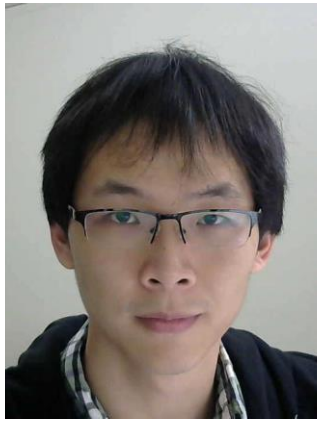
What is local optimality in nonconvex-nonconcave minimax optimization?

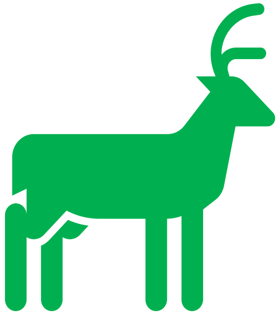
Praneeth Netrapalli
Microsoft Research India



Chi Jin
UC Berkeley

Michael I. Jordan
UC Berkeley

Minimax optimization/ Two player zero sum game

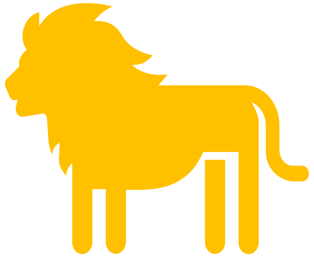


x

$$\min_{x} f(x, y)$$

y

$$\max_{y} f(x, y)$$



- Several applications in economics, evolutionary biology etc.
- Simultaneous vs sequential (Stackelberg) games
- Widely studied in the convex-concave setting
- All versions equivalent (Sion's minimax theorem [Sion 1958])

Minimax theorem [Sion 1958]

If $f(x, y)$ is convex in x and concave in y, then

$$\min_x \max_y f(x, y) = \max_y \min_x f(x, y)$$

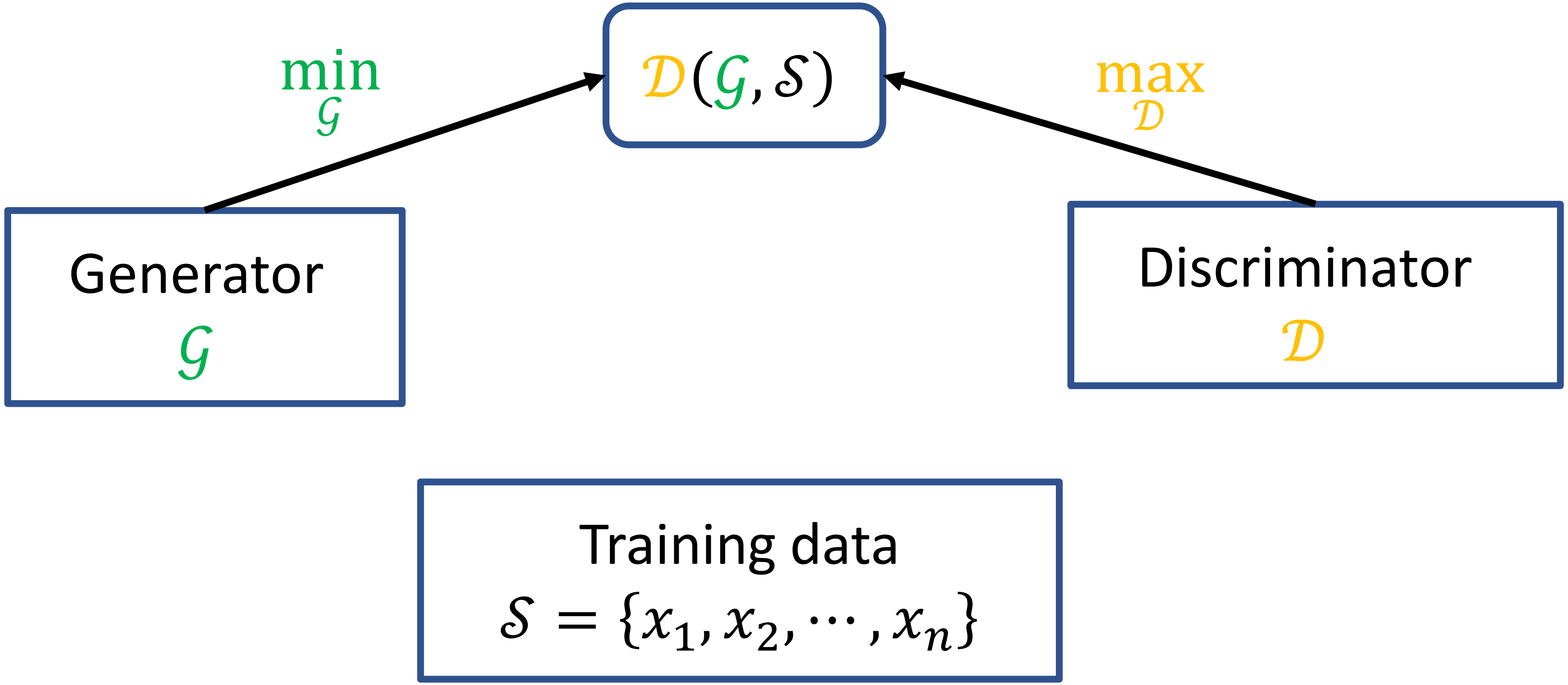
- Does not matter who plays first
- Optimal strategy (x^*, y^*) is called Nash equilibrium

