IFT 6085 - Lecture 14
The Numerics of GANs

Scribe(s): Joshua Romoff **Instructor:** Ioannis Mitliagkas

1 Summary

In the previous lecture, we discussed the Wasserstein Generative Adversarial Network (WGAN), which uses the Wasserstein distance instead of the traditional Jenson-Shannon divergence. WGANs help alleviate the zero-gradient problem that arises when we have a mismatch in support between the generative and true distribution.

In this lecture, we will focus on the training dynamics for GANs. Specifically, we will frame the training objective for GANs as a zero-sum game, and focus on one particular method, taken from [2], that helps GANs convergence to Nash Equilibria.

2 The Numerics of GANs

We assume that we are in the smooth two-player game setting where player 1 is trying to maximize $f(\phi, \theta)$, player 2 is trying to maximize $g(\phi, \theta)$, and f, g are both smooth with respect to (ϕ, θ). We also consider the GAN setting which can be cast as a Zero-sum game.

Definition 1 (Zero-sum games)**.**
$$g(\phi, \theta) = -f(\phi, \theta)$$

In zero-sum games, any gain by one player results in a loss to the other player and vice versa.

Convergence in this setting is considered to be a Nash Equilibria.

Definition 2 (Nash Equilibria (N.E))**.** *Point $\bar{x} = (\bar{\phi}, \bar{\theta})$ is a N.E if :*

$$\bar{\phi} \in \arg\max_{\phi} f(\phi, \bar{\theta})$$

and,

$$\bar{\theta} \in \arg\max_{\theta} g(\bar{\phi}, \theta)$$

In a N.E, neither Player 1 or Player 2 can improve given his opponents current parameters. We further define a **Local N.E** to be a point $\bar{x}$ where the above holds for some local neighbourhood.

Two player games can be framed in terms of Vector Fields.

Definition 3 (Vector Field)**.** *is a mapping V such that:*

$$V : \mathbb{R}^d \to \mathbb{R}^d$$

Definition 4 (Vector Field for 2-player games)**.** *A vector field defined over parameters (ϕ, θ):*

$$V(\phi, \theta) = \begin{bmatrix} \nabla_\phi f(\phi, \theta) \\ \nabla_\theta g(\phi, \theta) \end{bmatrix}$$

Definition 5 (Jacobian)**.** *The generalization of the derivative for vector valued functions. For $f : \mathbb{R}^n \to \mathbb{R}^m$, the Jacobian J is defined as:*

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Definition 6 (Jacobian of vector field for two player games)**.** *The generalization of the hessian from standard optimization:*

$$V' = \begin{bmatrix} \nabla^2_\phi f(\phi, \theta) & \nabla_{\phi,\theta} f(\phi, \theta) \\ -\nabla_{\phi,\theta} f(\phi, \theta)^T & \nabla^2_\theta f(\phi, \theta) \end{bmatrix}$$

Lemma 7. *The following holds for all* 0*-sum games:*

$V'(x)$ *is negative (semi)-definite iff* $\nabla^2_\phi f(\phi, \theta)$ *is negative (semi)-definite and* $\nabla^2_\theta f(\phi, \theta)$ *is positive (semi)-definite*

Corollary 8. *The following holds for all* 0*-sum games:*

$V'(x)$ *is negative semi-definite* $\forall$ *local N.E* $\bar{x}$

Proposition 9 (Classic Result)**.** *Let F be a continuously differentiable function from $\Omega \to \Omega$. Take $\bar{x} \in \omega$ such that:*

$$F(\bar{x}) = \bar{x} \quad \texttt{(Fixed Point)}$$

and,

$$|\lambda| < 1 \quad \forall \;\; \texttt{e-val} \;\; \texttt{(}\lambda\texttt{)} \;\; \texttt{of} \;\; F'(\bar{x})$$

then $\exists$ a set of neighbours U of $\bar{x}$ such that $\forall x \in U$:

$$F^{(k)}(x) \to \bar{x}$$

and,

$$||F^{(k)}(x) - \bar{x}|| = O(|\lambda_{max}|^k)$$

where $|\lambda_{max}|$ is the largest (absolute) eigenvalue of $F'(\bar{x})$.

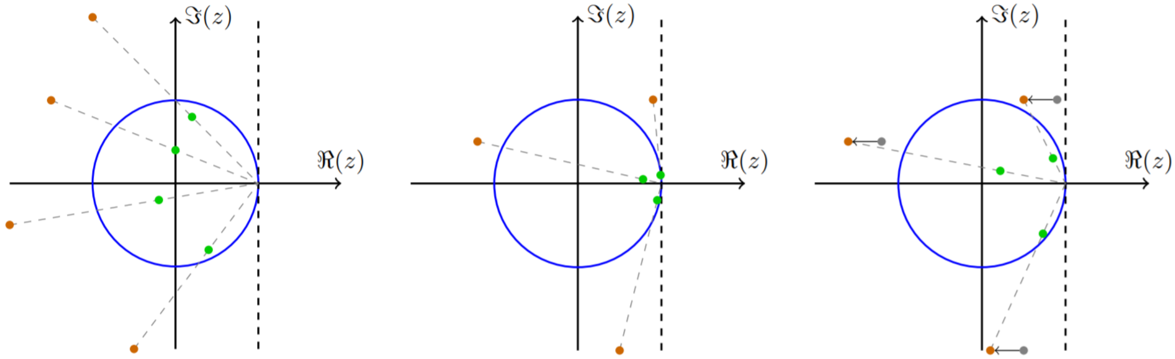
For the proof see proposition 4.4.1 of [1]. By the proposition above we understand that the convergence rate is dictated by the largest (absolute) eigenvalue.

3 Main Result

The paper uses a slight abuse of notation and denotes V', the Jacobian of the vector for two player games, as G' or A. Similarly, $G = V$.
In numerics we are typically interested in functions with the following form:

$$F(x) = x + hG(x)$$

where h is the learning rate.



(a) Illustration how the eigenvalues are projected into unit ball.

(b) Example where h has to be chosen extremely small.

(c) Illustration how our method alleviates the problem.

The Jacobian of the above function can be written as:

$$F'(x) = I + hG'(x)$$

The main issue is that G' is not symmetric and therefore the eigenvalues of F and G are not necessarily real, they can have an imaginary component. The authors analyze the above equation and uncover a sufficient and necessary condition for the conditions of proposition 9 to hold.

Lemma 10 (Main Result)**.** *Assume that $A \in R^{dxd}$, if $\forall$ e-val (λ) of A, the real part $Re(\lambda) < 0$ then:*

$$|\lambda| < 1 \quad \forall \ \texttt{e-val} \ (\lambda) \ \texttt{of} \ I + hA$$

iff,

$$h < \frac{1}{|Re(\lambda)|} \frac{2}{1 + (\frac{Im(\lambda)}{Re(\lambda)})^2}$$

So, in order for the eigenvalues to be within the unit-ball, the learning rate must be made sufficiently small. Figure 3 illustrates this effect. If either the real or imaginary component are very large then a very small learning rate is needed to project the eigenvalues back into the unit ball. The authors propose to add a gradient penalty that effectively pushes the eigenvalues to the left (Figure 3 (c)) thus allowing for larger learning rates while still staying within the unit ball.

References

[1] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods (Optimization and Neural Computation Series)*. Athena Scientific, 1 edition, 1996. ISBN 1886529043.

[2] L. M. Mescheder, S. Nowozin, and A. Geiger. The numerics of gans. *CoRR*, abs/1705.10461, 2017. URL `http://arxiv.org/abs/1705.10461`.