Carnegie Mellon University Computer Science Department

Gradient Descent GAN Optimization is Locally Stable

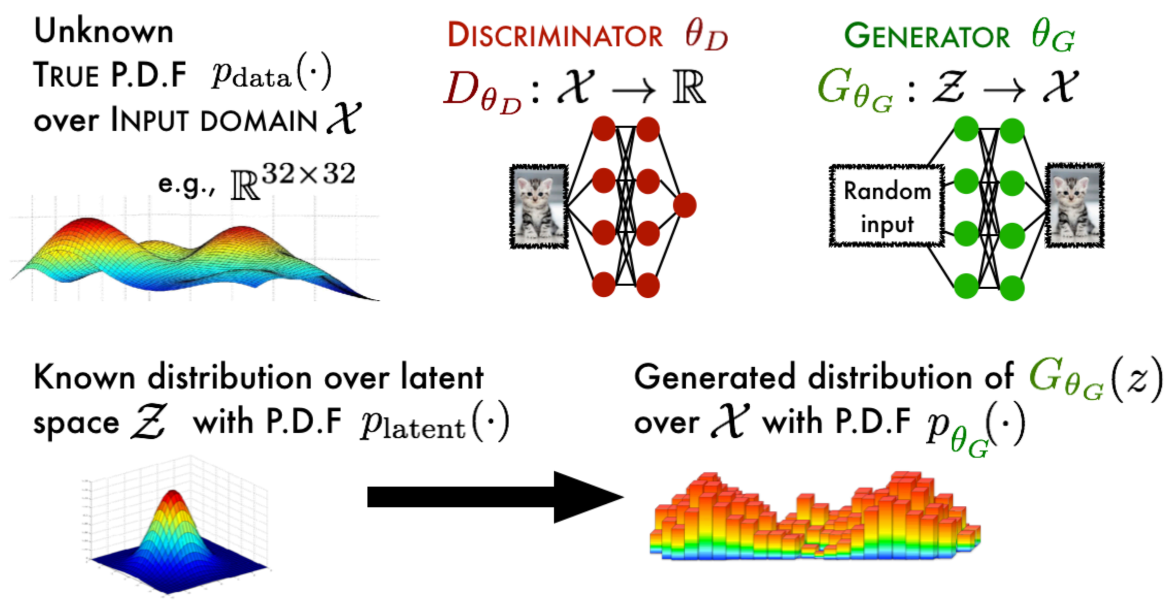
Vaishnavh Nagarajan, J. Zico Kolter

Attend our Oral Presentation!
Wed Dec 6th 11:05 – 11:20 AM
Hall C

We address a fundamental open question in GANs about the stability of the saddle point equilibrium, using tools from non-linear systems theory.

Generative Adversarial Networks (GANs)

An increasingly popular class of generative models – models that "understand" data to output new random data.
Formally, GANs learn distribution over the data.



$$\min_{\theta_G} \max_{\theta_D} V(\theta_G, \theta_D)$$

$$\mathbb{E}_{x\sim p_{\rm data}}\underbrace{[f(D(x))]}_{\text{How real x is, according to the discriminator}} + \mathbb{E}_{z\sim p_{\rm latent}}\underbrace{[f(-D(G(z)))]}_{\text{How "generated" } G(z) \text{ looks according to the discriminator}}$$

Traditional GAN	Wasserstein GAN (WGAN)
$f(t) = \log\left(\frac{1}{1+\exp(-t)}\right)$	$f(t) = t$

To use non-linear systems tools, we update both parameters **simultaneously** (not alternatively) and **infinitesimally** for the true (not empirical) objective.