$$x^* \in \mathrm{argmin}_x\, f(x, y^*) \quad \text{and} \quad y^* \in \mathrm{argmax}_y\, f(x^*, y)$$

- Extensive work on computing Nash equilibria in convex-concave setup
- Most machine learning applications are nonconvex-nonconcave

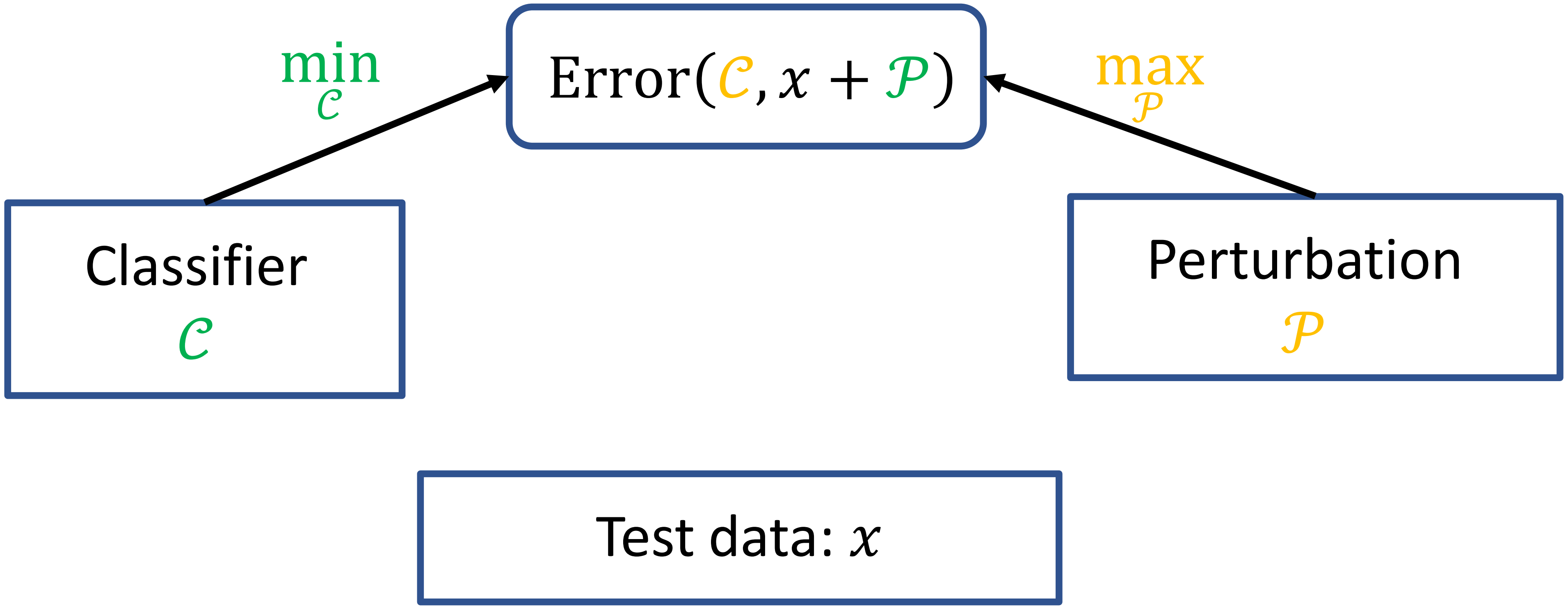
Machine learning applications

- Generative adversarial networks (for learning a distribution from samples) [Goodfellow et al. 2014]



Machine learning applications

- Robust machine learning (for learning models that are robust to attacks) [Madry et al. 2017]



Machine learning applications

- Mostly nonconvex-nonconcave
- Theory and understanding for convex-concave no longer apply

What can we say (and do) in this general setting?

