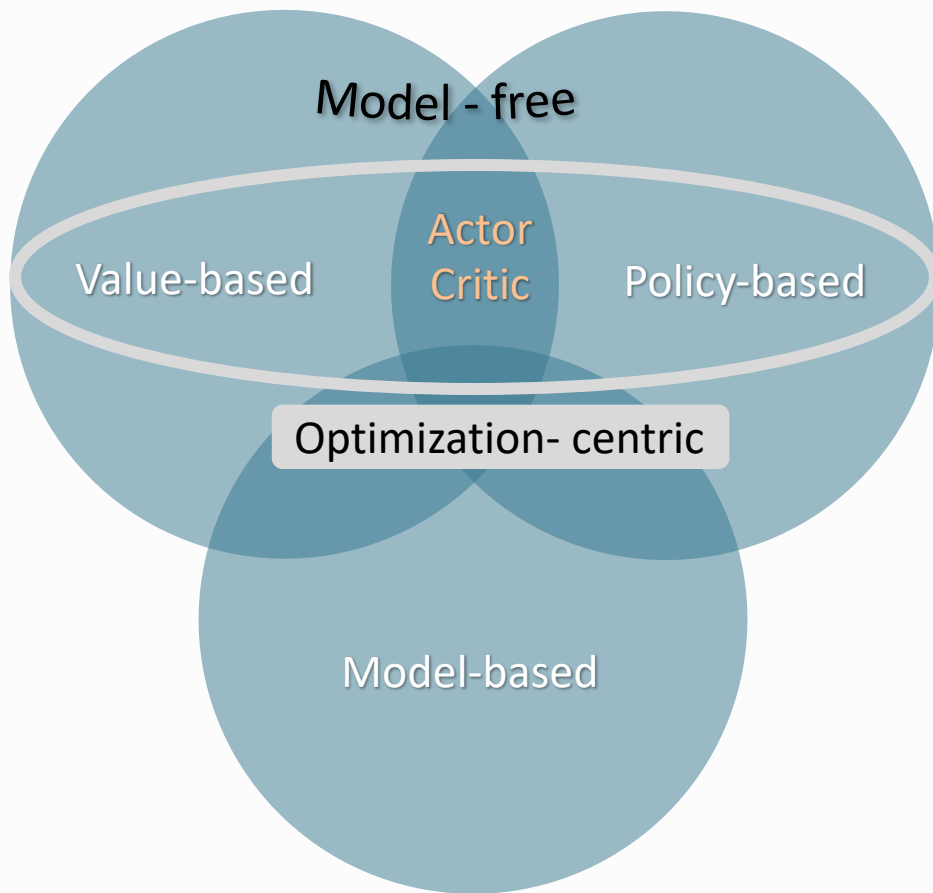


EPFL Summer School on Data Science, Optimization and Operations Research
August 15-20, 2021

Lecture 5: RL from Game Perspectives

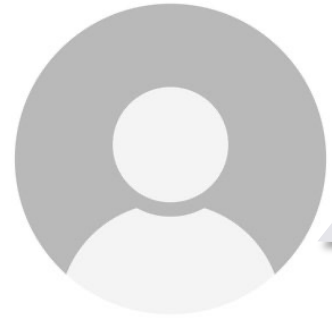
Niao He, D-INFK, ETH Zurich

Recap: Reinforcement Learning Approaches



- **Value-based RL**
 - Estimate the optimal value function $Q^*(s, a)$
 - Example: Q-learning
- **Policy-based RL**
 - Search directly the optimal policy $\pi^*(\cdot | s)$
 - Example: Policy Gradient Method
- **Model-based RL**
 - First estimate the model P, R and then do planning

Future of AI



“ Minimization is to current AI,
_____ is to future AI”

Min-max Optimization
Equilibrium Learning



+1



Outline of Lecture Series

Lecture 1

Introduction to RL

Lecture 2

RL from Control Perspectives
- Value-based RL

Lecture 3

RL from Optimization Perspectives
- *Policy-based RL*

Lecture 4

RL from Learning Perspectives
- *Deep RL*

Lecture 5

RL from Game Perspectives



John von Neumann

Zero-sum game $\min_{x \in X} \max_{y \in Y} f(x, y)$

Stackelberg game $\min_{x \in X} f(x, y)$
s.t. $y \in \operatorname{argmin}_{y \in Y} g(x, y)$

Outline of Lecture Series

Lecture 1

Introduction to RL

Lecture 2

RL from Control Perspectives
- Value-based RL

Lecture 3

RL from Optimization Perspectives
- *Policy-based RL*

Lecture 4

RL from Learning Perspectives
- *Deep RL*

Lecture 5

RL from Game Perspectives

- Capture many RL algorithms
- Widespread in many RL settings : imitation learning, multi-agent RL, meta-RL, etc.
- Profound applications in OR/ML domains: robust optimization, GANs, etc.

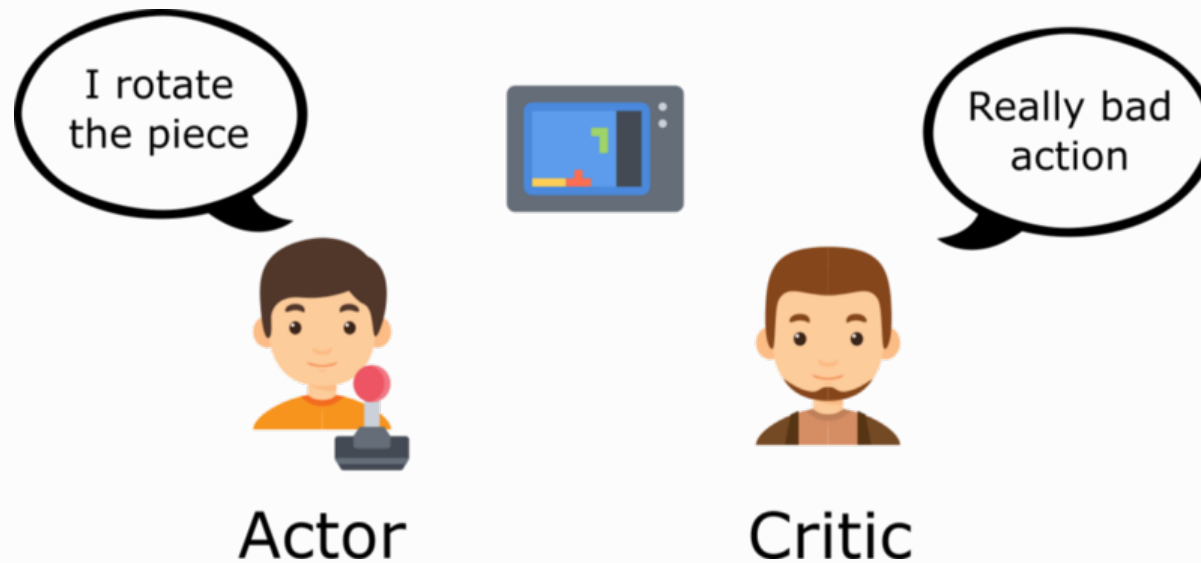
Zero-sum game $\min_{x \in X} \max_{y \in Y} f(x, y)$

Stackelberg game $\min_{x \in X} f(x, y)$
s.t. $y \in \operatorname{argmin}_{y \in Y} g(x, y)$

