GRADIENT DESCENT GAN OPTIMIZATION IS LOCALLY STABLE.

Vaishnavh Nagarajan, Zico Kolter

Computer Science Department,
Carnegie Mellon University

GENERATIVE ADVERSARIAL NETWORKS (GANs)

A breakthrough generative model
[Goodfellow et al., '14]

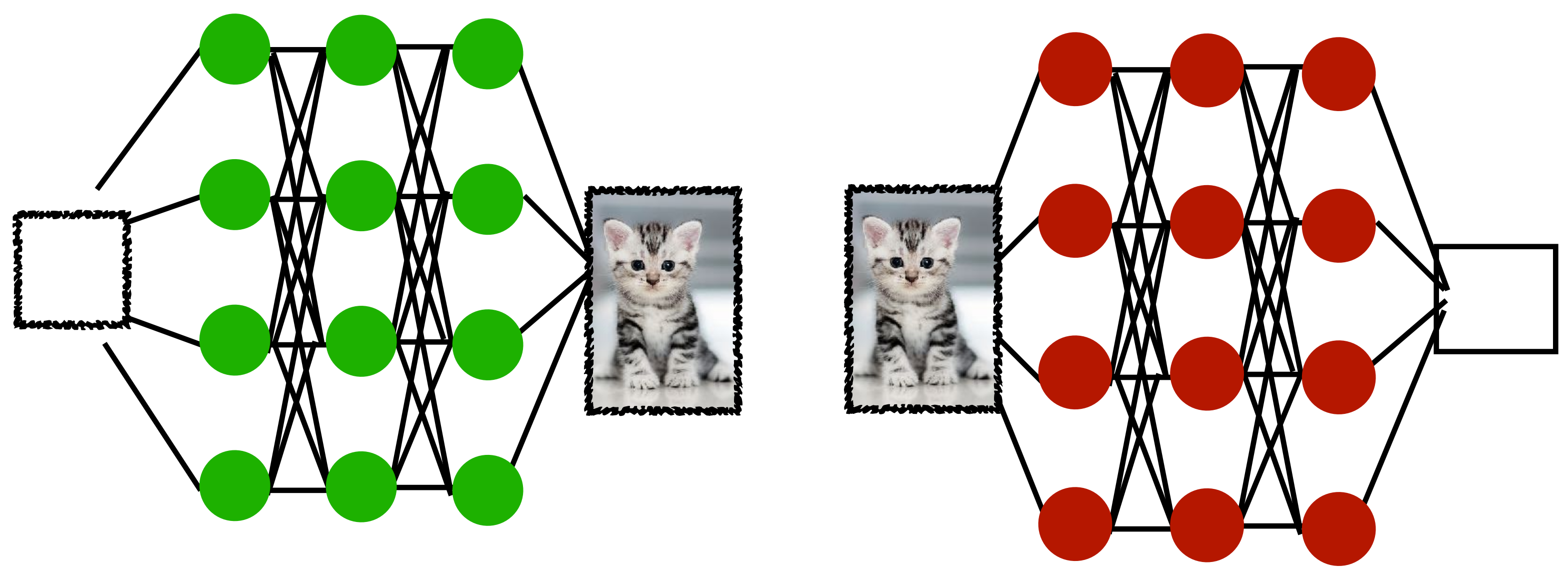
[Images of "fake" celebrities from Karras et al., '17]

We study
a **fundamental** question about
convergence of GAN optimization
using tools from non-linear
systems theory.

GENERATIVE ADVERSARIAL NETWORKS (GANS)

GENERATOR

θ_G



DISCRIMINATOR

θ_D

$$\min_{\theta_G} \max_{\theta_D} V(\theta_G, \theta_D) = \mathbb{E}_{x \sim p_{real}}[\log(D(x))] + \mathbb{E}_{z \sim p_{latent}}[\log(1 - D(G(z)))]$$