Repeat simultaneously:

time derivative

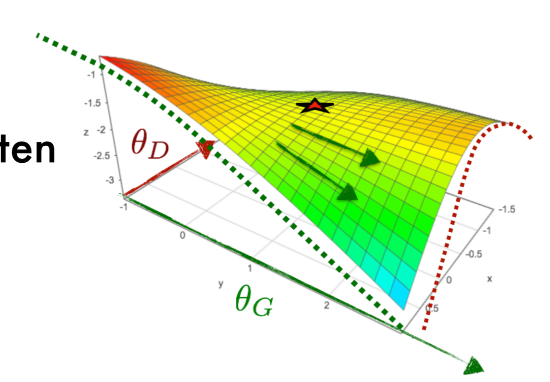
$$\dot{\theta}_D = \nabla_{\theta_D} V(\theta_G, \theta_D)$$
$$\dot{\theta}_G = -\nabla_{\theta_G} V(\theta_G, \theta_D)$$

until equilibrium:

$$\dot{\theta}_D = 0$$
$$\dot{\theta}_G = 0$$

Proving GAN stability is hard because of concave-concave-ity **arbitrarily close to equilibrium**, even for linear models!

Other proofs [3,4] require discriminator to be trained more often – effectively, closer to optimality, bringing system closer to pure minimization.



Toolbox: Non-Linear Systems

Consider a dynamical system $\dot{\theta} = h(\theta)$ for which $\theta^\star$ is an equilibrium point i.e., $h(\theta^\star) = 0$

Locally exponentially stable: if for *any* initialization sufficiently close to equilibrium, the system converges to equilibrium quickly. [1]

$$J = \left.\frac{\partial h(\theta)}{\partial \theta}\right|_{\theta^\star}$$

Linearization Theorem: The equilibrium is locally exponentially stable iff Jacobian at equilibrium has eigenvalues with **strictly negative real parts**.

Our extension: Eigenvalue can have zero real part as long as it corresponds to direction within a subspace of equilibria

Alternative stability analysis: LaSalle's invariance principle
Informally, equilibrium is exponentially stable if there exists a non-negative energy function ("Lyapunov function") that at any time either
i) strictly decreases
ii) remains constant only instantaneously,
and is zero only at equilibrium.

Assumptions

Some of our assumptions are strong, yet they shed light on what is required for stability.

1. Global equilibrium is realizable.

$$p_{\theta_G^\star}(\cdot) = p_{\rm data}(\cdot)$$
$$D_{\theta_D^\star}(x) = 0 \text{ for all } x$$

2. $\boxed{f''(0) < 0 \,\&\, f'(0) \neq 0}$

3.

$$h_D(\theta_D) = \mathbb{E}_{\rm data}[D^2_{\theta_D}(x)]$$
$$h_G(\theta_G) = \left.\left\| \mathbb{E}_{p_{\rm data}}[\nabla_{\theta_D} D_{\theta_D}(x)] - \mathbb{E}_{p_{\theta_G}}[\nabla_{\theta_D} D_{\theta_D}(x)] \right\|^2\right|_{\theta_D = \theta_D^\star}$$

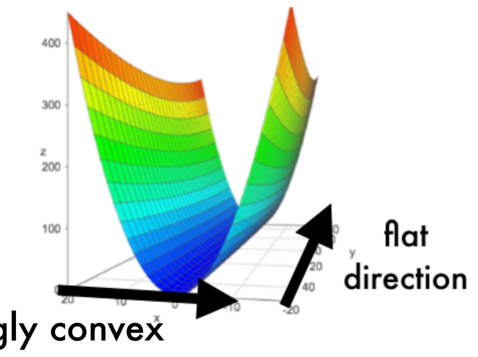
A notion of how "different" the parameters θ_D, θ_G are from equilibrium discriminator $\theta_D^\star$ and generator $\theta_G^\star$ respectively.

We assume at equilibrium, these locally convex functions are locally strongly convex – implying a locally unique equilibrium.

$$\boxed{\left.\nabla^2_{\theta_D} h_D(\theta_D)\right|_{\theta_D^\star} \succ 0 \qquad \left.\nabla^2_{\theta_G} h_G(\theta_G)\right|_{\theta_G^\star} \succ 0}$$

$$\nabla^2_{\theta_D} V(\theta_D, \theta_G^\star) \prec 0 \qquad \left.\nabla^2_{\theta_G} \|\nabla_{\theta_D} V(\theta_D, \theta_G)\|^2\right|_{\theta_D = \theta_D^\star} \succ 0$$

Relaxation: Zero eigenvalues must correspond to "flat" directions within a subspace of equilibria.