Actor-Critic RL

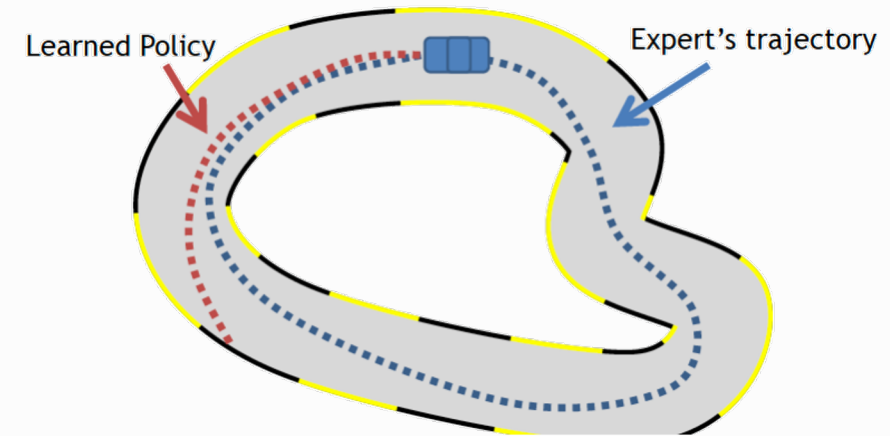
$$\begin{aligned} \max_{\pi} \quad & V^{\pi}(\mu) = \mathbb{E}_{\mu}[\langle Q_{\theta^*(\pi)}, \pi \rangle] \\ \text{s.t.} \quad & \theta^*(\pi) \in \operatorname{argmin}_{\theta \in \mathbb{R}^d} \|Q_{\theta} - R^{\pi} - \gamma P^{\pi} Q_{\theta}\|_{d_{\mu}^{\pi}}^2 \end{aligned}$$

$Q_{\theta}(s, a) = \phi(s, a)^{\top} \theta$ or neural nets



Imitation Learning

- Reinforcement learning
 - receive reward as a feedback signal
 - learn from interaction
- Imitation learning
 - receive no feedback signal
 - learn from expert demonstrations



Inverse Reinforcement Learning

$$\max_{c \in \mathcal{C}} \min_{\pi \in \Pi} \mathbb{E}_{\pi} [c(s, a)] - \mathbb{E}_{\pi_E} [c(s, a)] - H(\pi)$$

Expert policy: π_E

$$\text{Entropy: } H(\pi) := \mathbb{E}_{\pi} [-\log \pi(a|s)]$$

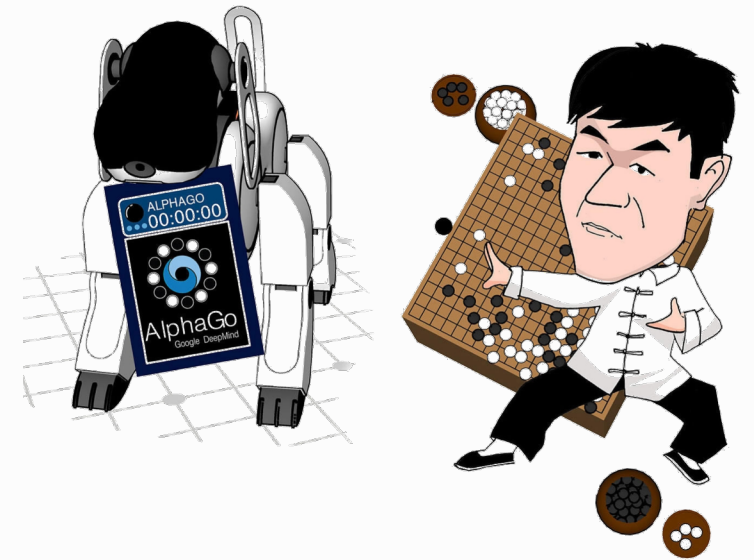
Two Player Markov Game

MDP Setting:

- Shared state $s \in S$
- Joint action $(a, b) \in A_1 \times A_2$
- Transition probability: $P(s'|s, a, b)$
- Rewards: $r_1(s, a, b) = -r_2(s, a, b) = r(s, a, b)$
- Policies: $\pi_1(a|s), \pi_2(b|s)$
- Agent 1 wants to minimize its value function
- Agent 2 wants to maximize its value function

$$V^{\pi_1, \pi_2}(s) = \mathbb{E}_{\substack{a_t \sim \pi_1(\cdot|s_t) \\ b_t \sim \pi_2(\cdot|s_t) \\ s_{t+1} \sim \mathcal{P}(\cdot|s_t, a_t, b_t)}} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t, b_t) \mid s_0 = s \right]$$

$$\min_{\pi_1} \max_{\pi_2} V^{\pi_1, \pi_2}(\mu) := \mathbb{E}_{s \sim \mu} [V^{\pi_1, \pi_2}(s)]$$



Equilibrium

(Zero-sum game)

$$\min_{x \in X} \max_{y \in Y} f(x, y)$$

(Stackelberg game)

$$\begin{aligned} \min_{x \in X} f(x, y) \\ \text{s.t. } y \in \operatorname{argmin}_{y \in Y} g(x, y) \end{aligned}$$

Nash equilibrium

$(x^*, y^*) \in X \times Y$ is a Nash equilibrium if
 $\forall x \in X, y \in Y :$

$$f(x^*, y) \leq f(x^*, y^*) \leq f(x, y^*)$$

Stackelberg equilibrium

Best response to best response: $\forall x \in X$

$$\min_{y \in BR(x^*)} f(x^*, y) \leq \min_{y \in BR(x)} f(x, y)$$

$$BR(x) := \{y : g(x, y) = \min_{y' \in Y} g(x, y')\}$$

The golden case

(Zero-sum game)

$$\min_{x \in X} \max_{y \in Y} f(x, y)$$

Minimax Theorem (von Neumann 1928)

If X and Y are convex and compact, f is continuous, convex in x , concave in y , then

$$\min_{x \in X} \max_{y \in Y} f(x, y) = \max_{y \in Y} \min_{x \in X} f(x, y).$$

(Stackelberg game)

$$\begin{aligned} &\min_{x \in X} f(x, y) \\ &\text{s.t. } y \in \operatorname{argmin}_{y \in Y} g(x, y) \end{aligned}$$

Existence of Stackelberg equilibrium

A Stackelberg equilibrium exists if f, g are continuous, X and Y are compact, and $g(x, \cdot)$ is strictly convex for all x .

See e.g., (Basar & Olsder, 1998)

Three critical regimes

(Zero-sum game):

$$\min_{x \in X} \max_{y \in Y} f(x, y)$$

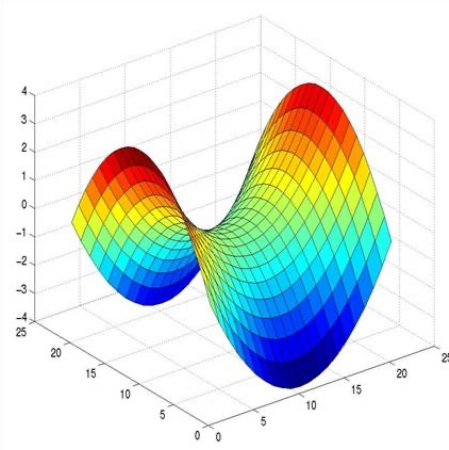
(General-sum game):