how well discriminator tells apart generated from real

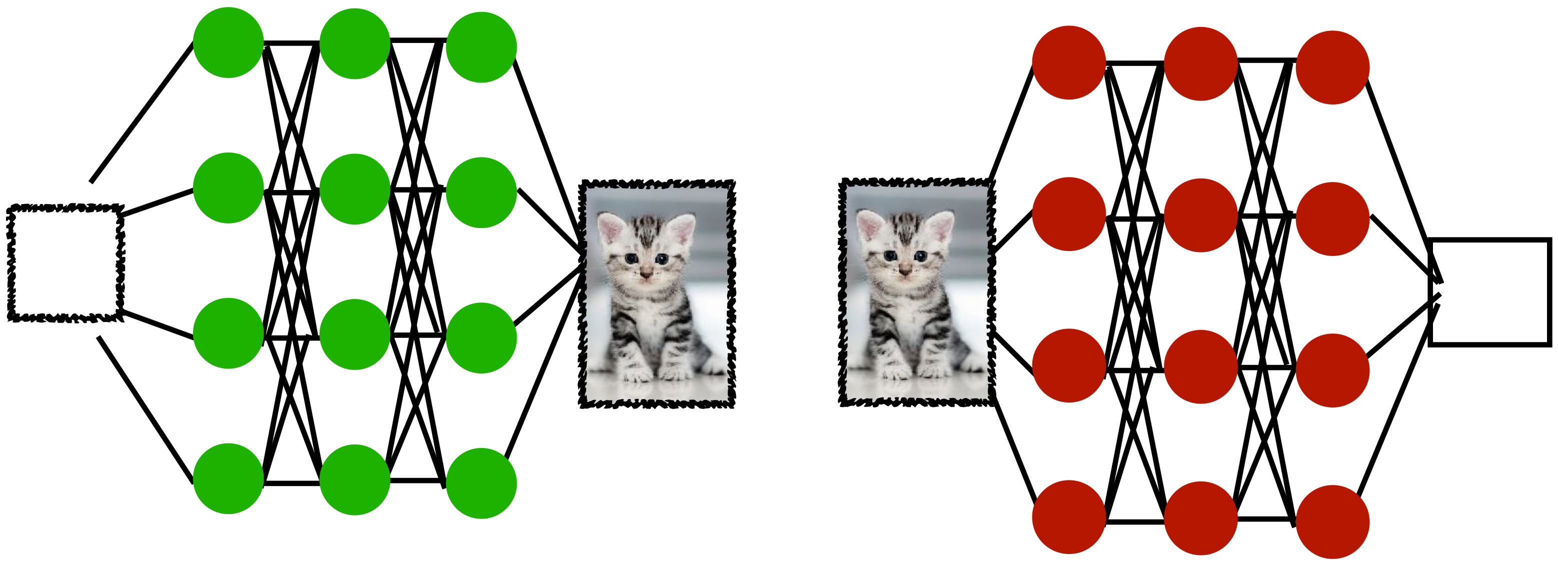
GAN OPTIMIZATION

Find (global) equilibrium of the game

i.e., saddle point of min-max objective.

Generative Adversarial Networks (GANs)

Generator
θ_G



Discriminator
θ_D

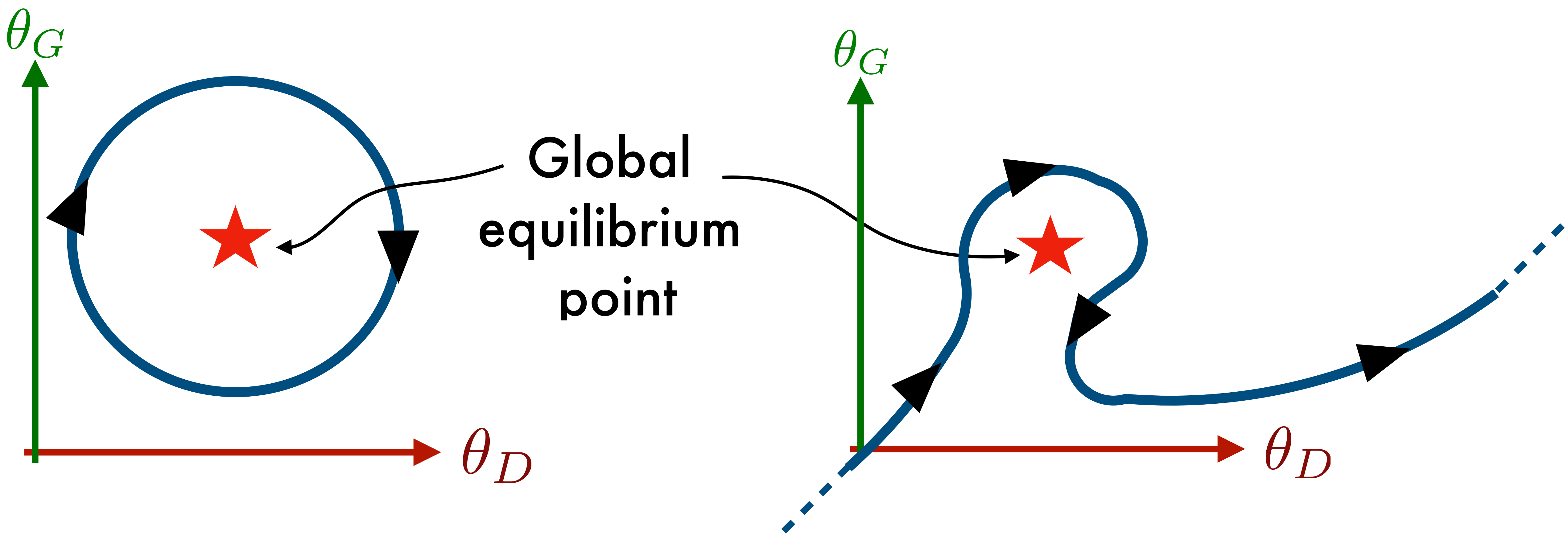
$$\min_{\theta_G} \max_{\theta_D} V(\theta_G, \theta_D) = \mathbb{E}_{x \sim p_{real}}[\log(D(x))] + \mathbb{E}_{z \sim p_{latent}}[\log(1 - D(G(z)))]$$

how well discriminator tells apart generated from real

Global Equilibrium: Generated distribution = Real distribution.

If this is realizable,
does it have "good convergence properties"?

GAN optimization typically **seems** to find a good solution. But has it really converged?



Open Question

Can we rule out cycling/unstable dynamics near equilibrium?

Is the equilibrium “locally exponentially stable”?

Informally, is *any* initialization sufficiently close to equilibrium guaranteed to converge under the optimization procedure?

“Minimum” requirement from the optimization procedure!