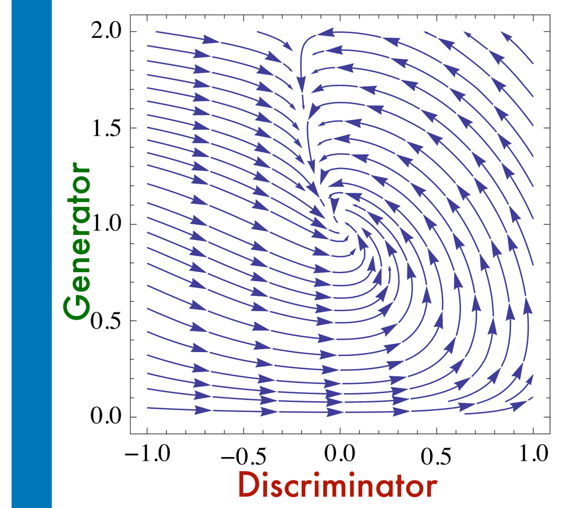


Results

Theorem: Under assumptions 1, 2, 3 the GAN system is locally exponentially stable.

The Jacobian at equilibrium is:

$$\begin{bmatrix} \nabla^2_{\theta_D} V(\theta_D, \theta_G) & \nabla_{\theta_G}\nabla_{\theta_D} V(\theta_D, \theta_G) \\ -\left(\nabla_{\theta_G}\nabla_{\theta_D} V(\theta_D, \theta_G)\right)^T & -\nabla^2_{\theta_G} V(\theta_D, \theta_G) \end{bmatrix} = \begin{bmatrix} \text{negative definite} & \text{full column rank} \\ \text{negative transpose} & 0 \end{bmatrix}$$



GAN dynamics for a small system learning a uniform distribution.

Key Lemma: Matrices of this form have eigenvalues with strictly negative real parts despite zero block!

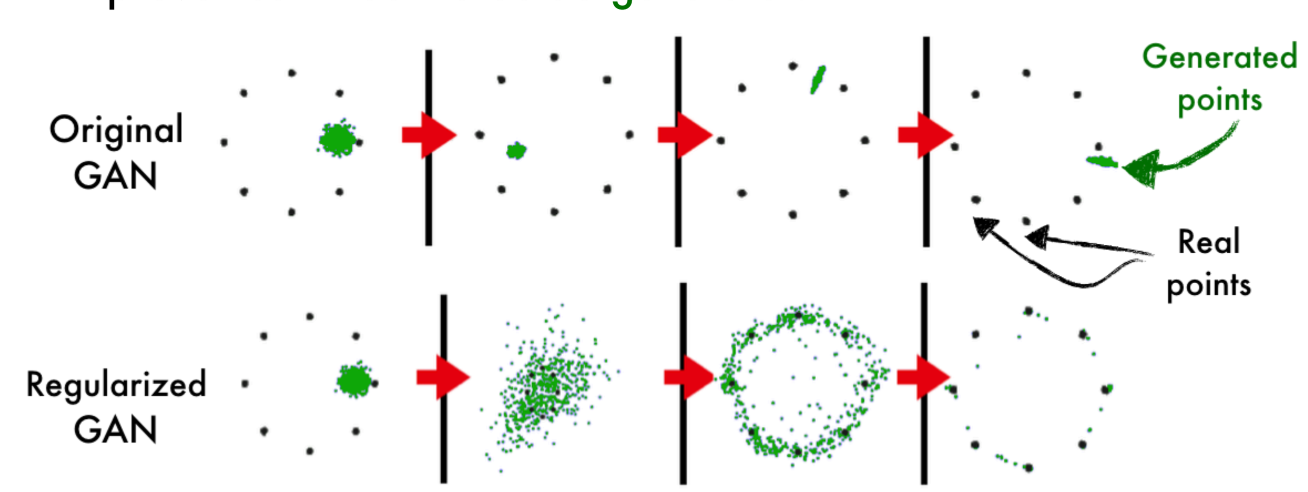
Decreasing, and only instantaneously non-decreasing energy function: Distance from equilibrium

Gradient Norm Based Regularization

$$\dot{\theta}_D = \nabla_{\theta_D} V(\theta_D, \theta_G)$$
$$\dot{\theta}_G = -\nabla_{\theta_G} V(\theta_D, \theta_G) - \eta \nabla_{\theta_G} \|\nabla_{\theta_D} V(\theta_D, \theta_G)\|^2$$

Generator minimizes (the objective + the norm of the discriminator's gradient).