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

$$\text{s.t. } y \in \operatorname{argmin}_{y \in Y} g(x, y)$$

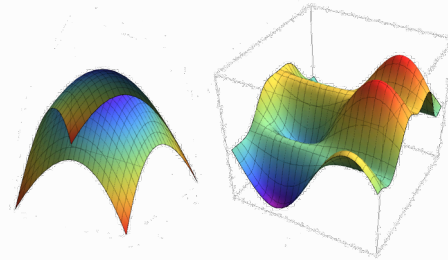
Convex-Concave

Extensive literature



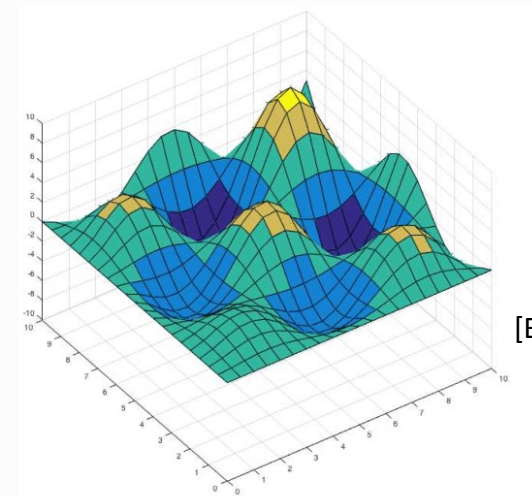
Nonconvex-Concave

Many recent literature



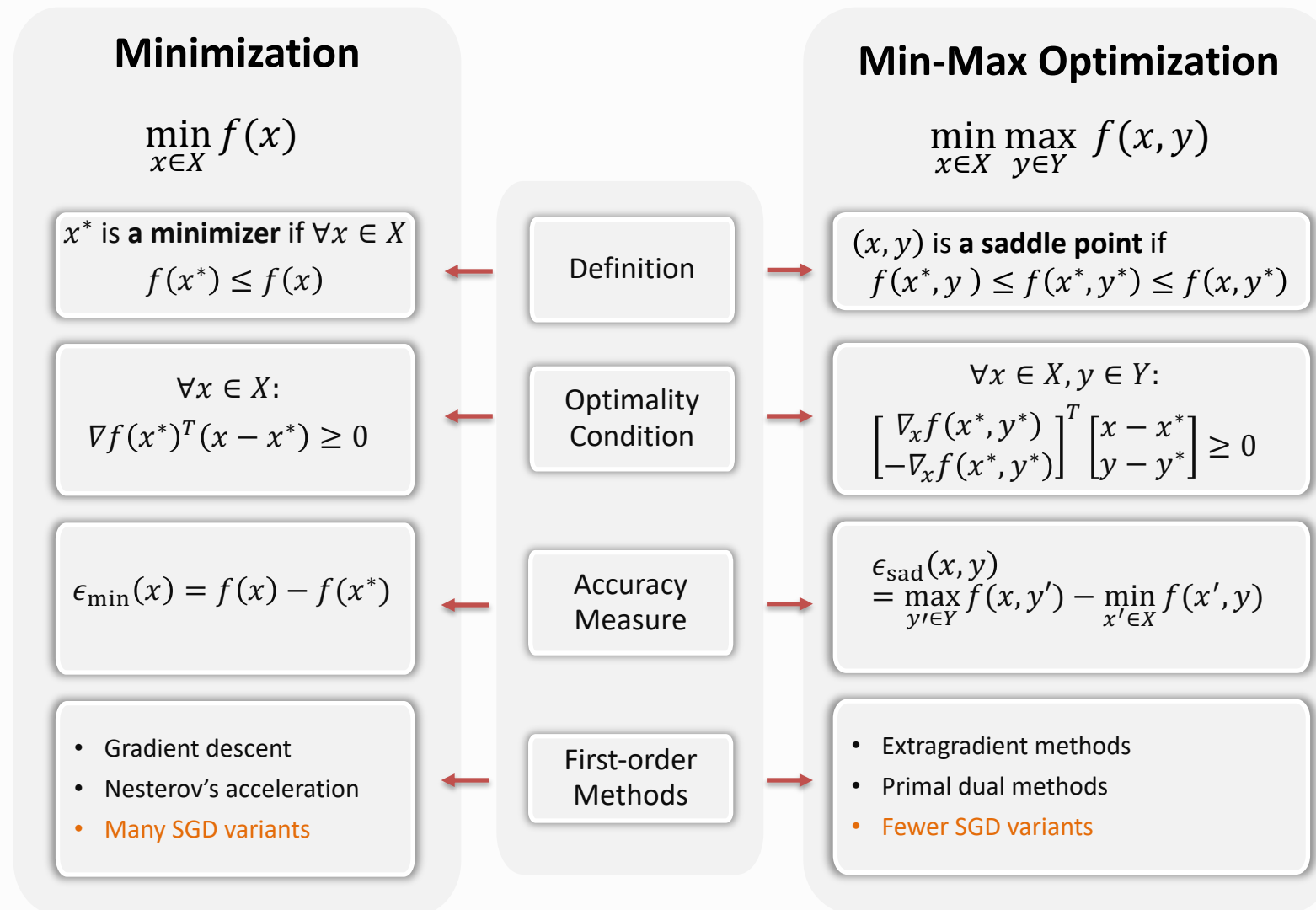
Nonconvex-Nonconcave

Few literature



[Bauso et al. 2017]

Convex-Concave Min-Max Optimization



A Unified Variational Inequality Framework

$$\text{VI}(Z, F) \quad \text{Find } \mathbf{z}^* \in X : \langle F(\mathbf{z}^*), \mathbf{z} - \mathbf{z}^* \rangle \geq 0, \forall \mathbf{z} \in Z$$

- Convex Minimization

$$\min_{x \in X} f(x) \iff \text{VI}(Z, F) \text{ with } \mathbf{z} = x; Z = X; F(\mathbf{z}) = \nabla f(x)$$

- Convex-concave saddle point problems

$$\min_{x \in X} \max_{y \in Y} f(x, y) \iff \text{VI}(Z, F) \text{ with } \mathbf{z} = (x, y); Z = X \times Y; F(\mathbf{z}) = \begin{bmatrix} \nabla_x f(x, y) \\ -\nabla_y f(x, y) \end{bmatrix}$$

- Convex Nash equilibrium problems

$$\min_{x_i \in X_i} f_i(x_i; x_{-i}), i = 1, \dots, K \iff \text{VI}(Z, F) \text{ with } \mathbf{z} = (x_1, \dots, x_K); Z = \prod_{i=1}^K X_i, F(\mathbf{z}) = [\nabla_{x_i} f_i(x_i; x_{-i})]_{i=1}^K$$

For convex problems, the operator F is **monotone**.

Solving VI with monotone operator

Algorithm	Iteration	Assumption	Convergence rate
Projection method [Sibony'70]	$\mathbf{z}_{t+1} = \Pi_Z(\mathbf{z}_t - \eta F(\mathbf{z}_t))$	Lipschitz + Strong Monotone	$O(e^{-t})$
Extragradient method [Korpelevich'76] [Nemirovski'04]	$\tilde{\mathbf{z}}_t = \Pi_Z(\mathbf{z}_t - \eta F(\mathbf{z}_t))$ $\mathbf{z}_{t+1} = \Pi_Z(\mathbf{z}_t - \eta F(\tilde{\mathbf{z}}_t))$	Lipschitz + Pseudo Monotone	$O(1/t)$ (ergodic)
Proximal point method [Martinet'70] [Rockafellar'76]	$\mathbf{z}_{t+1} = \text{solution to } \text{VI}(X, F + \eta(I - \mathbf{z}_t))$	(Strong) Monotone	$O(1/t)$ or $O(e^{-t})$

- Monotonicity: $\langle F(\mathbf{z}) - F(\mathbf{z}'), \mathbf{z} - \mathbf{z}' \rangle \geq 0, \forall \mathbf{z}, \mathbf{z}' \in Z$
- Strong monotonicity: $\langle F(\mathbf{z}) - F(\mathbf{z}'), \mathbf{z} - \mathbf{z}' \rangle \geq \mu \|\mathbf{z} - \mathbf{z}'\|$
- Pseudo monotonicity: $\langle F(\mathbf{z}'), \mathbf{z} - \mathbf{z}' \rangle \geq 0 \Rightarrow \langle F(\mathbf{z}), \mathbf{z} - \mathbf{z}' \rangle \geq 0$

