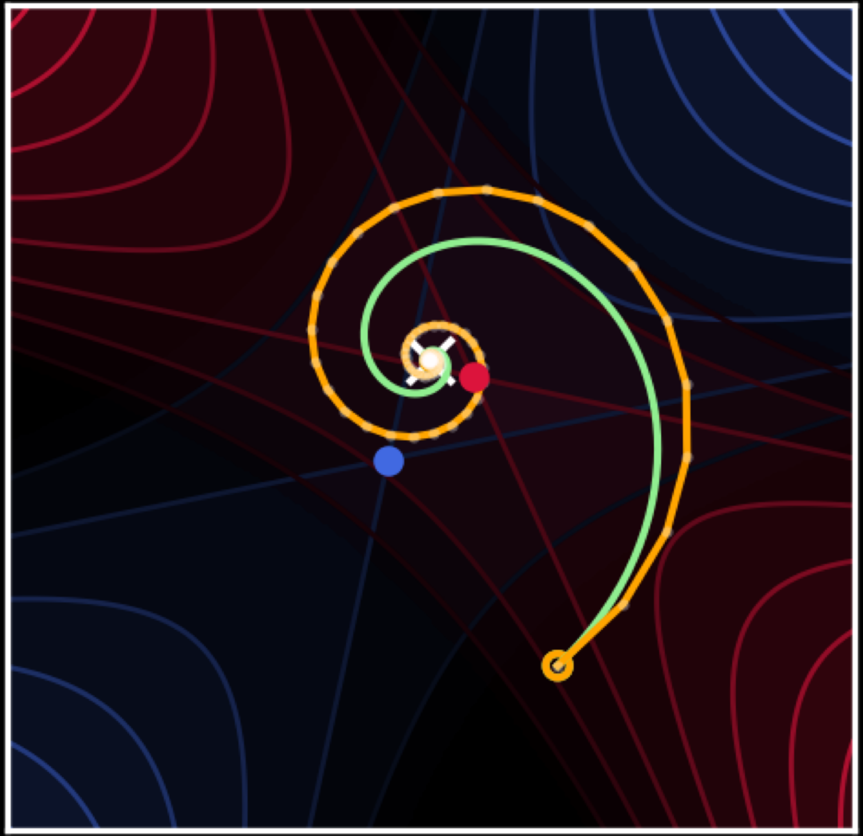
Learning Dynamics of Non-cooperative Agents in Dynamic Environments

Benjamin J. Chasnov

Electrical and Computer Engineering
University of Washington, Seattle WA

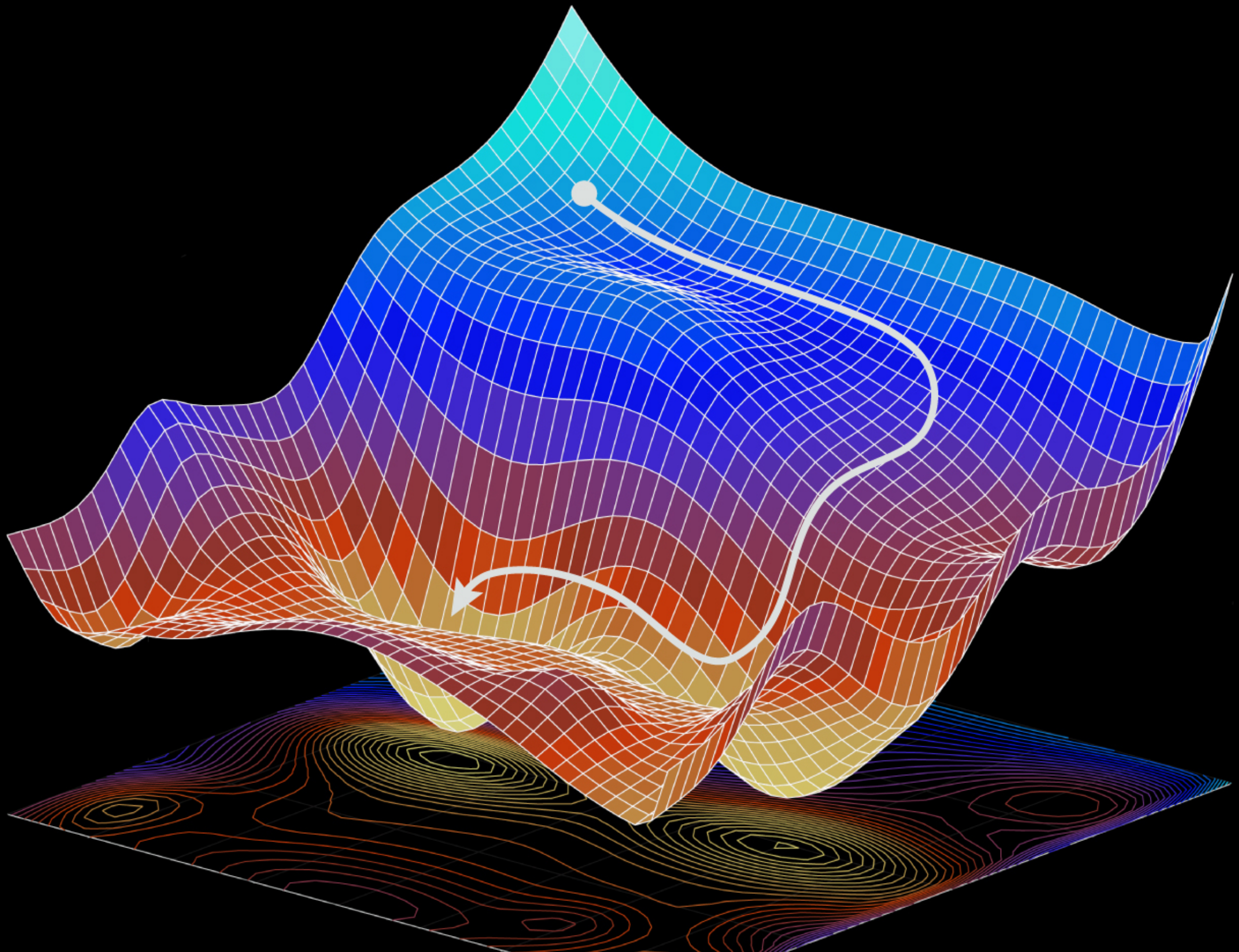
Qualifying Exam, May 2019



Advisors: *Dr. Samuel Burden, Dr. Lillian Ratliff*

Committee: *Dr. Maryam Fazel (chair), Dr. Behçet Açikmeşe, Dr. Kevin Jamieson*

Optimization-based agents will power our society




n agents

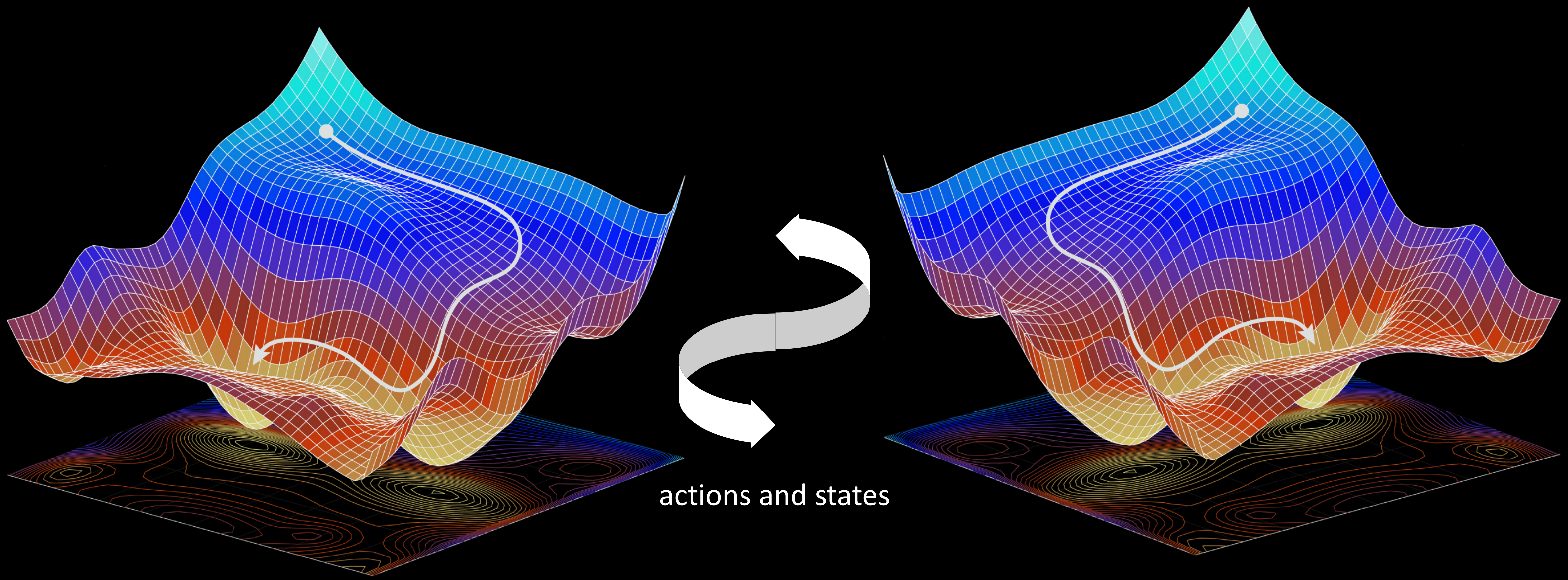
choose actions to minimize **total** cost