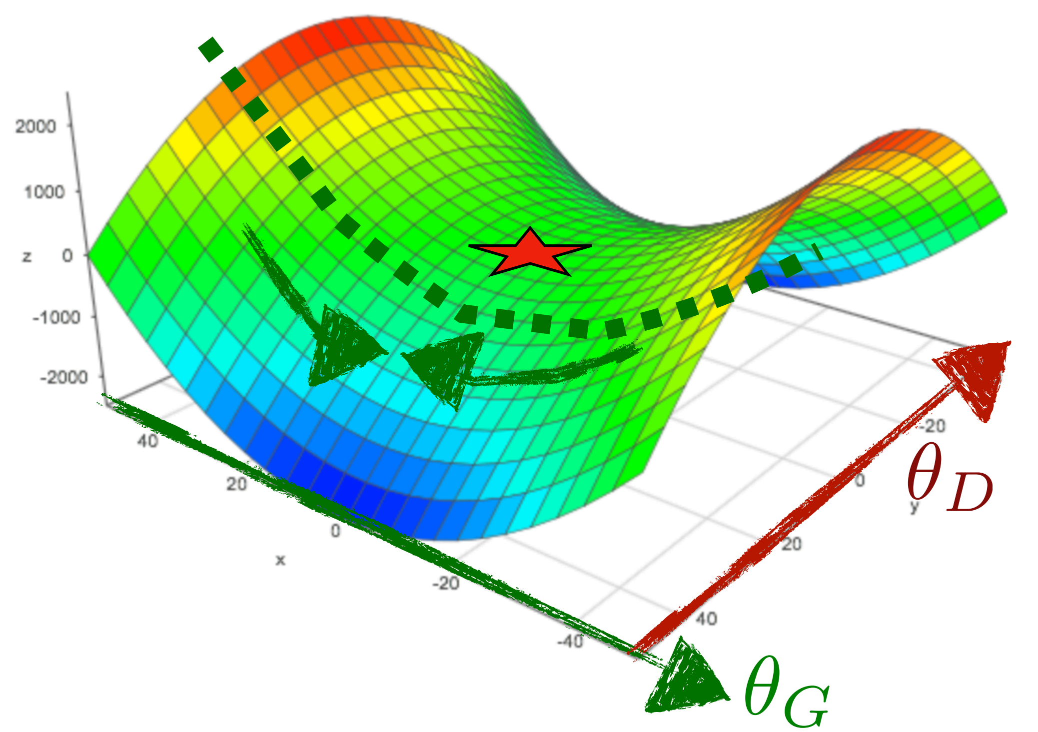
WHY IS PROVING GAN STABILITY HARD?

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{x \sim p_{real}}[\log(D(x))] + \mathbb{E}_{z \sim p_{latent}}[\log(1 - D(G(z)))]$$

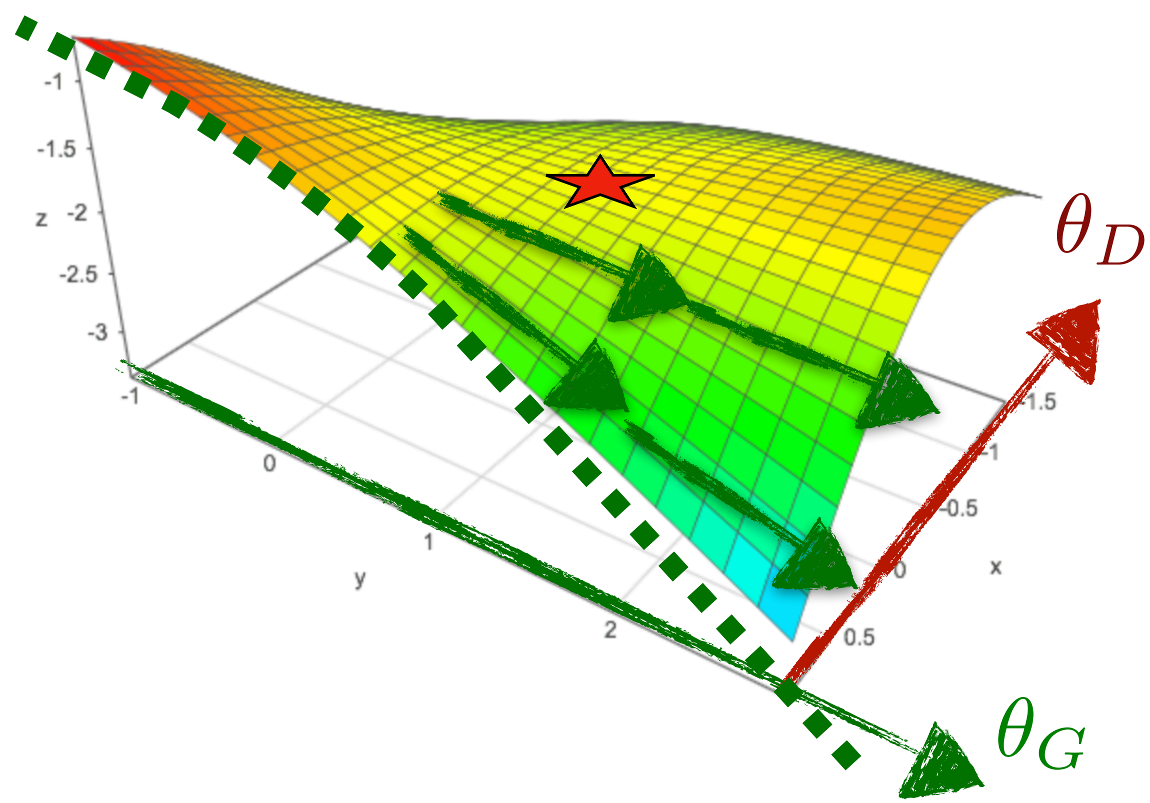
Concave at equilibrium :)

Concave even arbitrarily close to equilibrium
even for linear generator & discriminator **:(**

Not this (convex-concave)



But this (concave-concave)