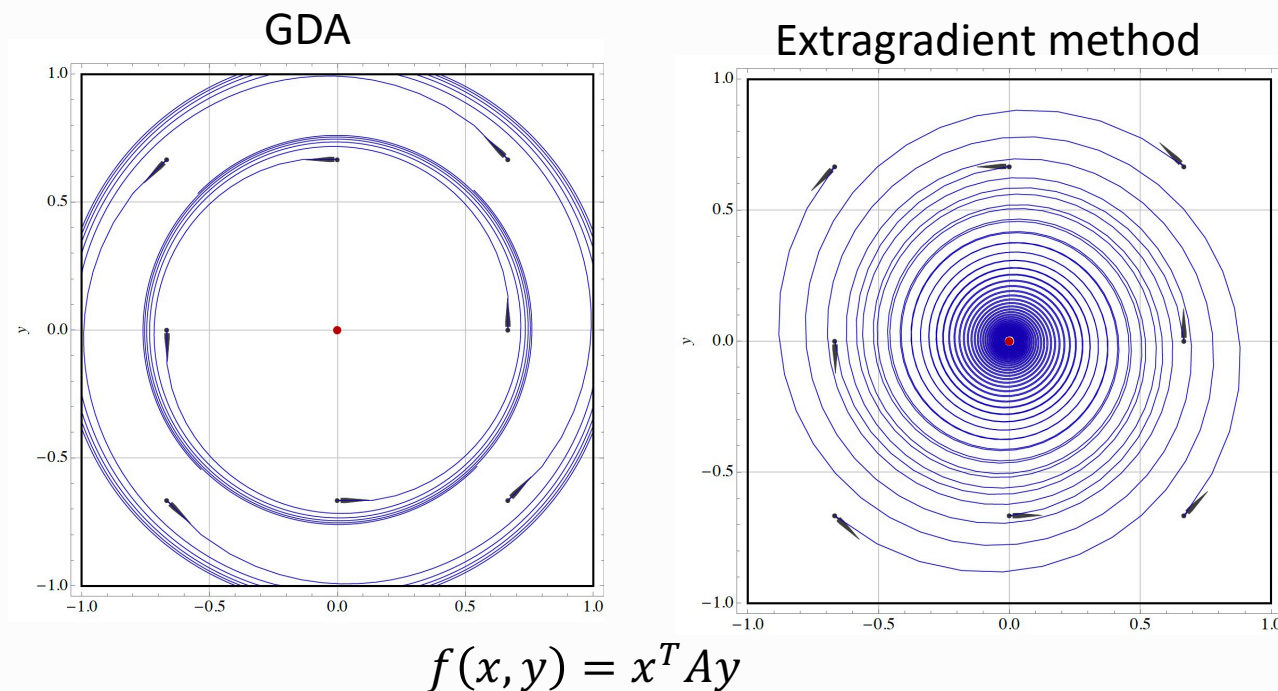
Algorithms for Min-Max Optimization

$$\min_{x \in X} \max_{y \in Y} f(x, y) \iff \text{VI}(Z, F) \text{ with } \mathbf{z} = (x, y); Z = X \times Y; F(\mathbf{z}) = \begin{bmatrix} \nabla_x f(x, y) \\ -\nabla_y f(x, y) \end{bmatrix}$$

(for simplicity, assume unconstrained):

- Gradient Descent Ascent (GDA): $\mathbf{z}_{t+1} = \mathbf{z}_t - \eta F(\mathbf{z}_t)$
- Proximal Point Algorithm (PPA): $\mathbf{z}_{t+1} = \mathbf{z}_t - \eta F(\mathbf{z}_{t+1})$
- Extragradient method (EG) $\mathbf{z}_{t+1} = \mathbf{z}_t - \eta F(\mathbf{z}_t - \eta F(\mathbf{z}_t))$
- Optimistic Gradient Descent Ascent (OGDA): $\mathbf{z}_{t+1} = \mathbf{z}_t - \eta(2F(\mathbf{z}_t) - F(\mathbf{z}_{t-1}))$
- Reflected-Forward-Backward Splitting (RFBS): $\mathbf{z}_{t+1} = \mathbf{z}_t - \eta F(2\mathbf{z}_t - \mathbf{z}_{t-1})$

GDA vs. Extragradient Method



- GDA with a constant stepsize could diverge or oscillate in cycles for bilinear problems.
- GDA with a decaying stepsize converges, but slower than extragradient.
- Last iterate of extragradient converges slower than the average solution. [Golowich et al, 2020].

*Figure source: Hsieh, Mertikopoulos, Cevher. The limits of min-max optimization algorithms. ICML 2021.

Recent Advance

- For convex-concave min-max optimization, the VI perspective might not capture refined structure.
 - Strong convexity: μ_x -(strongly)-convex, μ_y -(strongly)-concave
 - Lipschitz constants: L_{xx}, L_{xy}, L_{yy}
 - Linear coupling term
- Achieving the optimal complexity in each setting requires more sophisticated algorithms.

Problem Class (smooth objective)	Lower Bound (duality gap)	(near)-Optimal Algorithm
Convex-Concave	$O(L/\epsilon)$	Extragradient / Mirror Prox/OGDA [Nemirovski'2004; Daskalakis et al., 2018]
Strongly-Convex-Concave	$O(\sqrt{\kappa_x/\epsilon})$ [Ouyang and Xu, 2019]	DIAG/Minimax-APP/Catalyst [Thekumparampil et al. 2019; Lin et al. 2020; Yang, Zhang, Kiyavash, H., 2020]
Strongly-Convex-Strongly-Concave	$O(\sqrt{\kappa_x \kappa_y} \log 1/\epsilon)$ [Zhang et al 2019]	Minimax-APPA/Catalyst [Lin et al. 2020; Yang, Zhang, Kiyavash, H., 2020]

Notions of Optimality

Global solution:

- Global Nash equilibrium does not always exist.
- Even if it exists, finding global Nash equilibrium is NP-hard in general; indeed, PPAD-complete [Daskalakis et al. 2008]
- A global minimax point (Stackelberg equilibrium) always exists, but finding it is also NP-hard.

Local solution:

- Local Nash equilibrium may not exist in general. Example: $f(x, y) = \sin(x + y)$. [Jin et al. 2020]
- Even local minimax points may not exist. Example: $f(x, y) = y^2 - 2xy$. [Jin et al. 2020]
- The computation of such a point, even if exist, is also NP-hard. [Daskalakis et al. 2020]

Nonconvex-Concave Min-Max Optimization

Nonconvex Minimization

$$\min_x f(x)$$

- **ϵ -stationary point:**

$$\|\nabla f(x)\| \leq \epsilon$$

- **Gradient Descent:**

$$x_{t+1} = x_t - \eta \nabla f(x_t)$$

- **Convergence** [Nesterov, 2004]:

$$T(\epsilon) = O(\ell/\epsilon^2)$$

Nonconvex-Concave Minimization

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

- **ϵ -stationary point:**

$$\|\nabla \phi(x)\| \leq \epsilon; \quad \phi(x) = \max_y f(x, y)$$

- **Gradient Descent Ascent:**

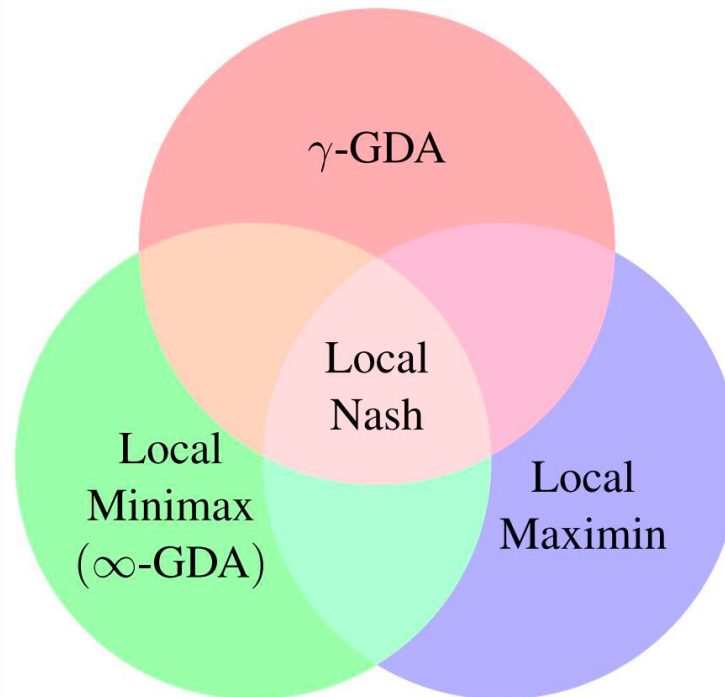
$$\begin{aligned} x_{t+1} &= x_t - \eta_x \nabla_x f(x_t, y_t) \\ y_{t+1} &= y_t + \eta_y \nabla_y f(x_t, y_t) \end{aligned}$$

- **Convergence** [Lin, Jin, Jordan, 2020a]:

$$T(\epsilon) = \begin{cases} O(\kappa^2/\epsilon^2) & \text{NC-SC} \\ O(\epsilon^{-6}) & \text{NC-C} \end{cases}$$

Where does GDA dynamics converge to?

$$\gamma = \eta_y / \eta_x$$



[Daskalakis and Panageas, 2018; Jin et al., 2020; Fiez and Ratli, 2020]

Can we improve over GDA?