$\underset{u=(u_1,\ldots,u_n)}{\text{minimize}}\ \text{cost}(u)$

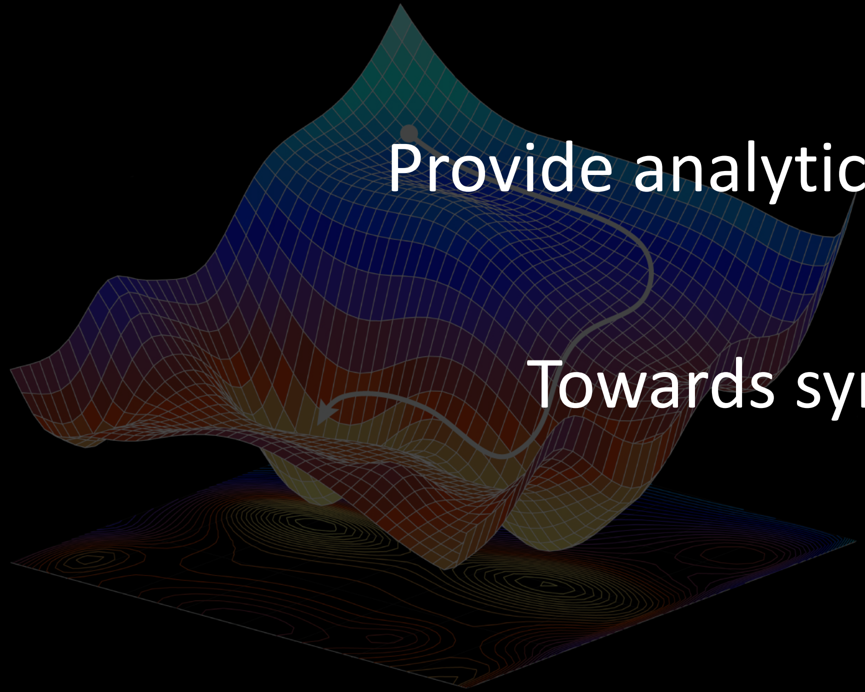
choose actions to minimize **self-interested** cost


$\underset{u_i}{\text{minimize}}\ \text{cost}_i(u)$

Coupled optimization-based agents

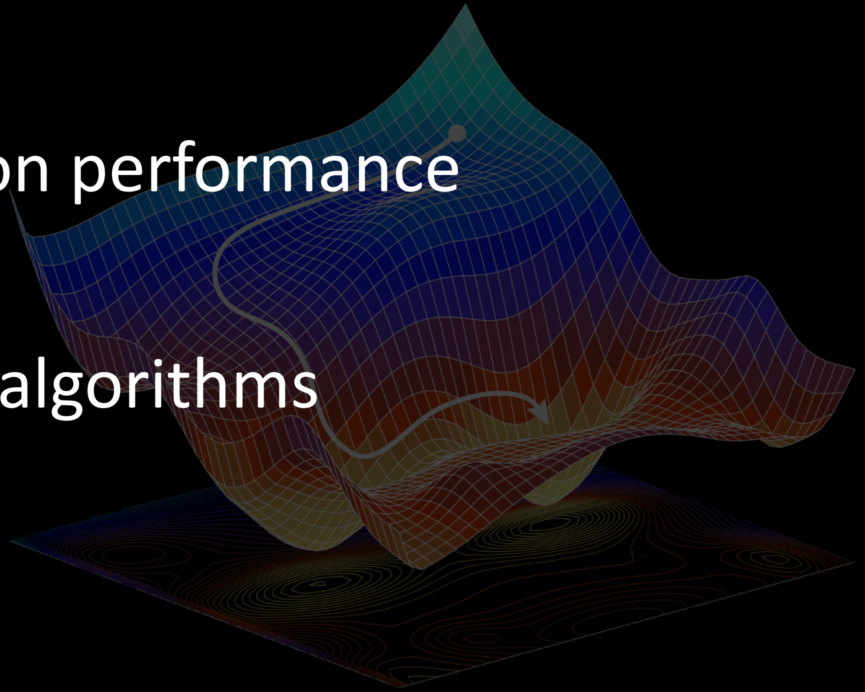


Coupled optimization-based agents



Provide analytical guarantees on performance

Towards synthesis of new algorithms



Example 1: ridesharing

Example 2:

Overview

- **Intro:** Non-cooperative learning agents
- **Part I:** Learning dynamics in games
  - A gradient-based method for solving games
  - Issues (non-Nash attractors, unstable Nash, limit cycles)
- **Part 2:** Towards games in dynamic environments
  - LQ games (feedback policy, open loop control)
  - Stochastic games
- Future extensions

Continuous game (2 players)