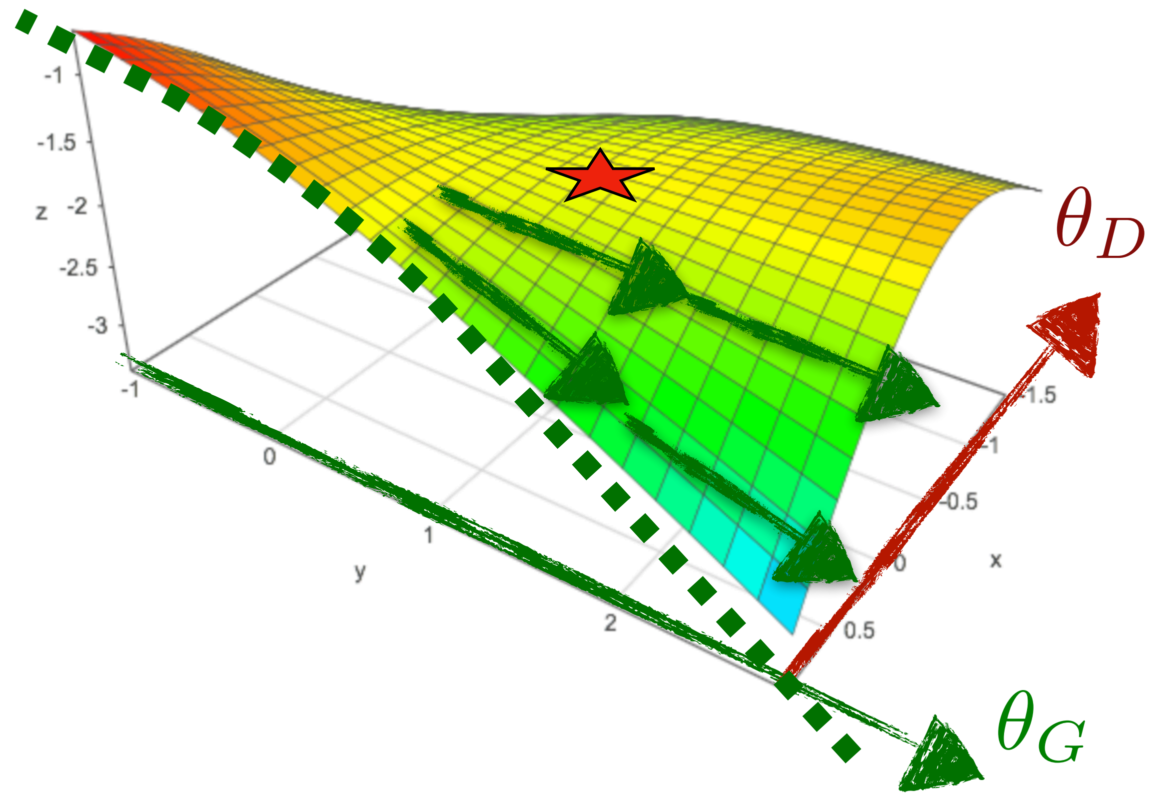
Why is proving GAN stability hard?

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{x \sim p_{real}}[\log(D(x))] + \mathbb{E}_{z \sim p_{latent}}[\log(1 - D(G(z)))]$$

Concave at equilibrium :)

Concave even arbitrarily close to equilibrium
even for linear generator & discriminator **:(**

But this (concave-concave)



Even arbitrarily close to equilibrium, updating only the generator – will diverge because of concavity!

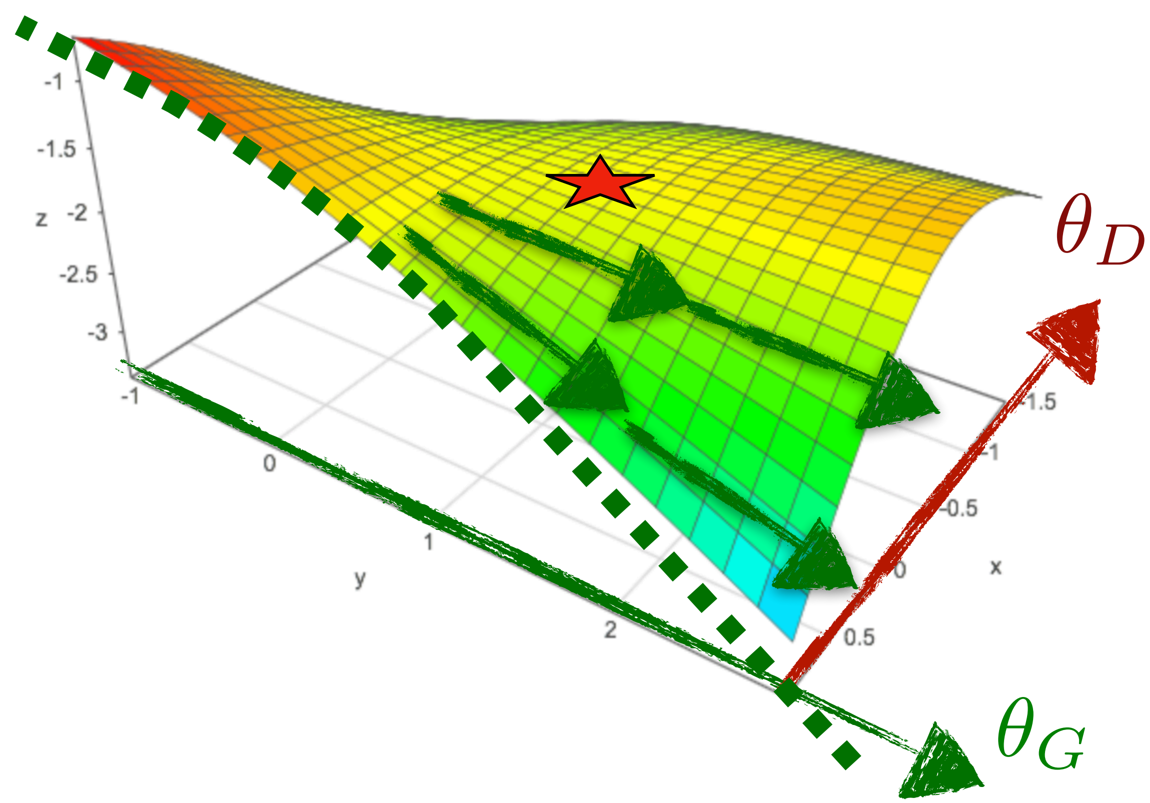
Why is proving GAN stability hard?