- Yes!

- For example, Minimax-PPA [Lin, Jin, & Jordan, 2020b]: $\mathcal{O}\left(\frac{\ell^2 D_{\mathbf{y}} \Delta_{\Phi}}{\epsilon^3} \log^2\left(\frac{\ell(\tilde{D}_{\mathbf{x}}^2 + D_{\mathbf{y}}^2)}{\epsilon}\right)\right)$

Algorithm 5 MINIMAX-PPA($g, \mathbf{x}_0, \mathbf{y}_0, \ell, \mu_{\mathbf{y}}, \epsilon, T$)

- 1: **Input:** initial point $\mathbf{x}_0, \mathbf{y}_0$, proximity ℓ , strongly-convex parameter μ , tolerance δ , iteration T .
 - 2: **Initialize:** set $\delta \leftarrow \frac{\epsilon^2}{(10\kappa_{\mathbf{y}})^4 \ell} \cdot \left(\frac{\epsilon}{\ell D_{\mathbf{y}}}\right)^2$.
 - 3: **for** $t = 1, \dots, T$ **do**
 - 4: denote $g_t(\cdot, \cdot)$ where $g_t(\mathbf{x}, \mathbf{y}) := f(\mathbf{x}, \mathbf{y}) + \ell \|\mathbf{x} - \mathbf{x}_{t-1}\|^2$.
 - 5: $\mathbf{x}_t \leftarrow \text{MAXIMIN-AG2}(g_t, \mathbf{x}_0, \mathbf{y}_0, 3\ell, \ell, \mu, \delta)$.
 - 6: **end for**
 - 7: sample s uniformly from $\{1, 2, \dots, T\}$.
 - 8: $\mathbf{y}_s \leftarrow \text{AGD}(-f(\mathbf{x}_s, \cdot), \mathbf{y}_0, \ell, \mu, \delta)$.
 - 9: **Output:** $(\mathbf{x}_s, \mathbf{y}_s)$.
-

Algorithm 3 MAXIMIN-AG2($g, \mathbf{x}_0, \mathbf{y}_0, \ell, \mu_{\mathbf{x}}, \mu_{\mathbf{y}}, \epsilon$)

- 1: **Input:** initial point $\mathbf{x}_0, \mathbf{y}_0$, smoothness ℓ , strongly convex module $\mu_{\mathbf{x}}, \mu_{\mathbf{y}}$ and tolerance $\epsilon > 0$.
 - 2: **Initialize:** $t \leftarrow 0, \tilde{\mathbf{x}}_0 \leftarrow \mathbf{x}_0, \eta \leftarrow \frac{1}{2\kappa_{\mathbf{x}} \ell}, \kappa_{\mathbf{x}} \leftarrow \frac{\ell}{\mu_{\mathbf{x}}}, \kappa_{\mathbf{y}} \leftarrow \frac{\ell}{\mu_{\mathbf{y}}}, \theta \leftarrow \frac{4\sqrt{\kappa_{\mathbf{x}} \kappa_{\mathbf{y}}} - 1}{4\sqrt{\kappa_{\mathbf{x}} \kappa_{\mathbf{y}}} + 1}, \tilde{\epsilon} \leftarrow \frac{\epsilon}{(10\kappa_{\mathbf{x}} \kappa_{\mathbf{y}})^7}$.
 - 3: **repeat**
 - 4: $t \leftarrow t + 1$.
 - 5: $\tilde{\mathbf{x}}_{t-1} \leftarrow \text{AGD}(g(\cdot, \tilde{\mathbf{y}}_{t-1}), \mathbf{x}_0, \ell, \mu_{\mathbf{x}}, \tilde{\epsilon})$.
 - 6: $\mathbf{y}_t \leftarrow \mathcal{P}_{\mathbf{y}}[\tilde{\mathbf{y}}_{t-1} + \eta \nabla_{\mathbf{y}} g(\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1})]$.
 - 7: $\tilde{\mathbf{y}}_t \leftarrow \mathbf{y}_t + \theta(\mathbf{y}_t - \mathbf{y}_{t-1})$.
 - 8: $\mathbf{x}_t \leftarrow \text{AGD}(g(\cdot, \mathbf{y}_t), \mathbf{x}_0, \ell, \mu_{\mathbf{x}}, \tilde{\epsilon})$.
 - 9: **until** $\|\mathbf{y}_t - \mathcal{P}_{\mathbf{y}}(\mathbf{y}_t + \eta \nabla_{\mathbf{y}} g(\mathbf{x}_t, \mathbf{y}_t))\|^2 \leq \frac{\epsilon}{(10\kappa_{\mathbf{x}} \kappa_{\mathbf{y}})^4 \ell}$ is satisfied.
 - 10: **Output:** $\mathcal{P}_{\mathcal{X}}(\mathbf{x}_t - (1/2\kappa_{\mathbf{y}} \ell) \nabla_{\mathbf{x}} g(\mathbf{x}_t, \mathbf{y}_t))$.
-

Algorithm 1 AGD($g, \mathcal{X}, \mathbf{x}_0, \ell, \mu, \epsilon$)

- 1: **Input:** initial point $\mathbf{x}_0 \in \mathcal{X}$, smoothness ℓ , strongly-convex module μ and tolerance $\epsilon > 0$.
 - 2: **Initialize:** set $t \leftarrow 0, \tilde{\mathbf{x}}_0 \leftarrow \mathbf{x}_0, \eta \leftarrow 1/\ell, \kappa \leftarrow \ell/\mu$ and $\theta \leftarrow \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$.
 - 3: **repeat**
 - 4: $t \leftarrow t + 1$
 - 5: $\mathbf{x}_t \leftarrow \mathcal{P}_{\mathcal{X}}[\tilde{\mathbf{x}}_{t-1} - \eta \nabla g(\tilde{\mathbf{x}}_{t-1})]$.
 - 6: $\tilde{\mathbf{x}}_t \leftarrow \mathbf{x}_t + \theta(\mathbf{x}_t - \mathbf{x}_{t-1})$.
 - 7: **until** $\|\mathbf{x}_t - \mathcal{P}_{\mathcal{X}}(\mathbf{x}_t - \eta \nabla g(\mathbf{x}_t))\|^2 \leq \frac{\epsilon}{2\kappa^2(\ell-\mu)}$ is satisfied.
 - 8: **Output:** $\mathcal{P}_{\mathcal{X}}(\mathbf{x}_t - \eta \nabla g(\mathbf{x}_t))$.
-

- But wait, is this desirable for practical implementation?