- Inspired by convex vs nonconvex optimization?
  - Local notions of optimality?
  - Algorithms?

Outline

- Existing notions of local optimality (and their drawbacks)
- New notion of local optimality – **local minimax**
- Gradient descent ascent – relation to **local minimax**
- Future directions

Existing notions of local optimality

- Local Nash equilibrium [Daskalakis and Panageas 2018; Mazumder and Ratliff 2018; Adolphs et al. 2018]
  - Replaces global min and global max with local versions

$$x^* \in \mathrm{LocalMin}_x f(x, y^*) \quad \text{and} \quad y^* \in \mathrm{LocalMax}_y f(x^*, y)$$

- First and second order conditions
  - Replaces optimality conditions with first order stationarity

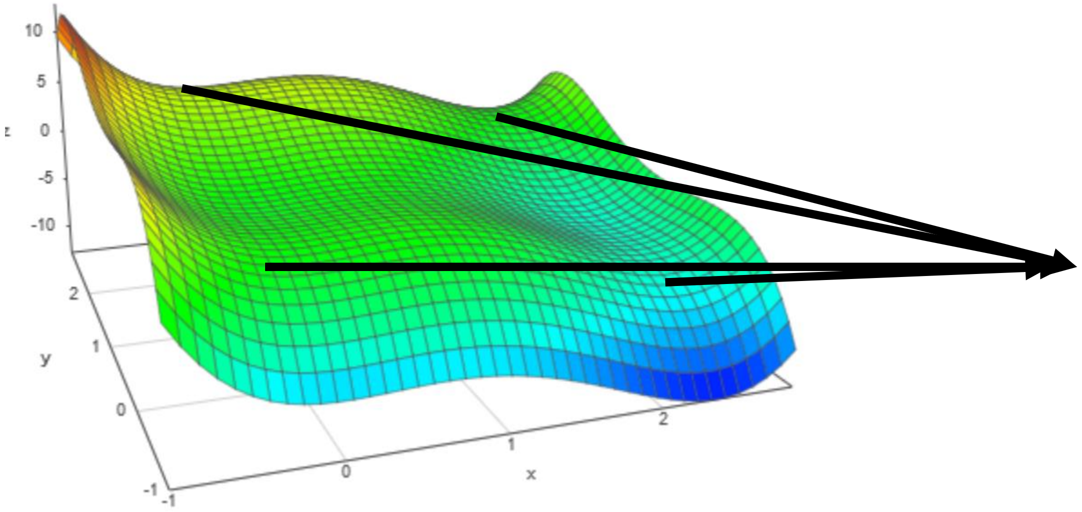
$$\nabla_x f(x^*, y^*) = 0 \quad \text{and} \quad \nabla_y f(x^*, y^*) = 0$$

  - Or second order stationarity

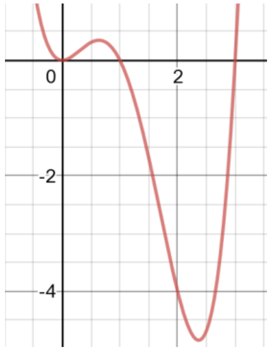
$$\nabla_{xx}^2 f(x^*, y^*) \succcurlyeq 0 \quad \text{and} \quad \nabla_{yy}^2 f(x^*, y^*) \preccurlyeq 0$$

Local Nash equilibrium

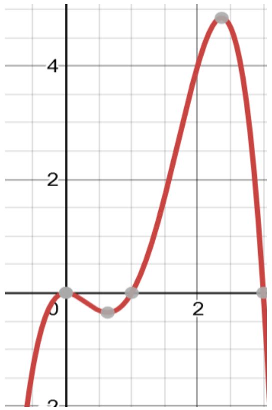
$x^* \in \mathrm{LocalMin}_x f(x, y^*)$ and $y^* \in \mathrm{LocalMax}_y f(x^*, y)$