Other proofs [Li et al., '17, Heusel et al., '17]: stability given discriminator is trained more often – closer to a pure minimization

In practice, seems to work without this assumption!

But this (concave-concave)

Even arbitrarily close to equilibrium, updating only the generator – will diverge because of concavity!



GAN Optimization

Infinitesimal, **simultaneous gradient descent:**
closer to practically used GAN training i.e., updates at similar frequency

time derivative

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

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

Computationally cheaper than alternate updates – fewer forward & backward passes

Despite a concave-concave objective,
despite not training discriminator to optimality at each step,
simultaneous gradient descent GAN equilibrium
is
"locally exponentially stable"
under suitable conditions.

Toolbox: Non-Linear Systems

Linearization Theorem: The equilibrium $\theta^\star$ of a non-linear system is locally exponentially stable if and only if its Jacobian at equilibrium

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

has eigenvalues with **strictly negative real parts.**

Proof Outline

Jacobian near equilibrium

$$\begin{bmatrix} \partial \dot{\theta}_D / \partial \theta_D & \partial \dot{\theta}_D / \partial \theta_G \\ \partial \dot{\theta}_G / \partial \theta_D & \partial \dot{\theta}_G / \partial \theta_G \end{bmatrix}$$

Proof Outline

Jacobian near equilibrium

$$\begin{bmatrix} \partial \dot{\theta}_D / \partial \theta_D & \partial \dot{\theta}_D / \partial \theta_G \\ \partial \dot{\theta}_G / \partial \theta_D & \text{positive semi-definite :(} \end{bmatrix}$$

because of concave-concavity!

Proof Outline

Jacobian at equilibrium

$$\begin{bmatrix} \partial\dot\theta_D / \partial\theta_D & \partial\dot\theta_D / \partial\theta_G \\ \partial\dot\theta_G / \partial\theta_D & 0 \quad :(\end{bmatrix}$$

Key question: Can this have all eigenvalues with strictly negative real parts despite the zero block?

Proof Outline

Jacobian at equilibrium

well-behaved

$$\begin{bmatrix} \partial \dot{\theta}_D / \partial \theta_D & \partial \dot{\theta}_D / \partial \theta_G \\ \partial \dot{\theta}_G / \partial \theta_D & 0 \end{bmatrix}$$

Key Lemma: Under some strong curvature assumptions, all eigenvalues have negative real parts despite the zero diagonal block!