1. Makes Jacobian "more stable"
2. Provides foresight to the generator: When norm of discriminator's gradient is minimized, discriminator cannot improve itself much to outdo generator.



3. Similar to 1-unrolled [2] approximation of the pure minimization problem.

Simultaneous Gradient Descent WGAN

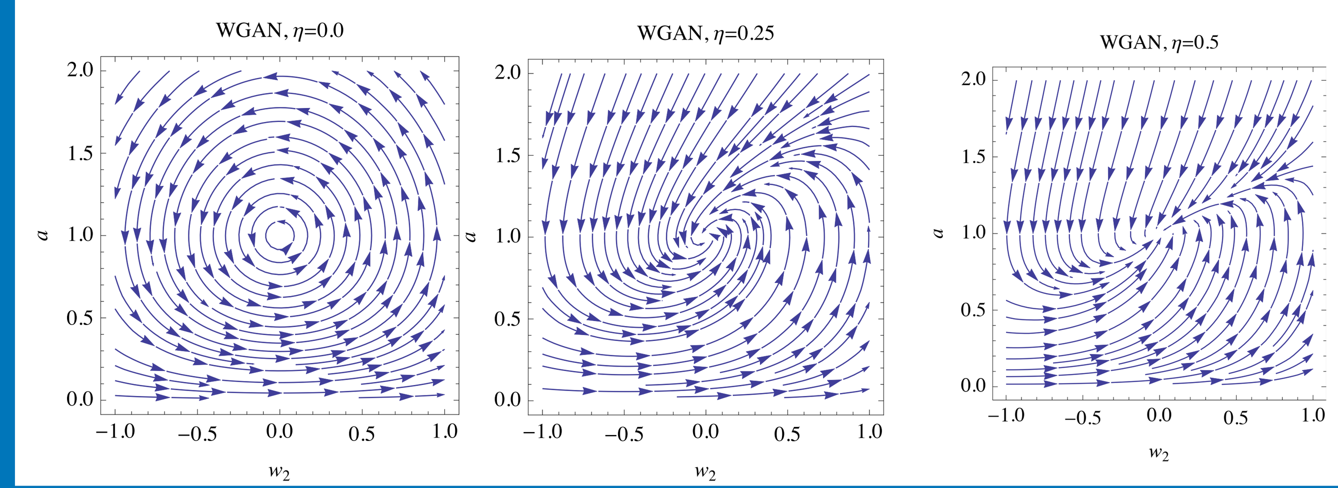
$$\begin{bmatrix} 0 & \text{full column rank} \\ \text{negative transpose} & 0 \end{bmatrix}$$

There exist simultaneous gradient descent WGAN systems which do not necessarily converge to the equilibrium.

Theorem: Under similar assumptions, the equilibrium of the regularized simultaneous gradient descent (W)GAN system is exponentially stable when η not too large.

$$\begin{bmatrix} 0 & \text{full column rank} \\ \text{negative transpose} & \text{negative definite} \end{bmatrix}$$

(even though the discriminator is not trained to optimality!)



Conclusion

Theoretical analysis of GAN dynamics using non-linear systems: despite lack of concave-convexity, GANs are locally stable under right conditions – perhaps why they work so well in practice! Regularizer can provably enhance local stability.

Open Questions

- Analyze other objectives and optimization techniques: unrolled GANs, f-GANs ...
- Relaxing some assumptions: Non-realizability: when discriminator is not linear in parameters? Without strong curvature? Intuitively: slower convergence, but still stable...
- Global convergence, at least for linear discriminator and generator!
- When do equilibria satisfying our conditions exist?

References

1. H. K Khalil. Non-linear Systems. Prentice-Hall, New Jersey, 1996.
2. L. Metz, et al., Unrolled generative adversarial networks. (ICLR 2017)
3. M. Heusel et al., GANs trained by a TTUR converge to a local Nash equilibrium (NIPS 2017)
4. I. J. Goodfellow et al., Generative Adversarial Networks (NIPS 2014)