Local Nash



$f(\cdot, y^*)$

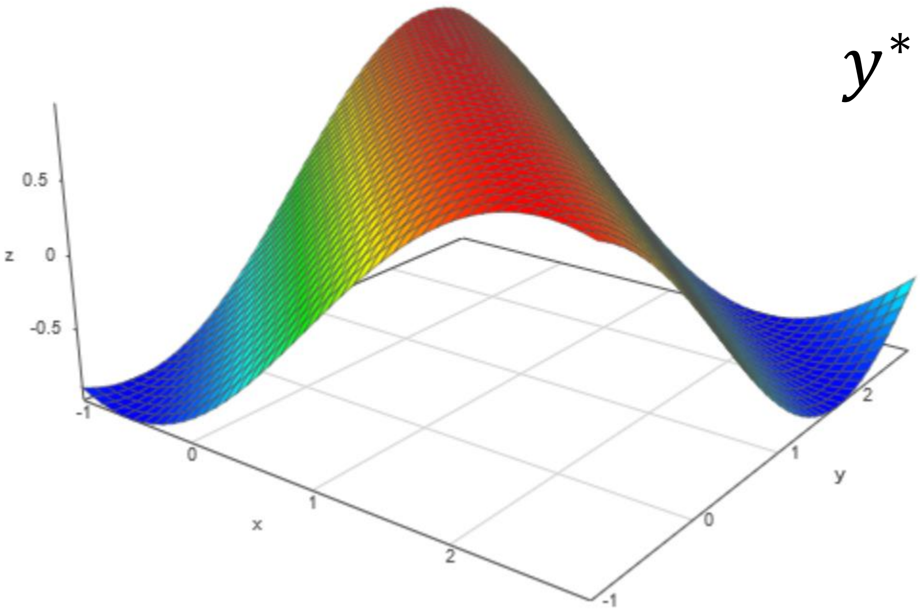


$f(x^*, \cdot)$

Local Nash equilibrium

- Unfortunately (both global and local Nash) do not always exist

$\sin(x + y)$



$$x^* \in \text{LocalMin}_x f(x, y^*) \Rightarrow x^* + y^* = \left(2k\pi + \frac{\pi}{2}\right)$$

$$y^* \in \text{LocalMax}_y f(x^*, y) \Rightarrow x^* + y^* = \left(2k\pi - \frac{\pi}{2}\right)$$

Local Nash – First and second order conditions

$$\nabla_x f(x^*, y^*) = 0 \quad \text{and} \quad \nabla_y f(x^*, y^*) = 0$$

$$\nabla^2_{xx} f(x^*, y^*) \succcurlyeq 0 \quad \text{and} \quad \nabla^2_{yy} f(x^*, y^*) \preccurlyeq 0$$

- Again, does not always exist e.g., $\sin(x + y)$

Main observation

- The local notions considered so far are inspired by simultaneous games
- In convex-concave setting, simultaneous/sequential does not matter
- In nonconvex-nonconcave setting, it is important
- **Takeaway**: Consider local notions of **sequential (aka Stackelberg) equilibria** (which is guaranteed to exist unlike Nash equilibrium)

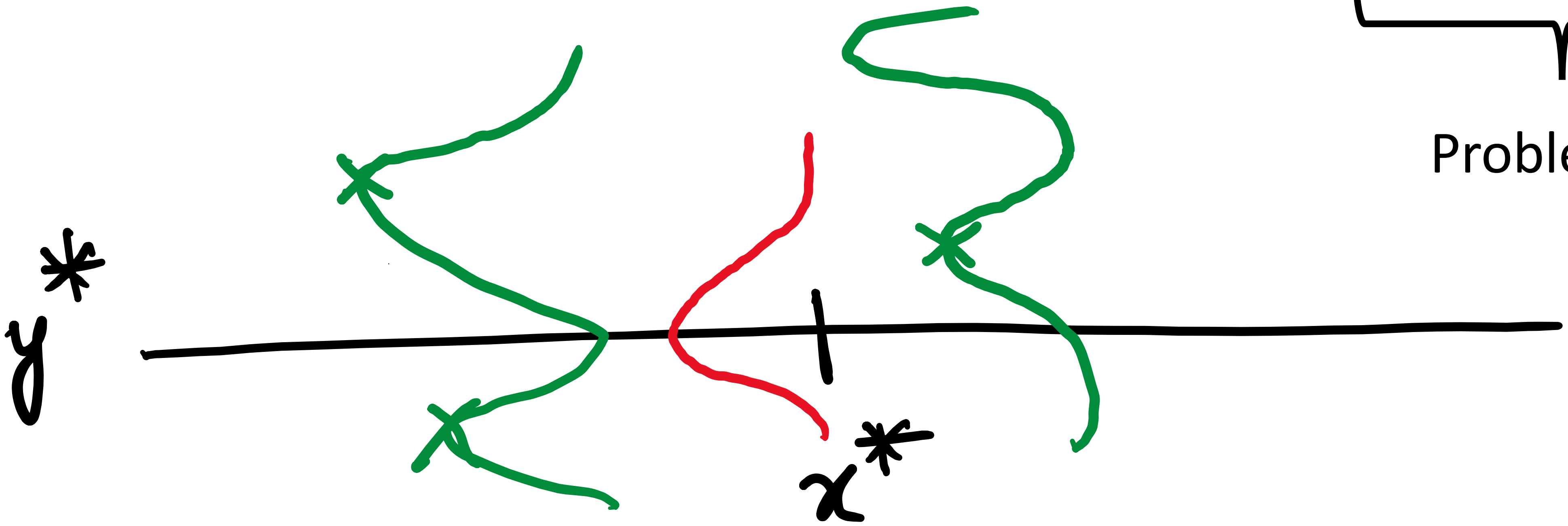
Global sequential/Stackelberg equilibrium