A 2-player continuous game consists of a joint action/strategy/choice-variable

$$u = (u_1, u_2) \in U_1 \times U_2 = U$$

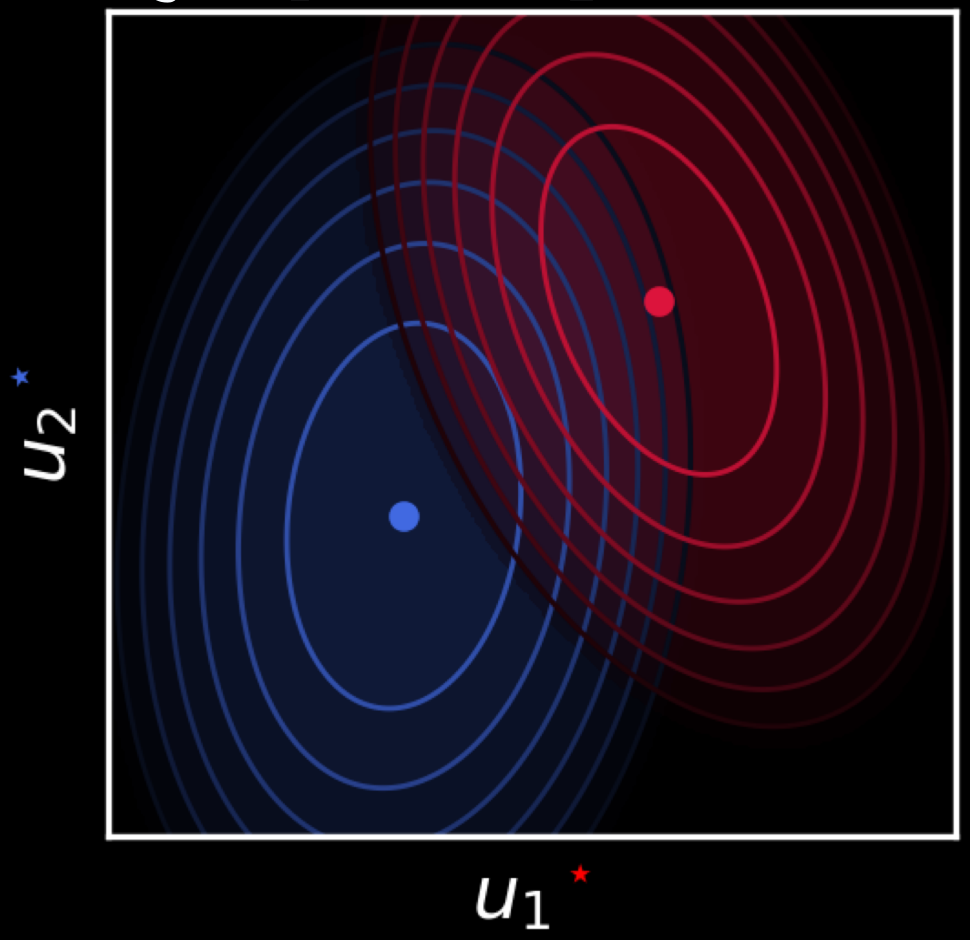
with agent 1's cost function

$$c_1(u) : U \to \mathbb{R}$$

and agent 2's cost function

$$c_2(u) : U \to \mathbb{R}$$

e.g. $U_1 = \mathbb{R}, \; U_2 = \mathbb{R}$



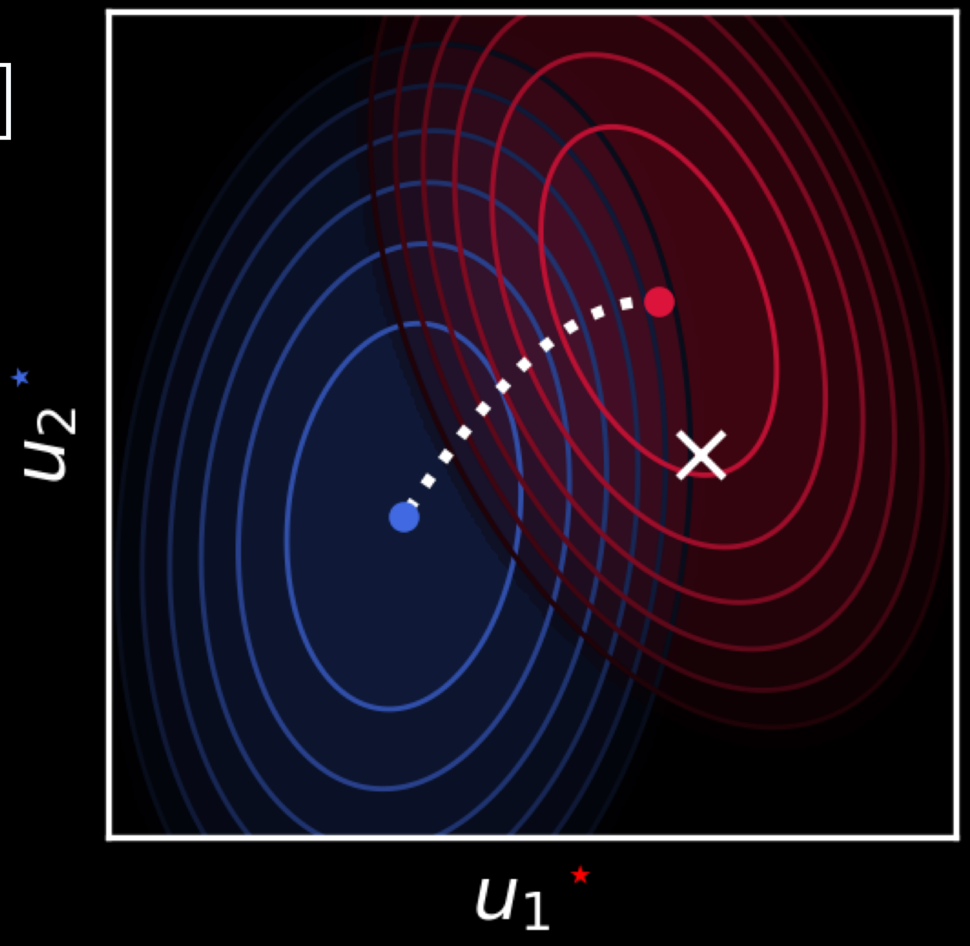
