^2) = \Re(\lambda) - \eta (\Re(\lambda))^2 + \eta (\Im(\lambda))^2$$

Which therefore gives:

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

Now we see that imaginary part has a role in stability(for $\eta > 0$)
Bilinear games will be stable for EG because real part is zero and imaginary part is positive.
Question: Why is this set larger?
Ans: This is because we can take η -ve to make $\Re(\lambda) - \eta(\Re(\lambda))^2$ always positive, also where the real part is zero, imaginary part is not zero.

$$\Re(\lambda) > 0$$

$$\eta = \frac{1}{2\Re(\lambda)}$$

which gives

$$\Re(\lambda) - \eta(\Re(\lambda))^2 + \eta(\Im(\lambda))^2$$

$$= \Re(\lambda) - \frac{\Re(\lambda)}{2} + \frac{\Im(\lambda)^2}{2\Re(\lambda)}$$

$$= \frac{\Re(\lambda)}{2} + \frac{Im(\lambda)^2}{2\Re(\lambda)} > 0$$

So whenever real part is positive, "$\Re(\lambda) - \eta(\Re(\lambda))^2 + \eta(\Im(\lambda))^2$" is positive.
Conclusion: Any point stable with Gradient Descent is stable with Extra Gradient.

5.1 EG is more stable

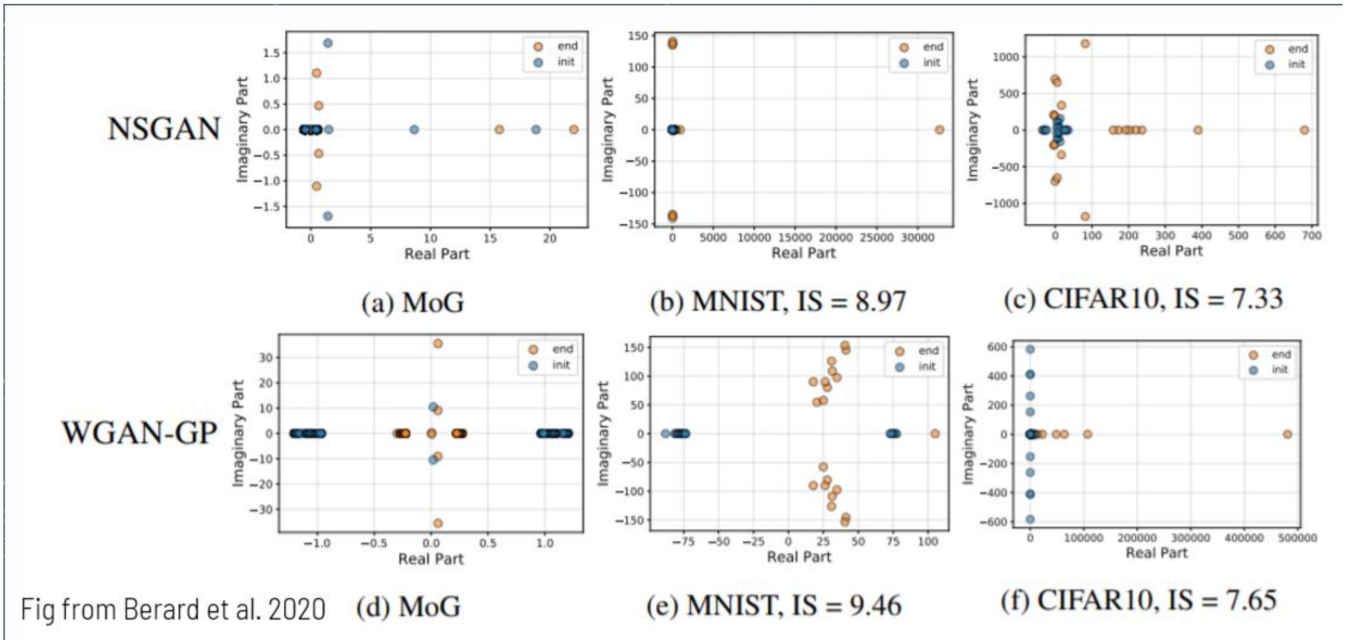
So basically EG is more stable that means locally differentiable stable stationary points is smaller than points stable for EG, i.e, if we take small η

$$\eta \leq min_\lambda \frac{1}{2\Re(\lambda)}$$

Exercise: Show that vector field of bilinear game is stable for EG but not for gradient method.
Soln. It is possible because of the imaginary part $\eta\Im(\lambda)^2$ EG is stable.
But for gradient method its not stable because diagonal element are not definite positive.