Proof Outline

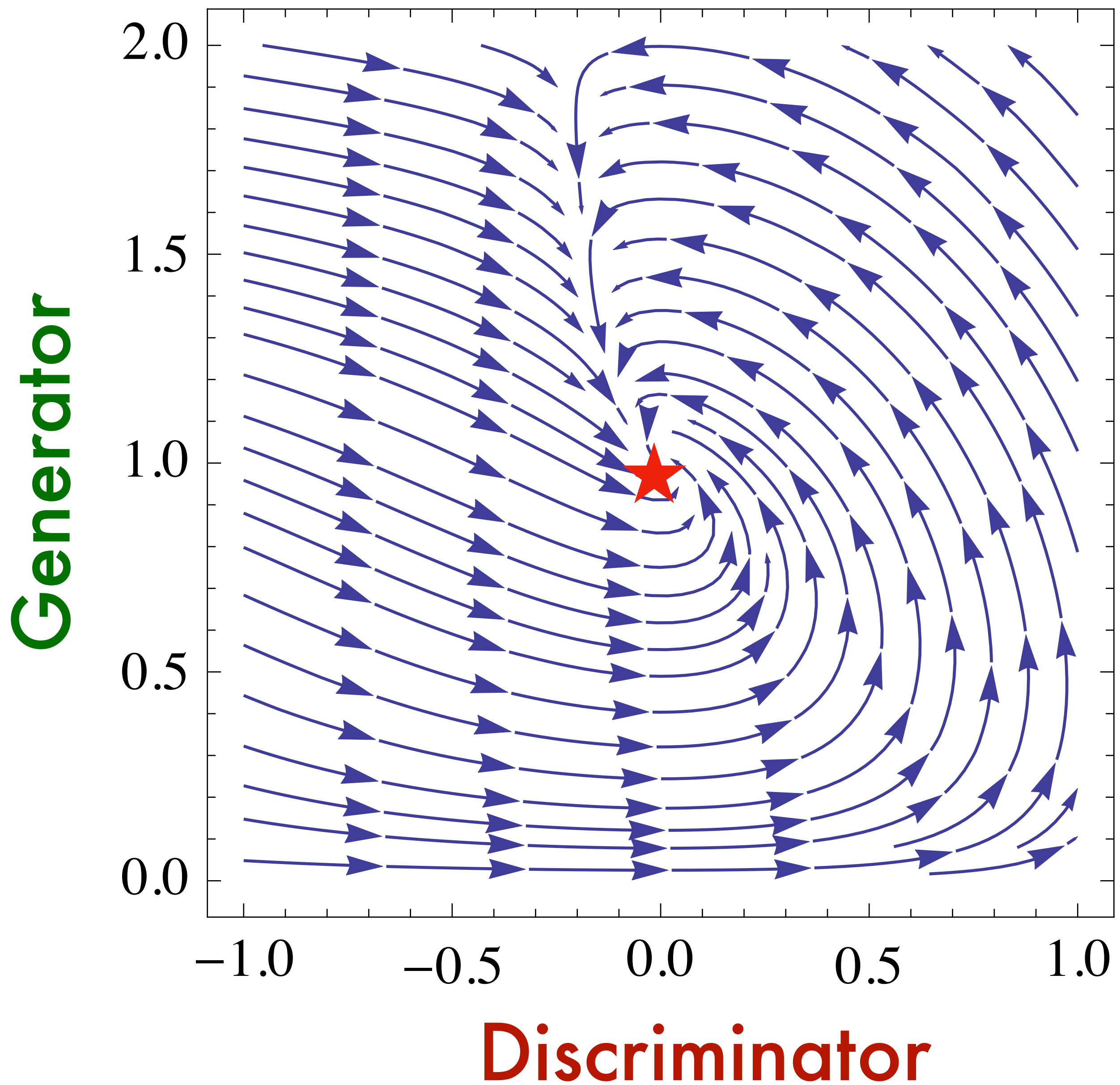
Jacobian at equilibrium

well-behaved

$$\begin{bmatrix} \partial\dot{\theta}_D / \partial\theta_D & \partial\dot{\theta}_D / \partial\theta_G \\ \partial\dot{\theta}_G / \partial\theta_D & 0 \end{bmatrix}$$

Thus, equilibrium is locally exponentially stable
despite the zero diagonal block!

ILLUSTRATION



A quadratic discriminator - linear generator system learning a uniform distribution.

The dynamics of simultaneous gradient descent GAN is quite non-linear. But it still converges!

Gradient-norm based Regularization

θ_D **maximizes** $V(\theta_D, \theta_G)$

θ_G **minimizes** $V(\theta_D, \theta_G) + \eta \left\| \nabla_{\theta_D} V(\theta_D, \theta_G) \right\|^2$

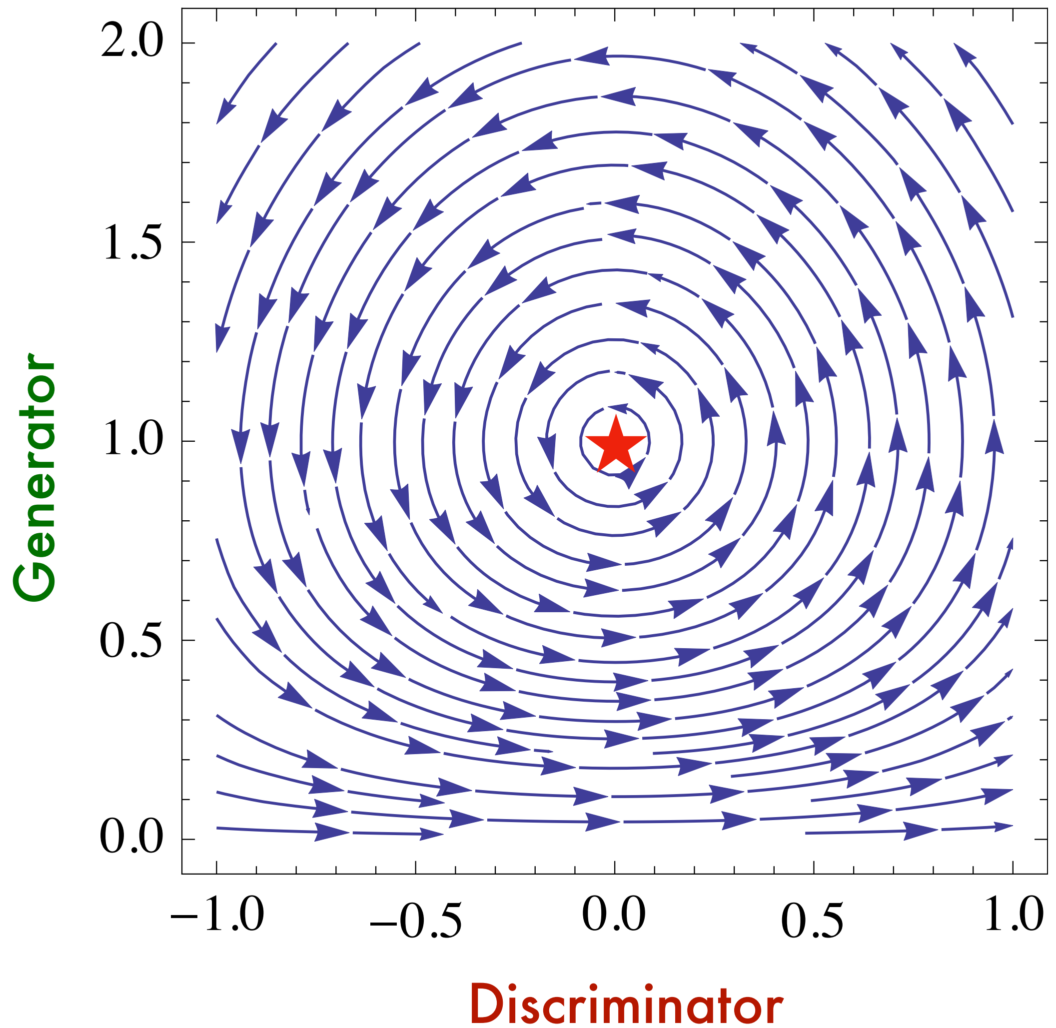
like a damping term

Generator minimizes the objective +
the norm of gradient w.r.t discriminator parameters.