Two different perspectives

Cooperative

$$\min_{u} \; \theta c_1(u) + (1-\theta)c_2(u), \quad \theta \in [0,1]$$

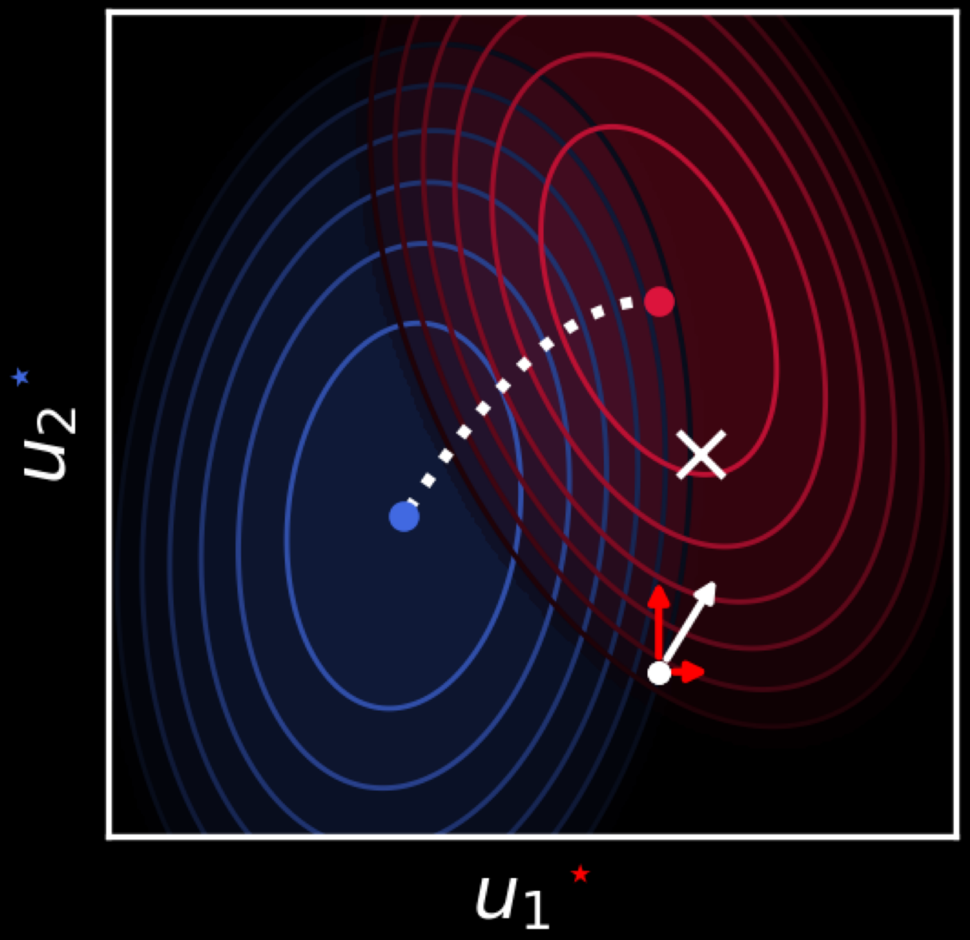
Non-cooperative

$$\min_{u_1} \; c_1(u) \text{ and } \min_{u_2} \; c_2(u)$$



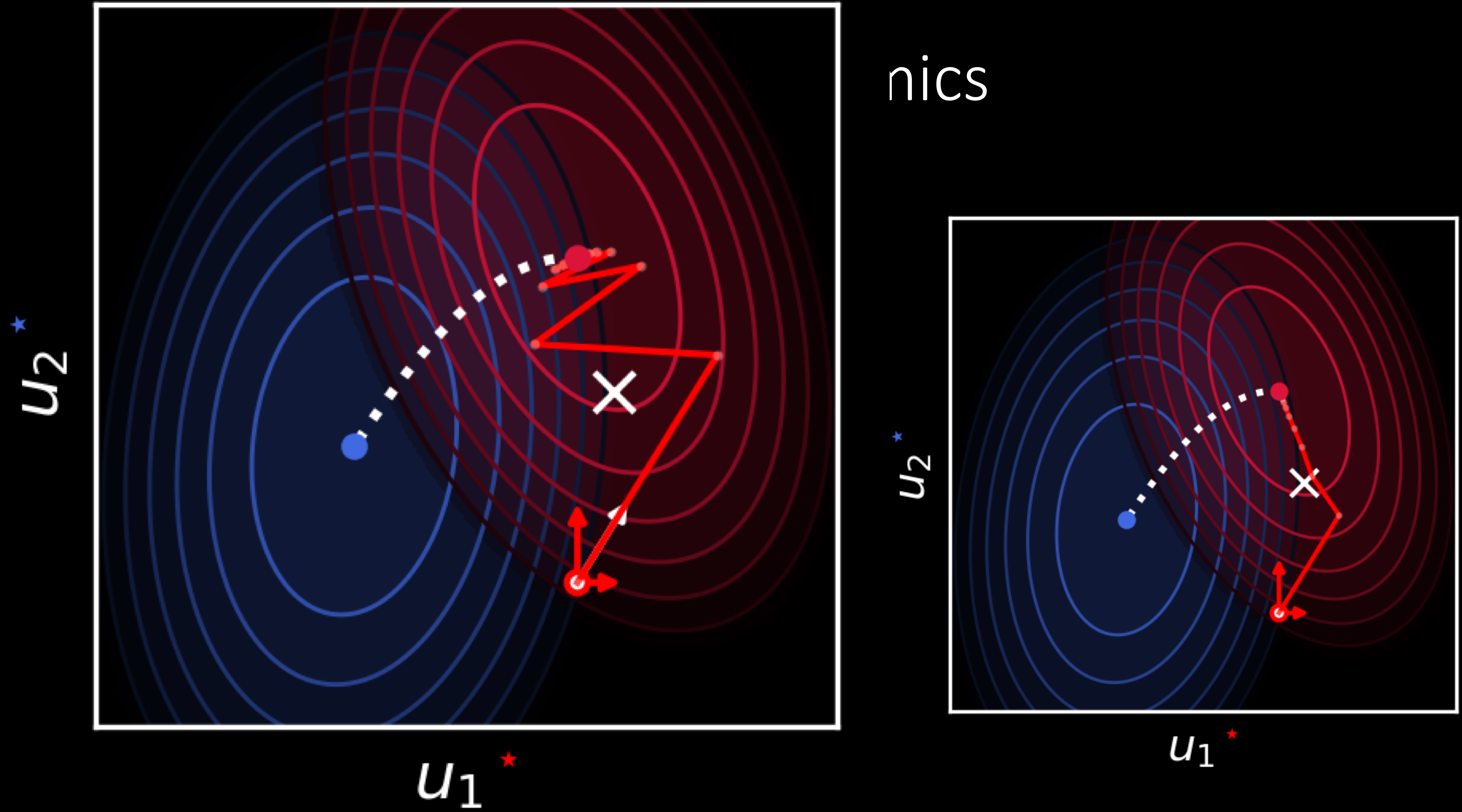
Gradient dynamics

$$u^+ = u - \gamma \begin{bmatrix} D_1 {\color{red}c_1}(u) \\ D_2 {\color{red}c_1}(u) \end{bmatrix}$$