A Better Recipe: the universal Catalyst framework

- Idea: dual PPA acceleration + existing method M for SC-SC (such as extragradient method)

Algorithm 1 Catalyst for Minimax Optimization

```

1: for all  $t = 1, 2, \dots, T$  do
2:   Set  $z_t = \eta_t v_{t-1} + (1 - \eta_t) y_{t-1}$ 
3:   Apply existing method  $\mathcal{M}$  to find an inexact solution  $(x_t, y_t)$  to the problem:
       
$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \left[ \tilde{f}_t(x, y) := f(x, y) + \frac{\tau_x}{2} \|x - \bar{x}_t\|^2 - \frac{\tau_y}{2} \|y - z_t\|^2 \right]$$

4:    $v_t = y_{t-1} + \frac{1}{\alpha_t} (y_t - y_{t-1})$ ;
5: end for

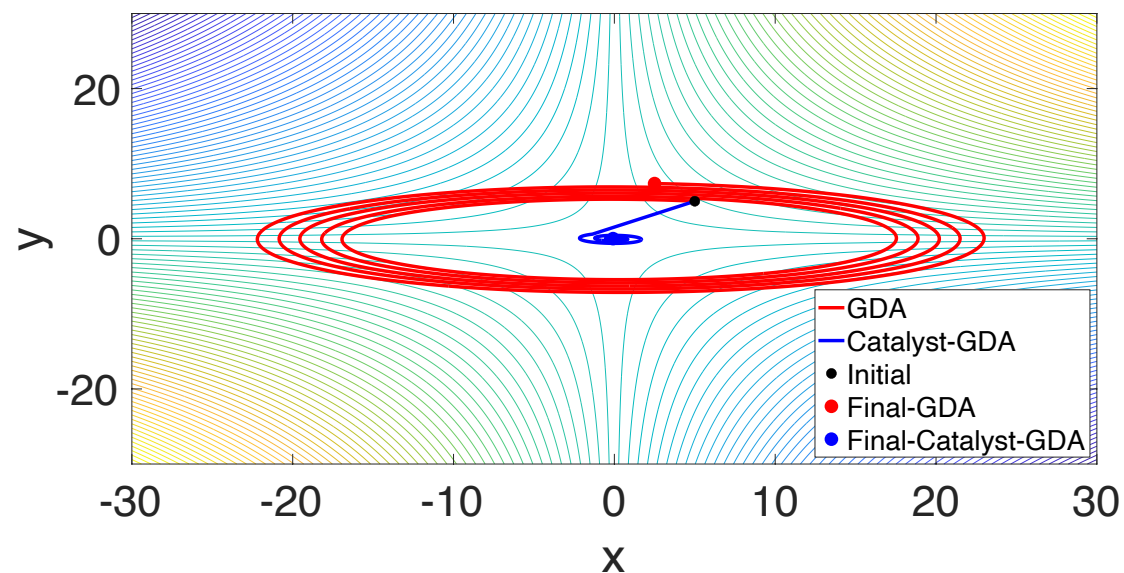
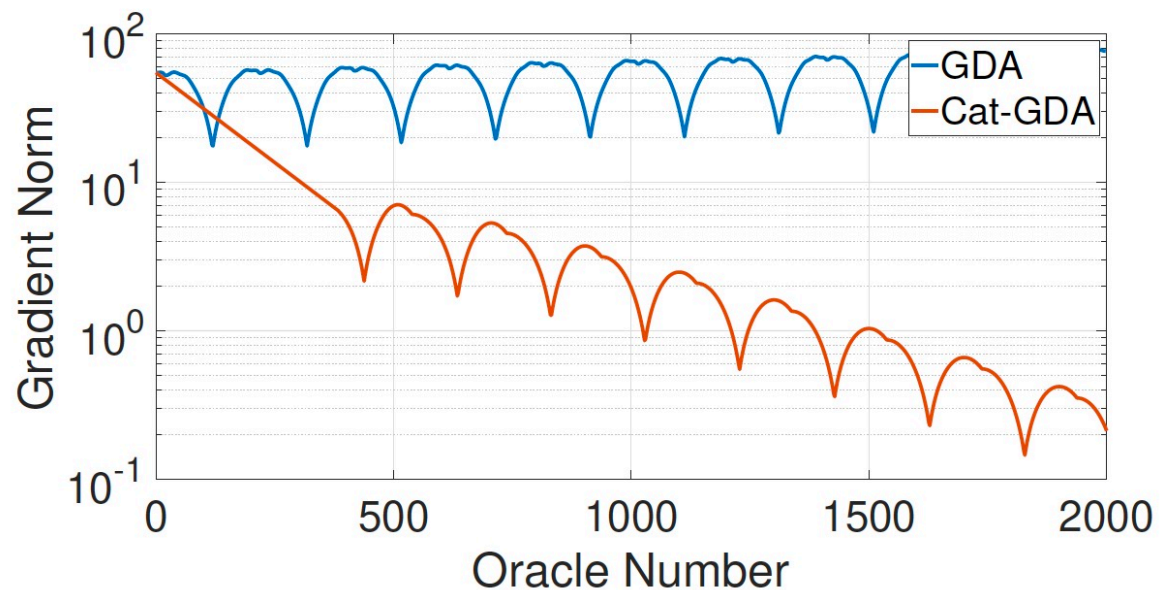
```

Setting	Choice	Catalyst + EG	SOTA complexity	Lower bound
SC-SC	$\tau_x = 0, \tau_y = \mu_x - \mu_y$	$\mathcal{O}(\sqrt{\kappa_x \kappa_y} \log(1/\epsilon))$	$\mathcal{O}(\sqrt{\kappa_x \kappa_y} \log^3(1/\epsilon))$	$\mathcal{O}(\sqrt{\kappa_x \kappa_y} \log(1/\epsilon))$
SC-C	$\tau_x = 0, \tau_y = \mu_x$	$\mathcal{O}(\sqrt{\kappa_x / \epsilon} \log(1/\epsilon))$	$\mathcal{O}(\sqrt{\kappa_x / \epsilon} \log^3(1/\epsilon))$	$\mathcal{O}(\sqrt{\kappa_x / \epsilon})$
NC-SC	$\tau_x = 2\ell, \tau_y = \ell - \mu_y$	$\mathcal{O}(\sqrt{\kappa} \epsilon^{-2})$	$\mathcal{O}(\sqrt{\kappa} \epsilon^{-2} \log^2(1/\epsilon))$	$\mathcal{O}(\sqrt{\kappa} \epsilon^{-2})$
NC-C	$\tau_x = 2\ell, \tau_y = \ell$	$\mathcal{O}(\ell^2 \epsilon^{-3} \log(1/\epsilon))$	$\mathcal{O}(\ell^2 \epsilon^{-3} \log^2(1/\epsilon))$	N/A

Toy Example

$$\min_x \max_y f(x, y) = F(x) + 10xy$$

$$F(x) = \begin{cases} -\frac{1}{2} \cos x - \frac{1}{2}x - \frac{1}{4}\pi & \text{for } x \leq -\frac{\pi}{2}, \\ -\cos x & \text{for } -\frac{\pi}{2} < x \leq \frac{\pi}{2}, \\ -\frac{1}{2} \cos x + \frac{1}{2}x - \frac{1}{4}\pi & \text{for } x > \frac{\pi}{2}. \end{cases}$$