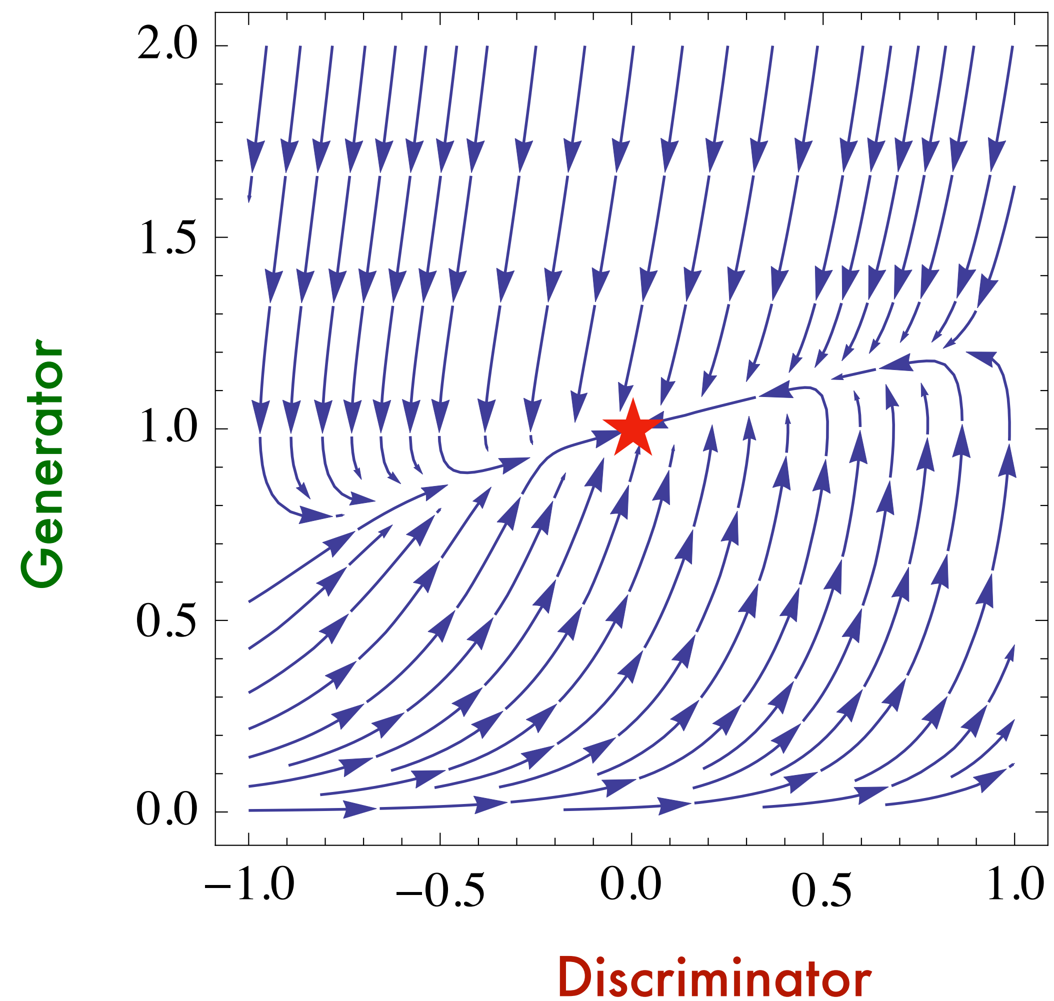
GRADIENT-NORM BASED REGULARIZATION

Provably enhances local stability.

Wasserstein GAN [Arjovsky'17]
under
simultaneous gradient descent



... with regularization



CONCLUSION

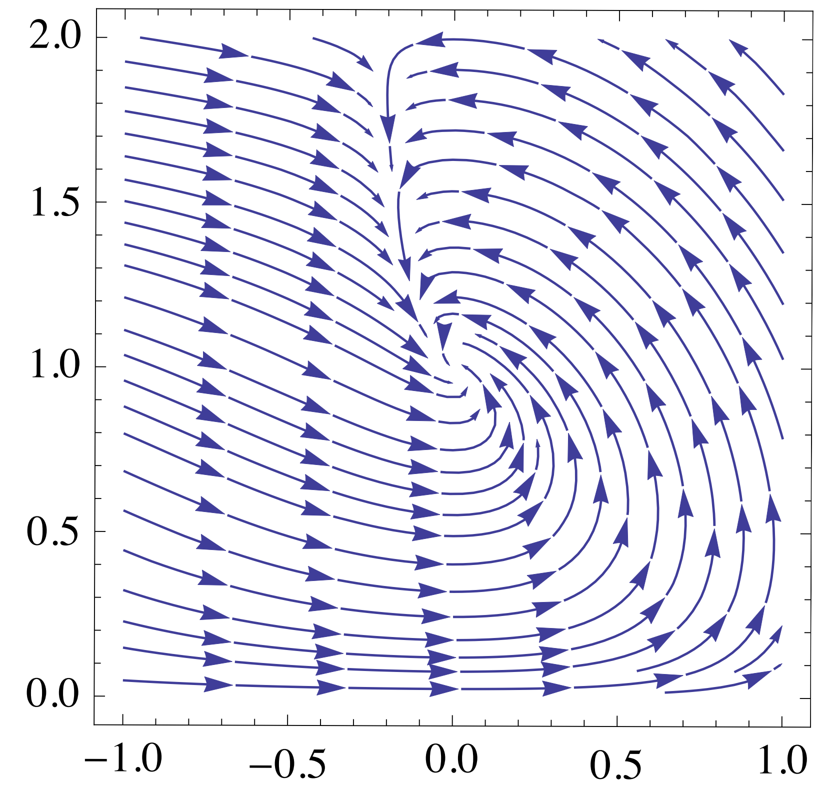
- Local stability of GANs using non-linear systems
- GAN objective is concave-concave, yet simultaneous gradient descent equilibrium is locally stable — perhaps why GANs have worked well in practice.
- Regularization term provably enhances local stability.