$$D_j c_i(u) \equiv \frac{\partial c_i(u)}{\partial u_j} \in \mathbb{R}^{d_j}$$

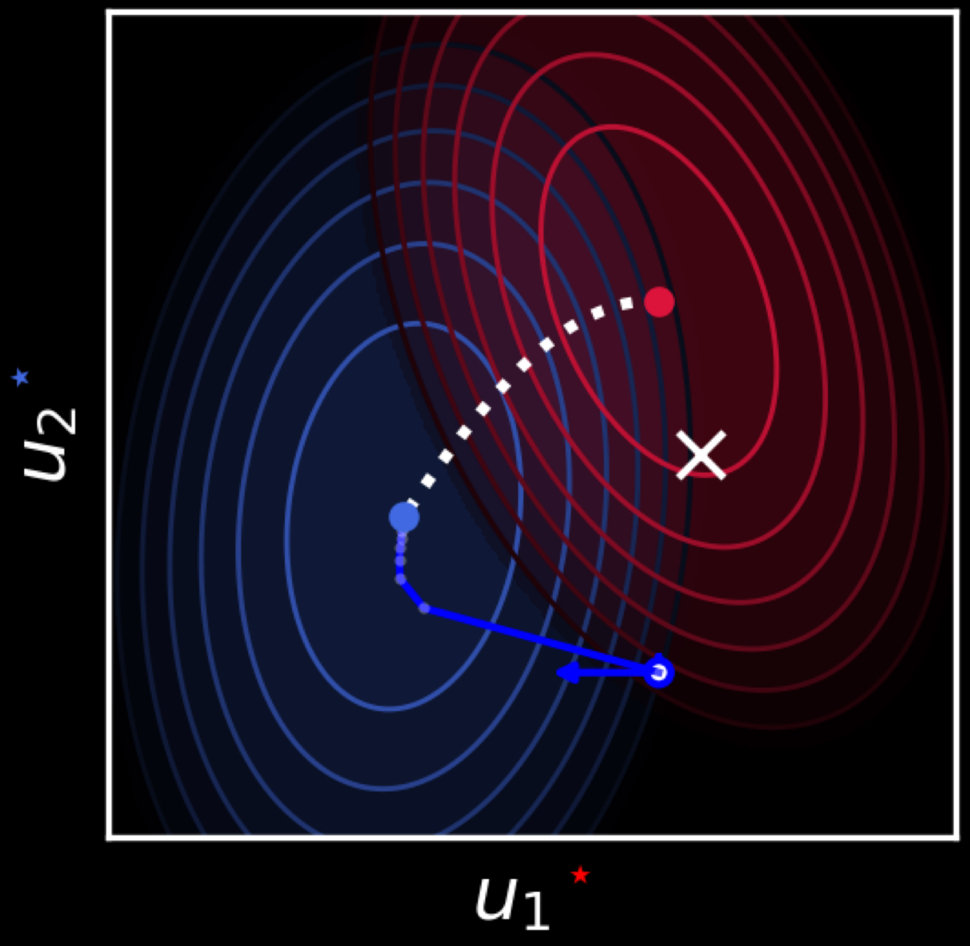
nics



Gradient dynamics

$$u^{+} = u - \gamma \begin{bmatrix} D_1 c_1(u) \\ D_2 c_1(u) \end{bmatrix}$$

$$u^{+} = u - \gamma \begin{bmatrix} D_1 c_2(u) \\ D_2 c_2(u) \end{bmatrix}$$

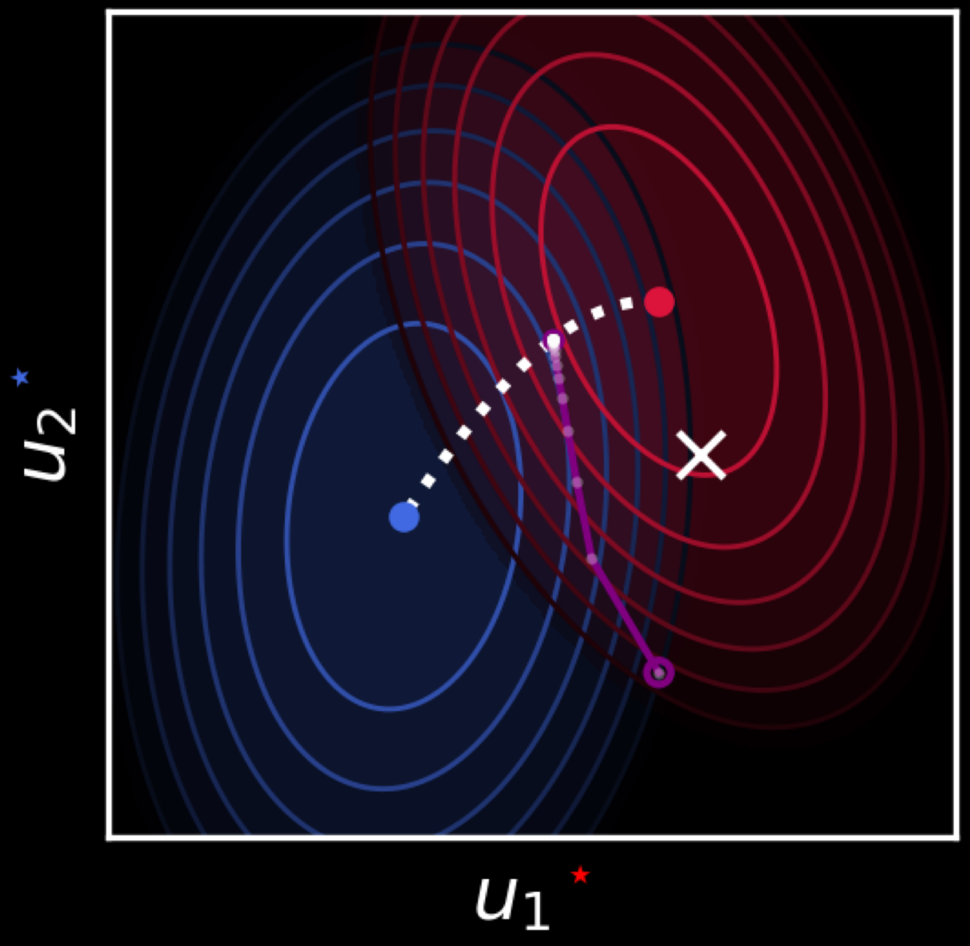


Cooperative dynamics

$$u^+ = u - \gamma \begin{bmatrix} D_1 c_1(u) \\ D_2 c_1(u) \end{bmatrix}$$

$$u^+ = u - \gamma \begin{bmatrix} D_1 c_2(u) \\ D_2 c_2(u) \end{bmatrix}$$

$$u^+ = u - \gamma \theta D c_1(u) + (1-\theta) D c_2(u)$$



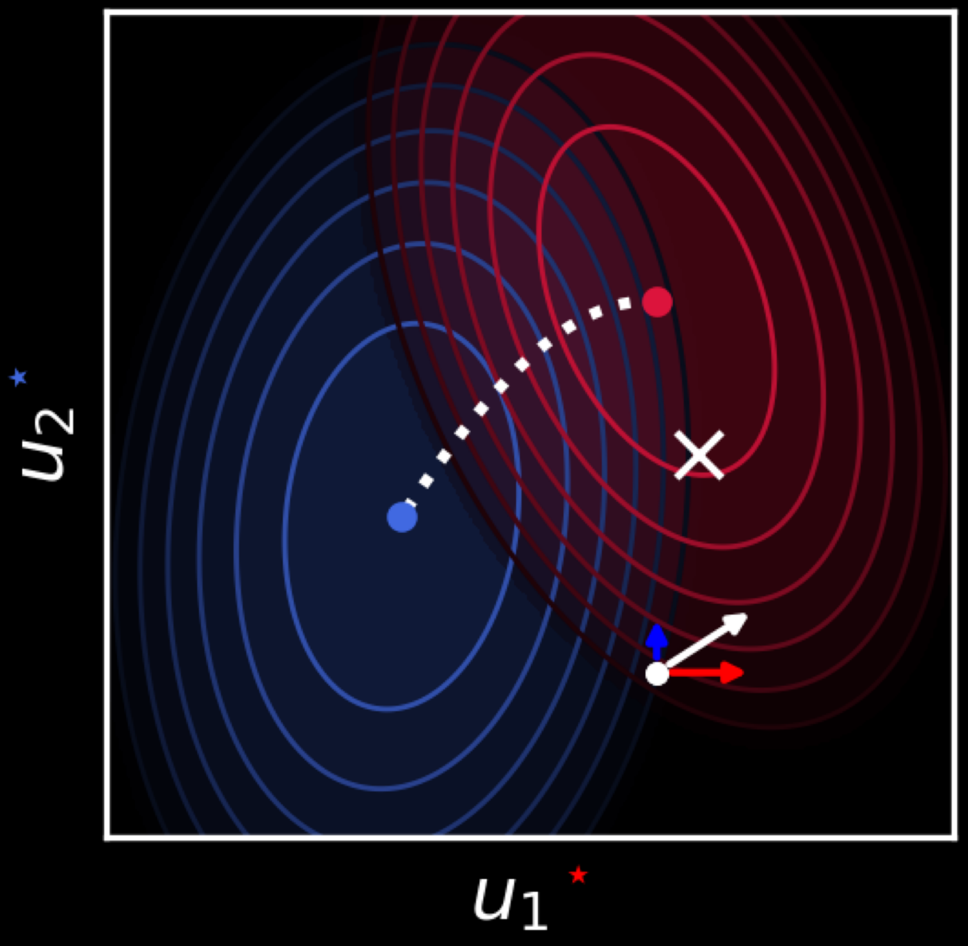
Game vector field

$$u^+ = u - \gamma \begin{bmatrix} D_1 c_1(u) \\ D_2 c_1(u) \end{bmatrix}$$

$$u^+ = u - \gamma \begin{bmatrix} D_1 c_2(u) \\ D_2 c_2(u) \end{bmatrix}$$

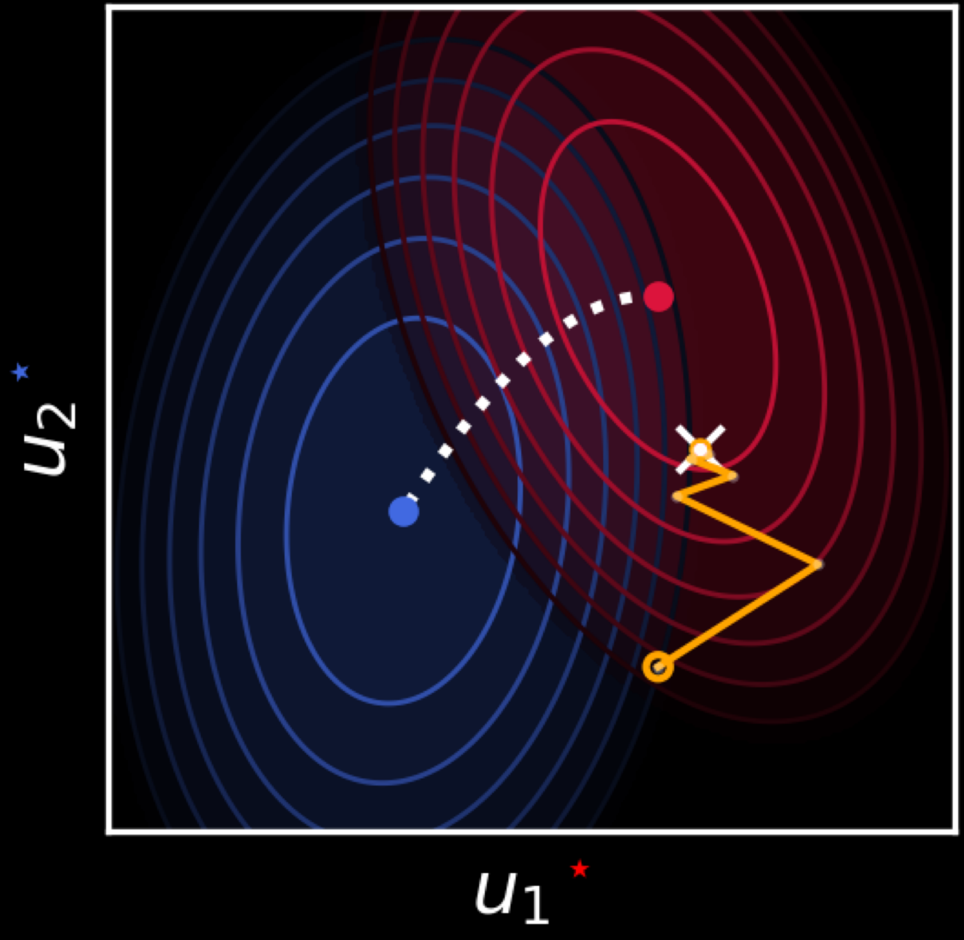
$$u^+ = u - \gamma \theta D c_1(u) + (1-\theta) D c_2(u)$$