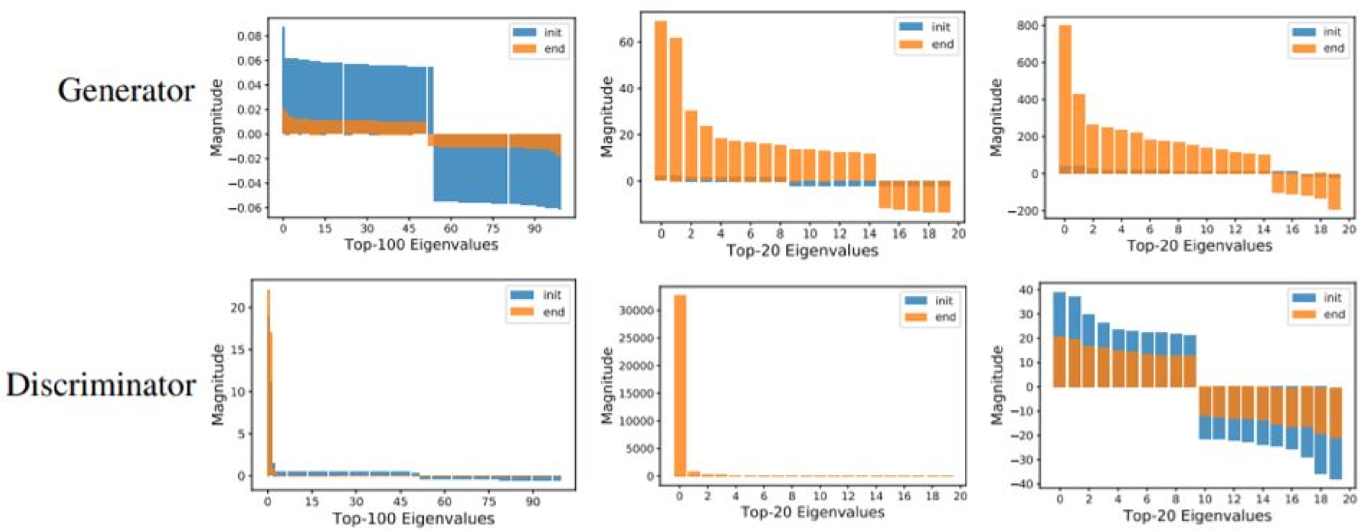
6 GAN in practice

In GAN, in practice we get the following plots for eigen values of jacobian. In practice we cannot plot all eigen values as it is not tractable, the jacobian is too large as it is the square of the number of parameter. But we can compute larger eigenvalues in terms of magnitude.



In plot we can see that real part of all eigen values is positive as it is on the right side of the plane, not at the initialization but at the optimum.

If we observe the below plots:



we are at saddle point for the generator. In case of discriminator, half the times we are at local minima and half the times we are at saddle point.

Question: Any intuition as to why the mixture of Gaussian with WGAN-GP has negative eigen values at end of the training?

Ans: If we look at the scale of the first figure of Discriminator plot of WGAN-GP, we can see that the imaginary part is very big. Since we are using extra gradient:

$$\Re(\lambda) - \eta(\Re(\lambda))^2 + \eta(\Im(\lambda))^2 > 0$$

We have real part -ve but imaginary part is large enough to compensate for -ve real part.

End of training: The real part of (largest) eigenvalues is positive. It seems that we find stable stationary points. (Atleast approximately)

The matrix S1 and S2 are not positive. It seems that we find stable stationary points that are not local nash equilibria.

Question: What can we say about the discriminator? It seems that for two of the three datasets, the eigenvalues for the discriminator stay positive.
Ans: We don't have to be in case where we have saddle points for everyone. The discriminator always has a flow in the way it discriminates. So generator has a way to fool the specific discriminator. That's why we have a saddle point for the generator. Discriminator does not have any way to fool the generator.