$$y^* \in \operatorname{argmax}_y f(x^*, y) \quad \text{and} \quad x^* \in \operatorname{argmin}_x \left(\max_y f(x, y) \right)$$

- In essence, fix the order $\min_x \max_y f(x, y)$ (or viceversa) and solve $\min_x g(x)$, where $g(x) \stackrel{\text{def}}{=} \max_y f(x, y)$
- Also known as global minimax
- Always exists under mild conditions
- Finding it is of course hard in general

Towards local Stackelberg equilibrium

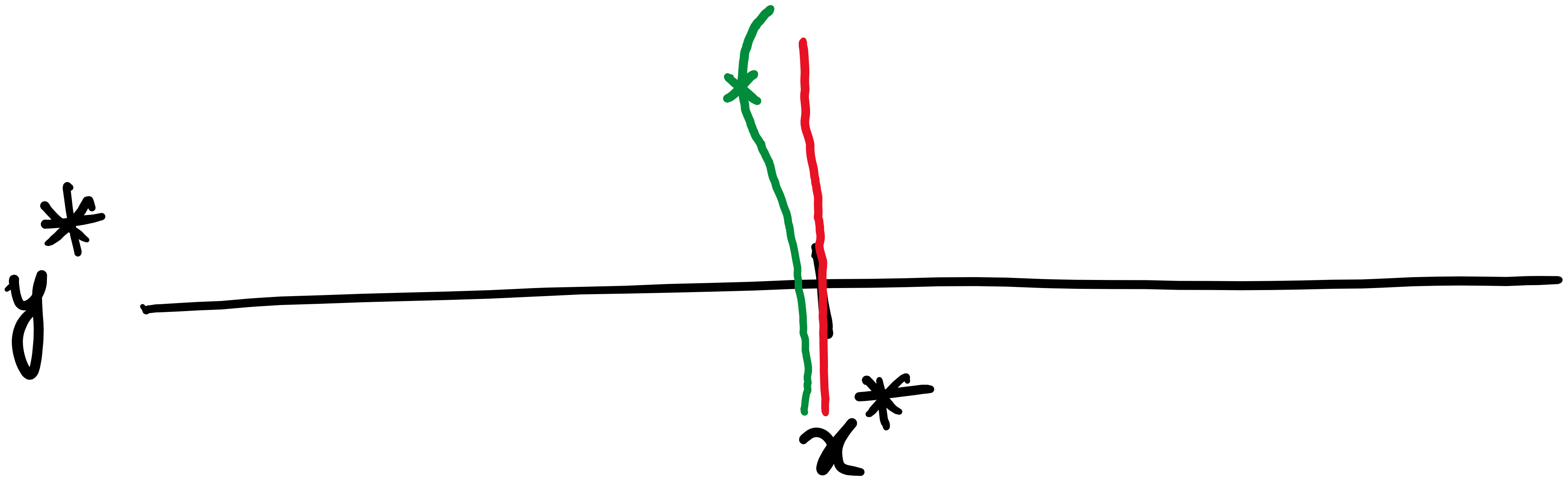
$y^* \in \mathrm{LocalMax}_y f(x^*, y)$ and $x^* \in \mathrm{LocalMin}_x \left(\underbrace{\mathrm{LocalMax}_{y \in B(y^*)} f(x, y)}_{\text{Problematic}} \right)$