GDA and Catalyst-GDA with constant stepsizes

Nonconvex-Nonconcave Min-Max Optimization

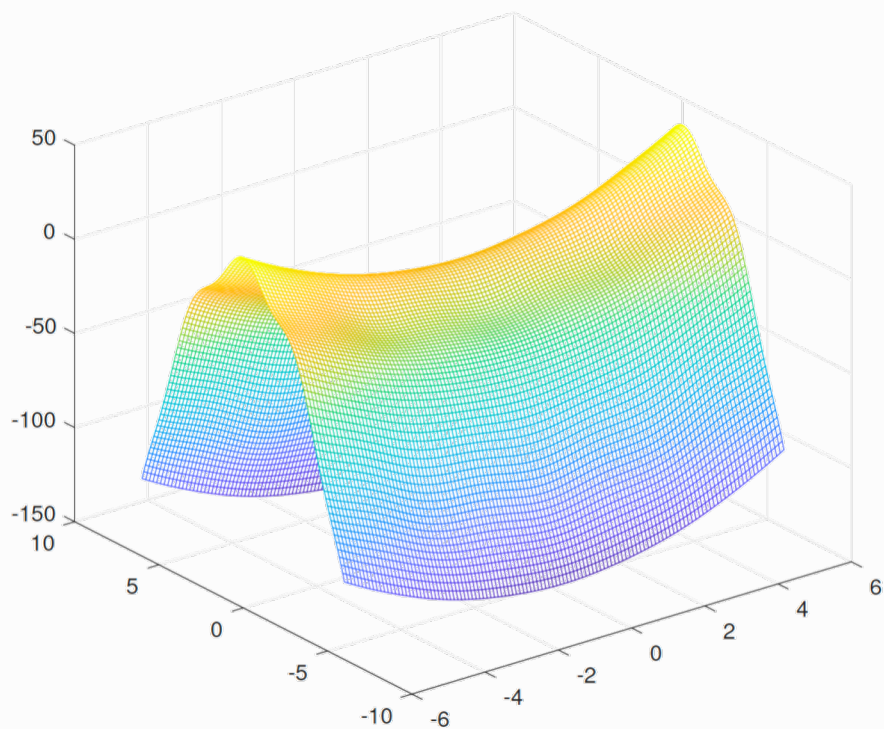
- In general, finding a first-order local Nash equilibrium is NP-hard.
- What should be a reasonable goal here?
 - **Route 1:** Efficient non-polynomial algorithms that converge to local Nash?
 - **Route 2:** Weaker solutions such as first-order stationary points?
 - **Route 3:** Special structure?
(Lojasiewicz conditions, pseudo-monotonicity, variational coherence, interaction dominant regime, etc.)

Two-sided PL Conditions

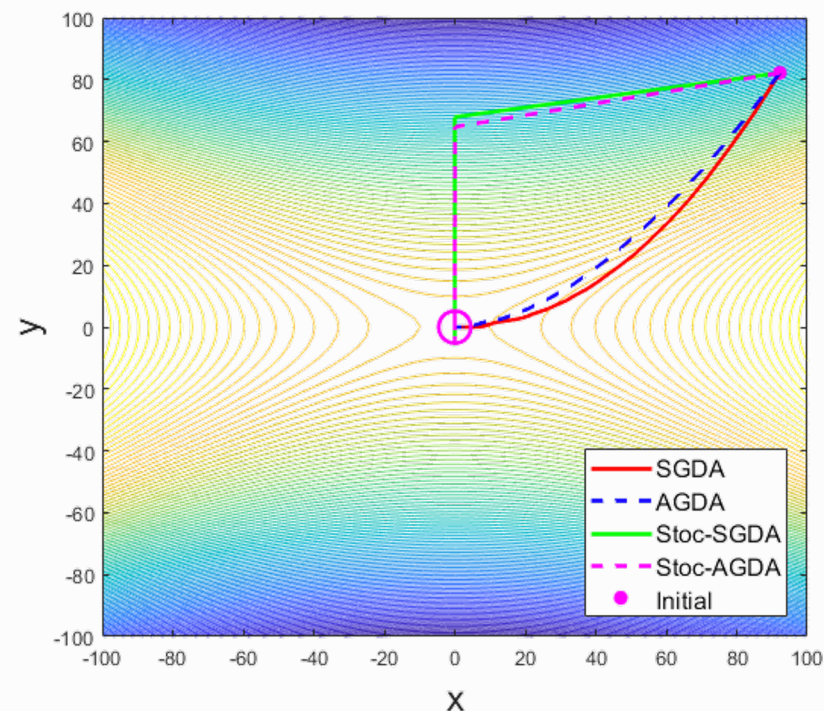
[Yang, Kiyavash, H., 2020]

$$\|\nabla_x f(x, y)\|^2 \geq 2\mu_1[f(x, y) - \min_x f(x, y)]$$

$$\|\nabla_y f(x, y)\|^2 \geq 2\mu_2[\max_y f(x, y) - f(x, y)]$$



$$f(x, y) = x^2 + 3 \sin^2 x \sin^2 y - 4y^2 - 10 \sin^2 y$$



Two-sided PL Conditions

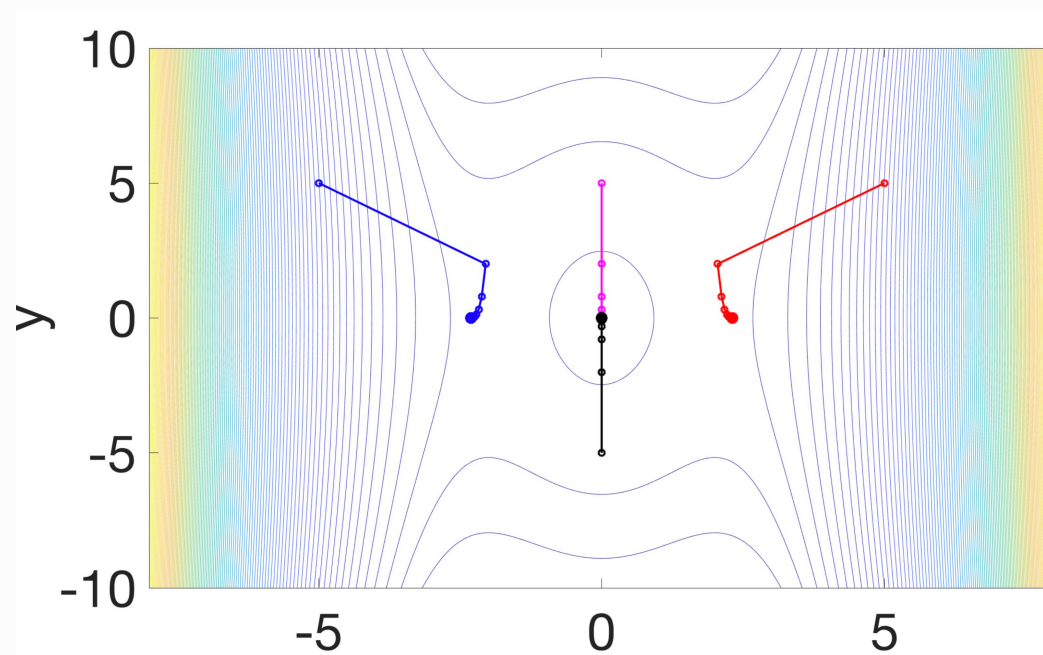
- **Alternating GDA converges linearly to a global minimax point.** [Yang, Kiyavash, H., 2020]
- Application I : Imitation learning for linear quadratic control (LQR)
- **Application II: Zero-sum linear quadratic game (LQG)**

$$\min_{\{u_t\}} \max_{\{v_t\}} \mathbb{E}_{x_0} \left[\sum_{t=0}^{\infty} x_t^T Q x_t + u_t^T R^u u_t - v_t^T R^v v_t \right]$$
$$\text{s.t. } x_{t+1} = Ax_t + Bu_t + Cv_t$$

One-sided PL Condition

$$\|\nabla_y f(x, y)\|^2 \geq 2\mu[\max_y f(x, y) - f(x, y)]$$

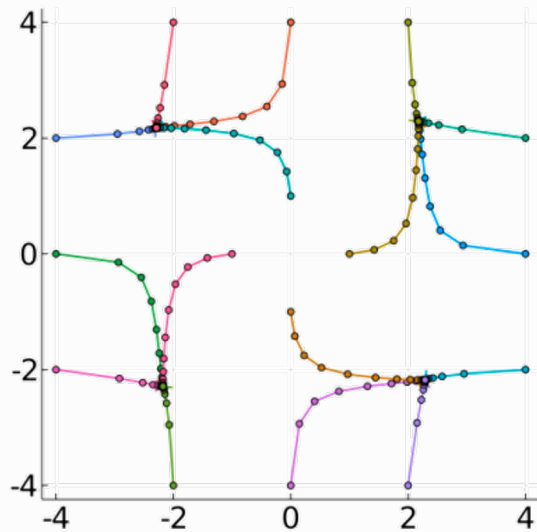
- **Alternating GDA converges to an ϵ -first order equilibrium point within $O(\kappa^2/\epsilon)$ iterations.**
[Yang, Kiyavash, H., 2020]