$$u^+ = u - \gamma \begin{bmatrix} D_1 c_1(u) \\ D_2 c_2(u) \end{bmatrix}$$



Non-cooperative perspective

$$u^+ = u - \gamma \begin{bmatrix} D_1 c_1(u) \\ D_2 c_2(u) \end{bmatrix}$$



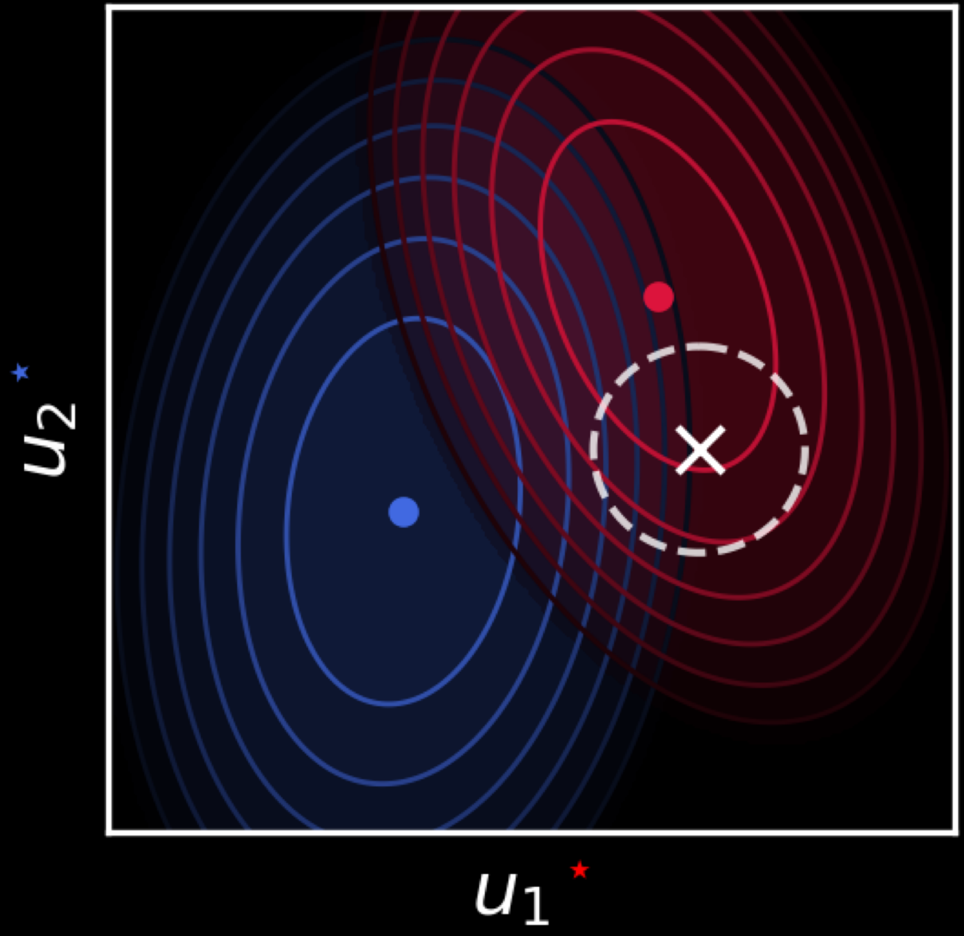
Definition: differential Nash equilibrium

First order conditions

$$D_1 c_1(u^*) = 0, \;\; D_2 c_2(u^*) = 0$$

Second order conditions

$$D_{11} c_1(u^*) > 0, \;\; D_{22} c_2(u^*) > 0$$

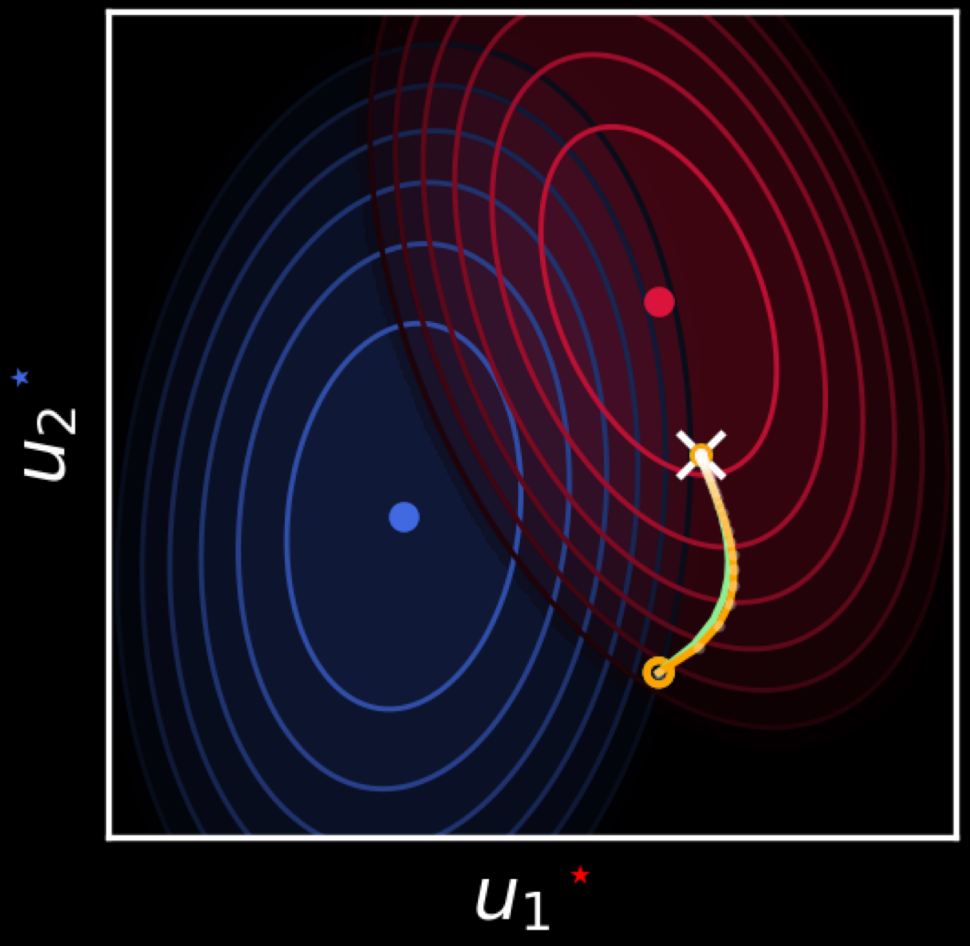


Part I: Learning dynamics in games

$$u^+ = u - \gamma \begin{bmatrix} D_1 c_1(u) \\ D_2 c_2(u) \end{bmatrix}$$

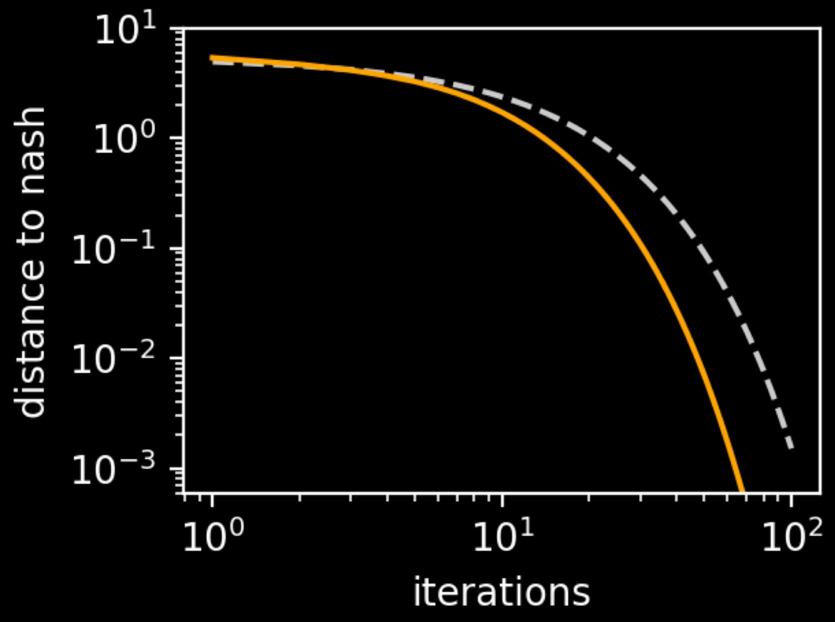
$\Updownarrow$ (with appropriate γ)