7 Conclusion

1. We can analyze stability using $\nabla F(w^*)$ and it can be seen as a dynamic equilibrium.
2. We can use block decomposition:

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

It is useful to determine differentiable nash equilibrium, i.e, to have diagonal as definite positive ($\nabla^2_\theta L_1(\theta^*, \phi^*)$ and $\nabla^2_\phi L_2(\theta^*, \phi^*)$)
3. Slightly weaker notion(sufficient second order condition) of stability/equilibrium.
4. The optimization method change the stability conditions. For instance, EG stabilizes the bilinear game, also games with negative real part for eigen values.

Question: In experimental results why do some have all imaginary points and some have all real eigen values?
Ans. It is maybe because of the scale that end points have imaginary part, but probably even the init points have imaginary part.
Prove: Locally differentiable stable stationary point

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

,
are not differentiable Nash equilibrium.

Soln: Bilinear game:

$$J = \nabla F(\omega) = \begin{pmatrix} 0 & B \\ -B^T & 0 \end{pmatrix}$$

[From previous slides]
Theorem:

$$J = -J^T \implies \Re(\lambda) = 0; \forall \lambda \in Sp(J)$$

Lets prove the above equation:

$$J_j = \lambda_j$$

$$j * J_j = \lambda ||j||^2_{\mathbb{C}}; j = x + iy \in \mathbb{C}^n$$

$$\Re(j * J_j) = \Re(\lambda)||j||^2_{\mathbb{C}}$$

$$x^T Jx + y^T Jy = \Re(\lambda)||j||^2_{\mathbb{C}}$$

–eqn(1)
Preposition: $J = -J^T \implies x^T Jx = 0$
Proof: $a = x^T Jx = (x^T Jx)^T = x^T J^T x = -x^T Jx = -a \implies a = 0$
Therefore eqn 1 becomes:

$$\Re(\lambda)||j||^2_{\mathbb{C}} \implies \Re(\lambda) = 0$$

Reminder:
Bilinear game:

$$min_\theta max_\phi \theta^T B\phi$$

,
where: $\theta \in \mathbb{R}^n, B \in \mathbb{R}^{n*m}, \phi \in \mathbb{R}^m$
Conclusion:

$$\nabla F(\omega) = \begin{pmatrix} 0 & B \\ -B^T & 0 \end{pmatrix} = -\nabla F(\omega)$$

, i.e, skew symmetric.

$$\implies \Re(\lambda) = 0; \forall \lambda \in Sp(\nabla F(\omega))$$

This implies, not stable for GD but stable for EG if: $\Im(\lambda) > 0; \forall \lambda \in Sp(\nabla F(\omega))$
Therefore, equivalent to say B is invertible.

Theorem: there exists

$$J = \begin{pmatrix} S_1 & -B^T \\ B & S_2 \end{pmatrix}$$

such that $S_1 > 0, S_2 \ngtr 0$,i.e, Not a local nash.
But $\Re(\lambda) > 0 \forall \lambda \in Sp(J)$ i.e, Stable for GD
Proof:
We will consider the following matrix:

$$J = \begin{pmatrix} 2 & 0 & a \\ 0 & 1 & 0 \\ -a & 0 & -1 \end{pmatrix}$$

where $S_1 = 2$ and $S_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$
We have to find value of 'a'.
$Tr(J) = \lambda_1 + \lambda_2 + \lambda_3 = 1$
We use wolfram and find that the value of a = 2.
Question: What is the relation of the non zero sum game matrix:

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

Ans: When A and B equals to zero, spectrum of F is spectrum of S1 and S2.
Question: Since in practice we do not reach the Nash Equilibrium but the models still work well, Is it worth trying to reach the Nash equilibrium anyways?
Ans: We do not know. Methods to reach Nash Equilibrium:
(a) Follow paper by Adolphs et al. (2018); Mazumdar et al. (2019)
(b) Use second order information.

References

[1] H. Berard, G. Gidel, A. Almahairi, P. Vincent, and S. Lacoste-Julien. A closer look at the optimization landscapes of generative adversarial networks, 2020.