Towards local Stackelberg equilibrium

$y^* \in \mathrm{LocalMax}_y f(x^*, y)$ and $x^* \in \mathrm{LocalMin}_x \left(\underset{y \in B(y^*)}{\mathrm{LocalMax}} f(x, y) \right)$

- $\underset{y \in B(y^*)}{\mathrm{LocalMax}} f(x, y)$ can be achieved far away from y^*



Local Stackelberg equilibrium

$y^* \in \mathrm{LocalMax}_y f(x^*, y)$ and $x^* \in \mathrm{LocalMin}_x \left(\max_{y \in B_\epsilon(y^*)} f(x,y) \right)$

$$g_{\epsilon,y^*}(x) \stackrel{\text{def}}{=} \max_{y \in B_\epsilon(y^*)} f(x,y)$$

Can also call it

Local minimax

$$x^* \in \mathrm{LocalMin}_x \, g_{\epsilon,y^*}(x) \quad \forall \; \epsilon \le \epsilon_0$$

Also independently by [Fiez et al. 2019]

Some results on local minimax

- May also not exist e.g., $y^2 - 2xy$ on $[-1,1] \times [-1,1]$
  - Reason: Set of local maxima (in y) is discontinuous as a (set) function of x

- Local Nash equilibria $\subseteq$ Local minimax

- Since global minimax always exists, it implies global minimax not always local minimax

First and second order conditions

- First order: $\nabla_x f(x^*, y^*) = 0$ and $\nabla_y f(x^*, y^*) = 0$
- 2nd order sufficient:

$$\nabla^2_{xx} f - \nabla_{xy} f \left(\nabla_{yy} f\right)^{-1} \nabla_{yx} f \succ 0 \quad \text{and} \quad \nabla^2_{yy} f(x^*, y^*) \prec 0$$

- Also need not exist
  - 2nd order Nash ⊆ 2nd order local minimax

Quick recap

- Existing notions of local optimality for minimax problems inspired by equilibrium notions for simultaneous games
- We introduce a new notion of local optimality inspired by equilibrium notion for sequential games; more relevant for nonconvex-nonconcave settings
- Local minimax suffers from nonexistence issues but
  - Local Nash $\subseteq$ Local minimax
  - When it exists, it is more relevant for practical minimax problems

Algorithms

Gradient descent ascent

$$x_{t+1} = x_t - \eta \nabla_x f(x_t, y_t)$$
$$y_{t+1} = y_t + \eta \nabla_y f(x_t, y_t)$$

- Algorithm again inspired by simultaneous games
- In practice, multiple steps of ascent for one step of descent – signifying the order $\min_x \max_y f(x, y)$
- Need not converge – could cycle; several alternatives proposed e.g., optimistic gradient methods, extra gradient methods etc. [Nemirovski 2004; Daskalakis et al. 2017]

Fixed points of gradient descent ascent

- Widely used in practice with out any averaging
- Motivates the study of fixed points
- We study the flow version of γ-GDA

$$\dot{x} = -\nabla_x f(x, y)$$
$$\dot{y} = +\gamma \nabla_y f(x, y)$$

- γ indicates the number of ascent steps per descent steps

Stable fixed points

- A fixed point of a dynamical system (such as γ-GDA flow) is called stable if points close to it converge to it.
    - Jacobian of the dynamical system has spectral radius < 1.

- Set of stable fixed points changes with γ

- For $\min_x \max_y f(x, y)$, we are interested in $\gamma \gg 1$

Main result

Local minimax points $\cong$ Stable fixed points of ∞-GDA

- Equality holds up to some degenerate points
- Gives a game theoretic meaning to limit points of GDA
- Can extend results to approximate local minimax points and stable fixed points of $\gamma(< \infty)$-GDA

Summary

- Minimax optimization/Two player zero sum games important
- Very little understanding in nonconvex-nonconcave setting
  - Sequential games quite important
- Propose local notions of sequential equilibria; existing works only do for simultaneous equilibria
- Show a close relationship between local minimax and GDA

Future directions

- Our results unsatisfactory – nonexistence a serious issue
- Other notions of local optimality?
  - Computational restrictions on the adversary
- Is convergence to a point important? Can we harness limit cycles in the nonconvex-nonconcave setting?
- Better algorithms?