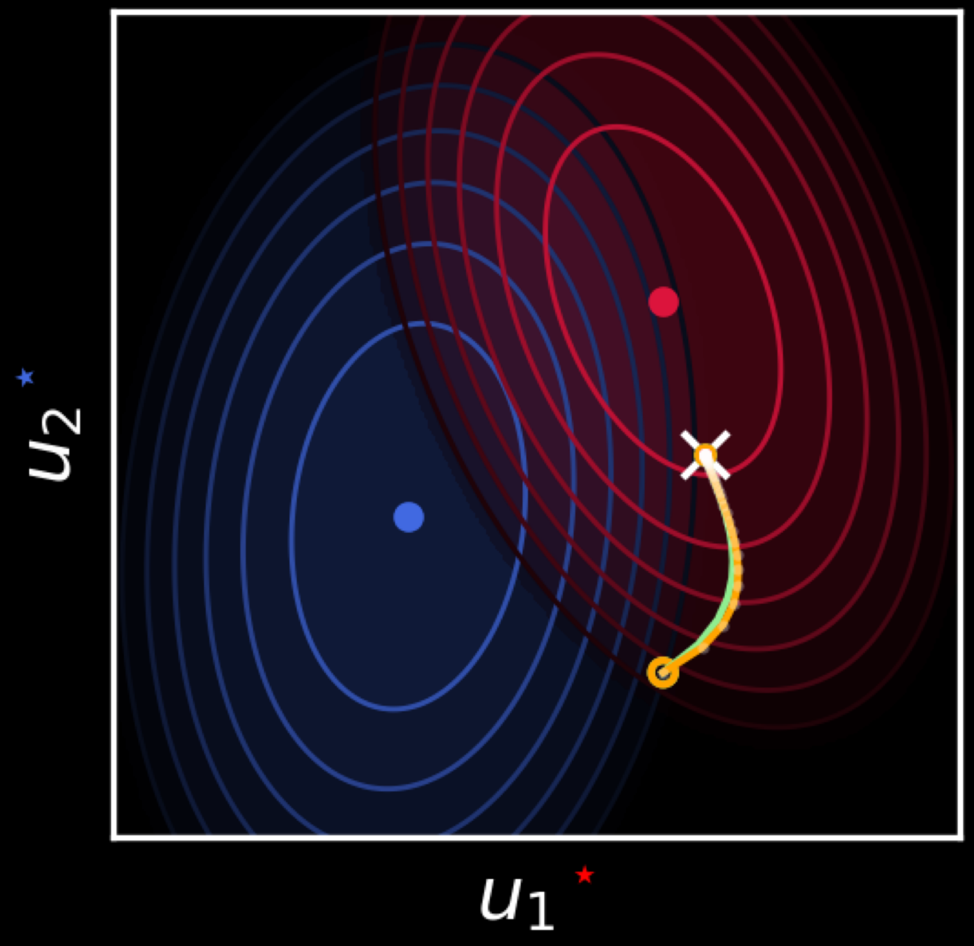
$$\dot{u} = -\omega(u)$$



Non-asymptotic convergence guarantees

$$u^+ = u - \gamma \begin{bmatrix} D_1 c_1(u) \\ D_2 c_2(u) \end{bmatrix}$$





Contraction of learning dynamics

$$u^+ = u - \gamma \begin{bmatrix} D_1 c_1(u) \\ D_2 c_2(u) \end{bmatrix}$$

$$= [I - \gamma J(u)]u$$

Fixed points of vector field *ω(u)*

$$D_1 c_1(u^*) = 0, \ D_2 c_2(u^*) = 0$$

Jacobian of vector field *ω(u)*

$$J = D\omega = \begin{bmatrix} D_{11} c_1 & D_{12} c_1 \\ D_{21} c_2 & D_{22} c_2 \end{bmatrix}$$

Proposition: if $\sup_{\gamma} \|I - \gamma J\| < 1$, then $u(k) \to u^*$

Learning dynamics in games

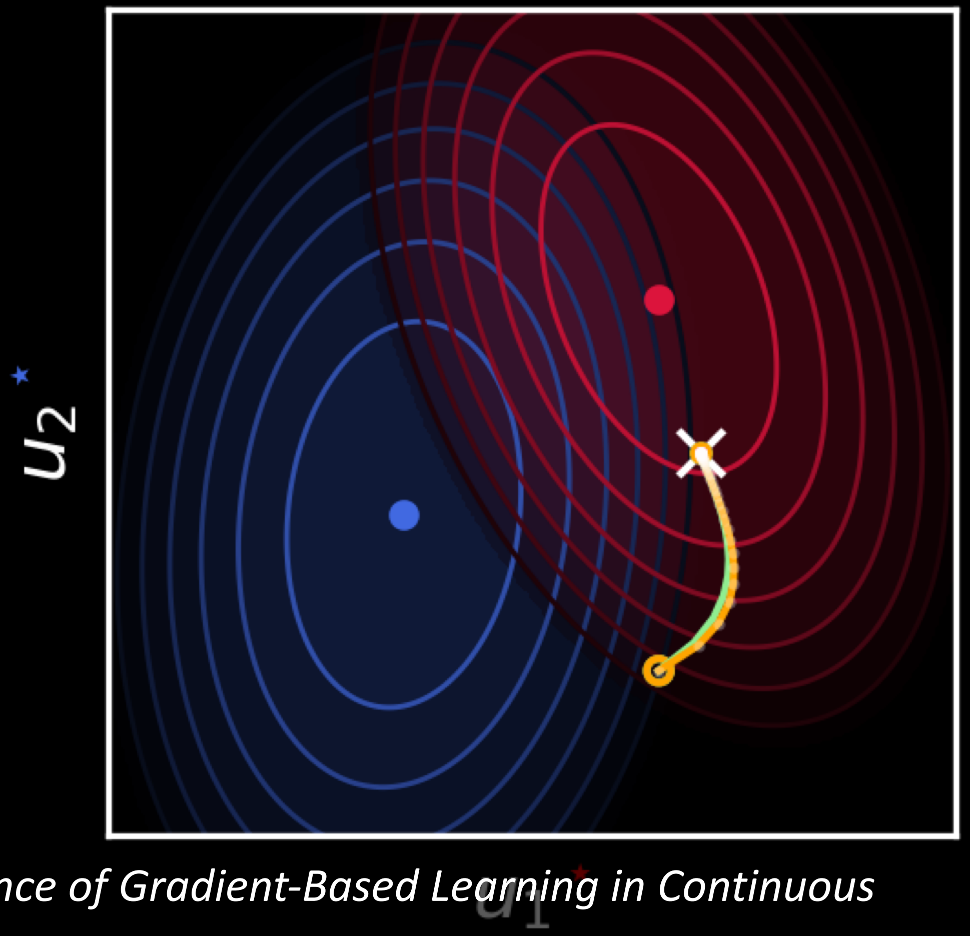
Theorem: With learning rate $\gamma = \alpha/\beta^2$ where singular values $\alpha,\ \beta$ are

$$\alpha = \min_{u \in B_r(u^*)} \sigma_{\min}(J(u) + J(u)^T)/2$$

$$\beta = \max_{u \in B_r(u^*)} \sigma_{\max} J(u)$$

and $u^{(1)}$ is initialized in a region of attraction of a local Nash equilibrium, then the iterates $u^{(k)}$ will be bounded by

$$\|u^{(k)} - u^*\| \leq \exp(-\sqrt{\tfrac{\alpha}{2\beta}}k)\|u^{(1)} - u^*\|$$



[1] Chasnov, Ratliff, Calderone, Mazumdar, Burden, *"Finite-Time Convergence of Gradient-Based Learning in Continuous Games."* AAAI Workshop on Reinforcement Learning in Games (2019).