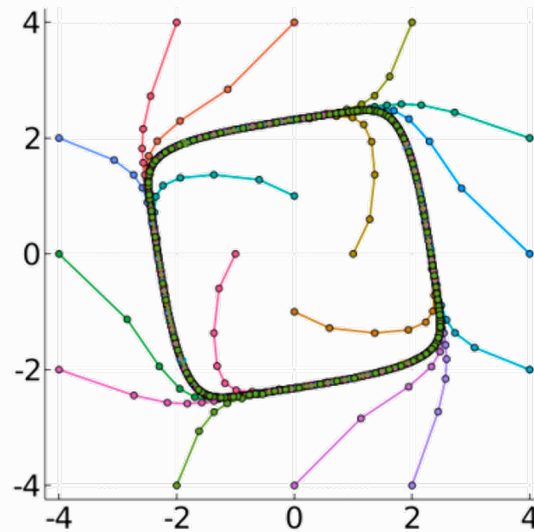
Interaction Dominance

[Grimmer et al., 2020]

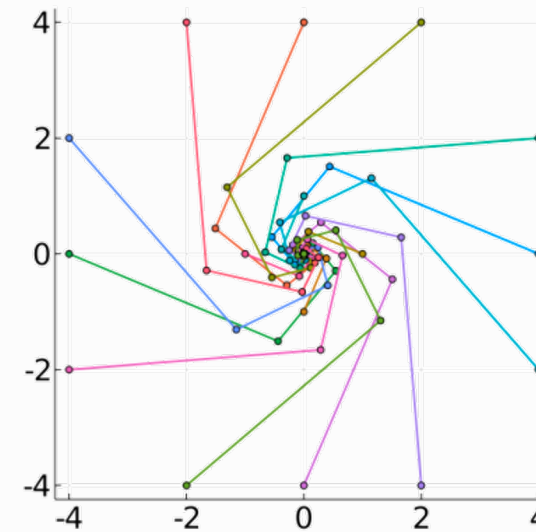
$$\min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m} L(x, y) = f(x) + x^T A y - g(y)$$
$$\nabla^2 f(x) + \frac{A A^T}{2(\beta + \rho)} \succeq \alpha I \quad \left(\text{or} \quad \nabla^2 g(y) + \frac{A^T A}{2(\beta + \rho)} \succeq \alpha I \right)$$



(a) $A = 1$



(b) $A = 10$



(c) $A = 100$

Proximal point algorithm: local linear convergence, divergence, global linear convergence

Open Questions

- Much work needs to be done in the nonconvex-(non)concave regime.
 - Second-order method for local Nash equilibrium or local minimax points?
 - Stochastic and adaptive methods?
 - The landscape of the stable limit points for different algorithms?
 - Avoid spurious critical points?
 - Information-theoretic lower bounds?
 - Fundamental limits of single-loop algorithms?
 - Leverage stochastic dynamics for global optimality?



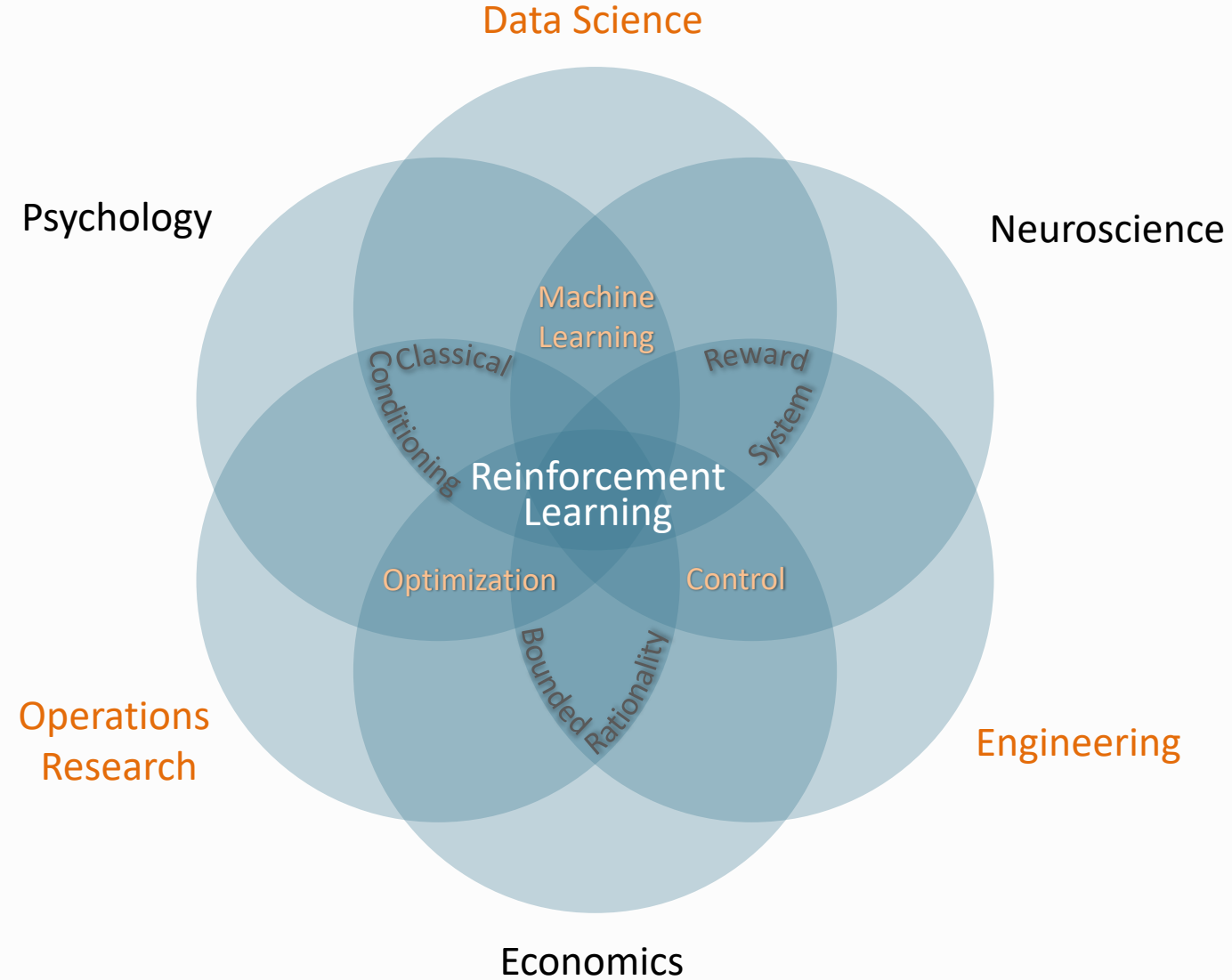
Back to Reinforcement Learning

- Primal-dual RL
- Actor-critic RL
- Inverse RL
- Multi-agent RL
-

- Zero-sum game
- Stackelberg game
- Repeated game
- Markov game
-

Reinforcement
Learning

Game
Theory



Acknowledgement



Semih Cayci
(UIUC)



Donghwan Lee
(KAIST)



Bo Dai
(Google Brain)



R. Srikant
(UIUC)



Junchi Yang
(ETH Zurich)

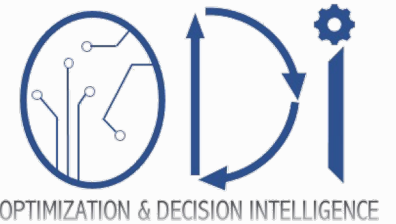


Siqi Zhang
(UIUC)



Negar Kiyavash
(EPFL)

And many other collaborators!



<https://odi.inf.ethz.ch/>