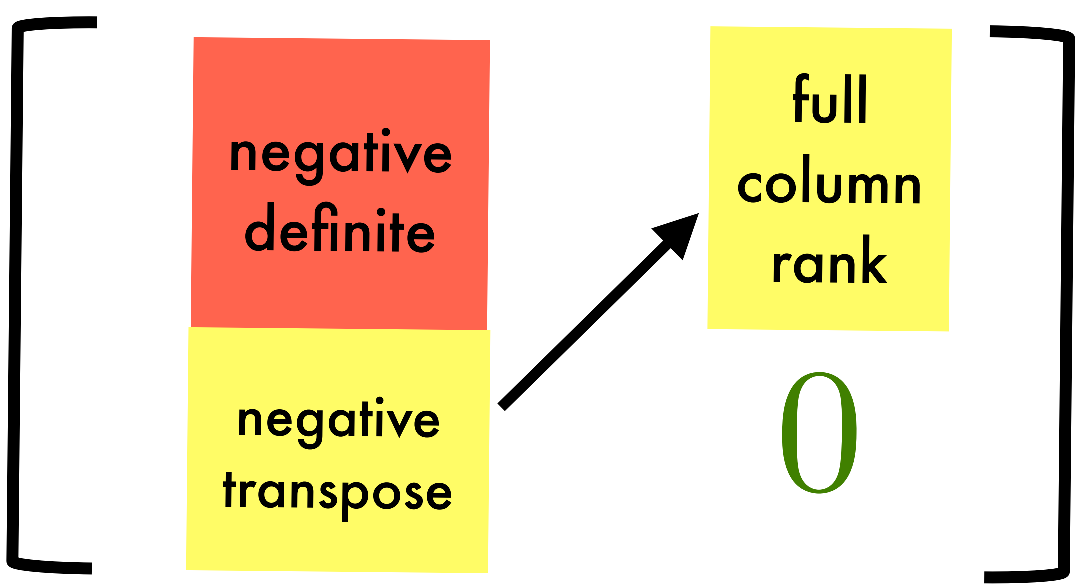
Open Questions

- Analyze other objectives and optimization techniques: f-GANs, unrolled GANs ...
- Relaxing some assumptions e.g., non-realizable case.
- Global convergence, at least for simple architectures

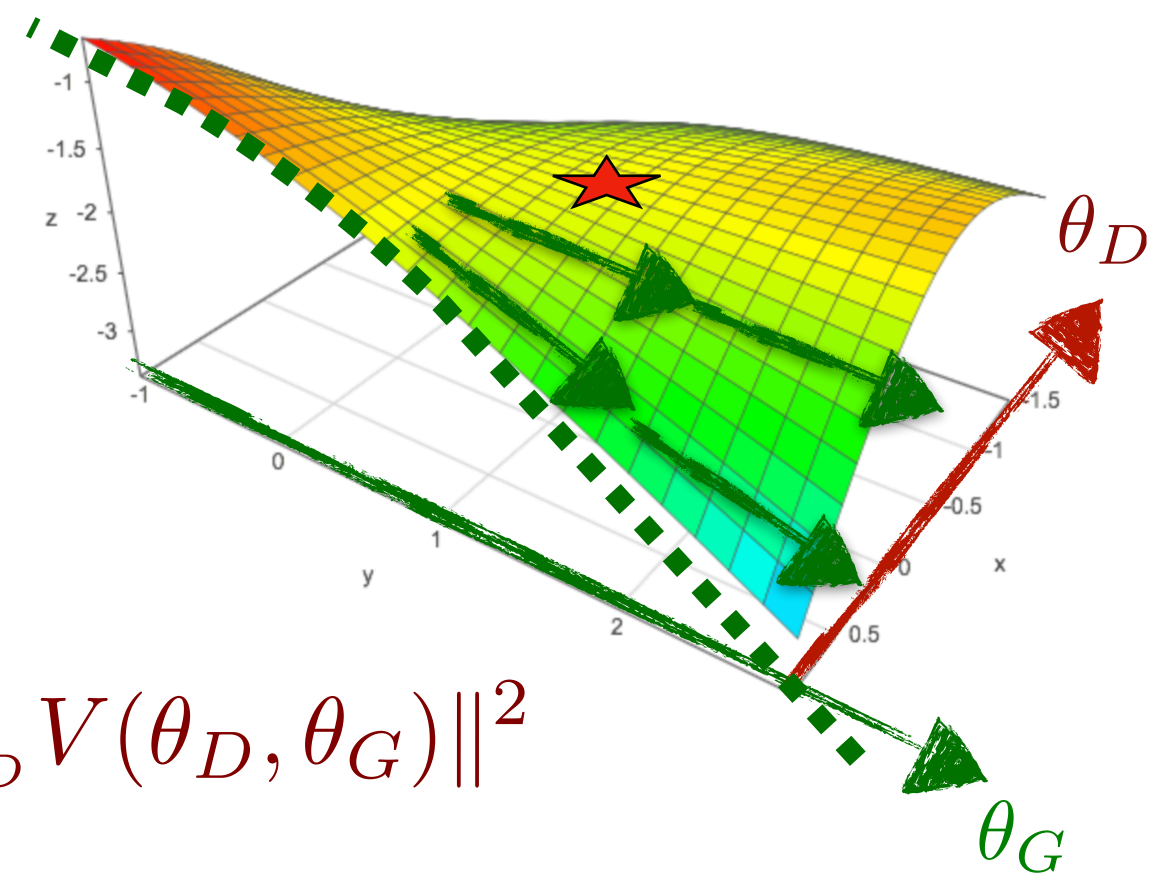
- Many other theoretical questions e.g., when do equilibria satisfying our conditions exist?
- Many other powerful tools in non-linear systems theory!

THANK YOU.
QUESTIONS?
POSTER #99





concave-concave:



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