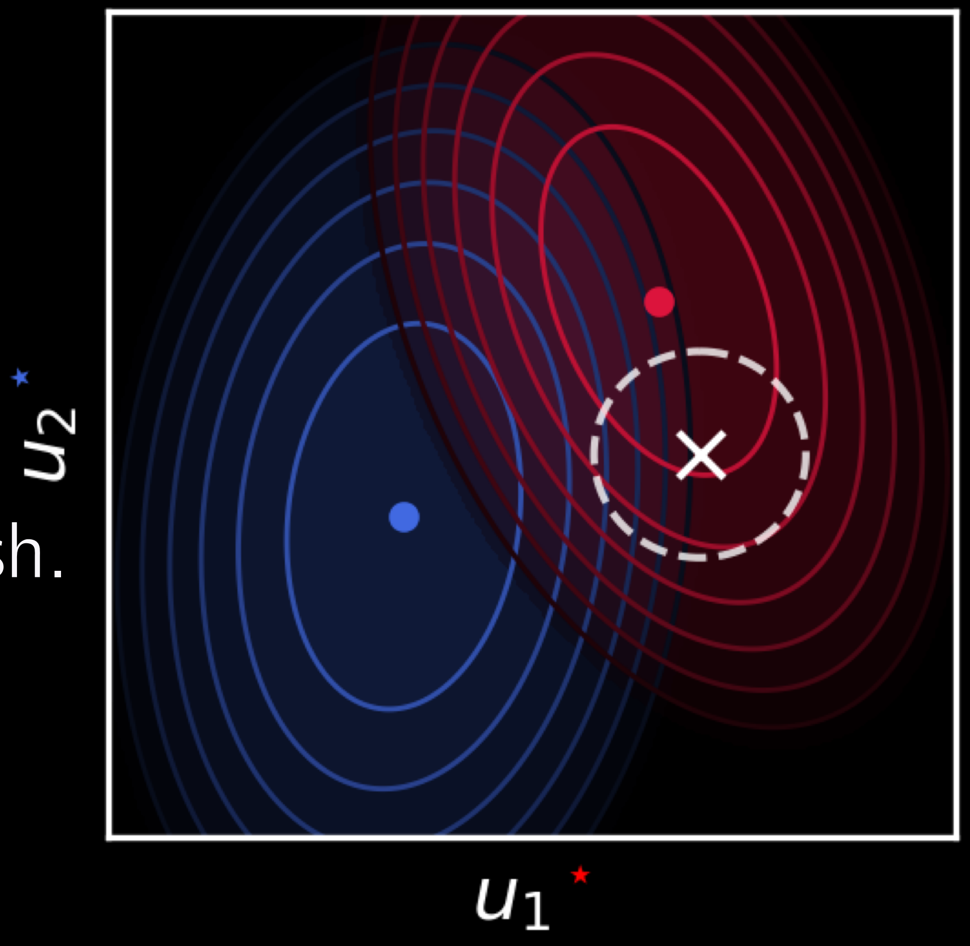
Spectrum of the Jacobian

$$\dot{u} = -\omega(u)$$
$$= -\underbrace{J(u)}_{}\, u$$

If $\mathrm{spec}(J) \subset \mathbb{C}^\circ_+$ at u^*, then u^* is stable.

If $\mathrm{blockdiag}_i(J) > 0$ at u^* $\forall i$, then u^* is Nash.

$$J = D\omega = \begin{bmatrix} D_{11}c_1 & D_{12}c_1 \\ D_{21}c_2 & D_{22}c_2 \end{bmatrix}$$



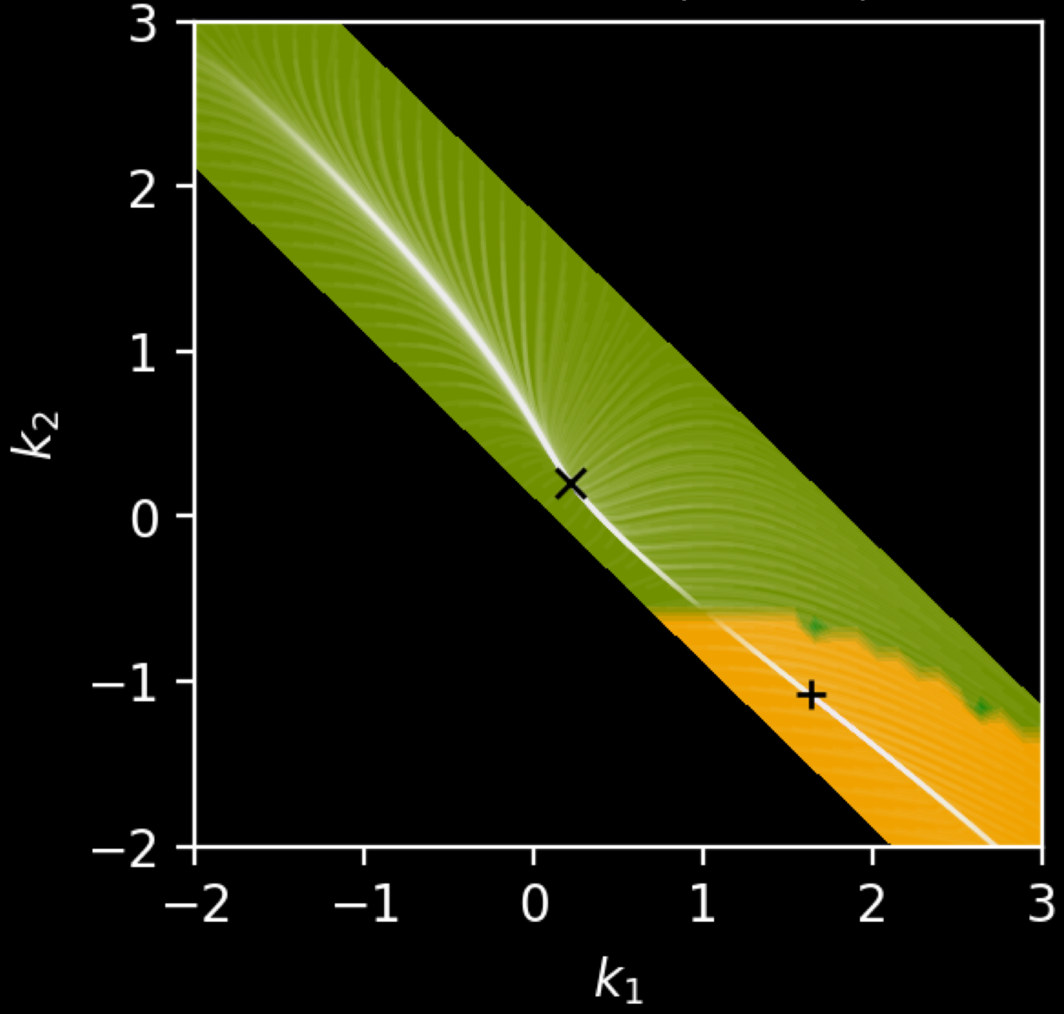
Issue 1: not all stable equilibria are Nash

$$\mathrm{spec}(J) \subset \mathbb{C}_+^\circ$$

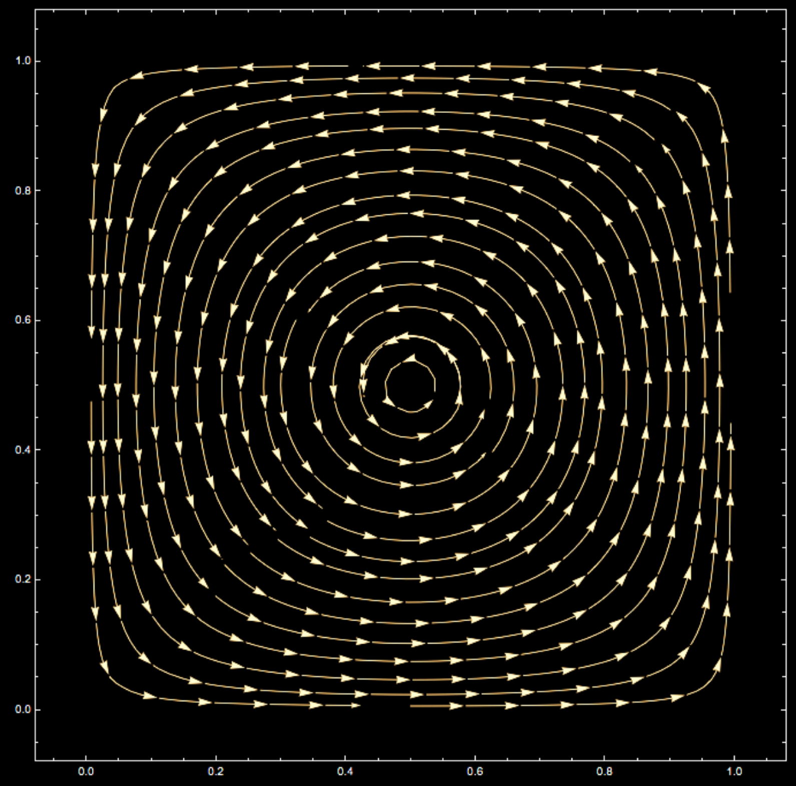
$$\underset{\text{Nash}}{J(u^*)} = \begin{bmatrix} + & \\ & + \end{bmatrix}$$

$$\underset{\text{Non-Nash}}{J(u^*)} = \begin{bmatrix} + & \\ & - \end{bmatrix}$$

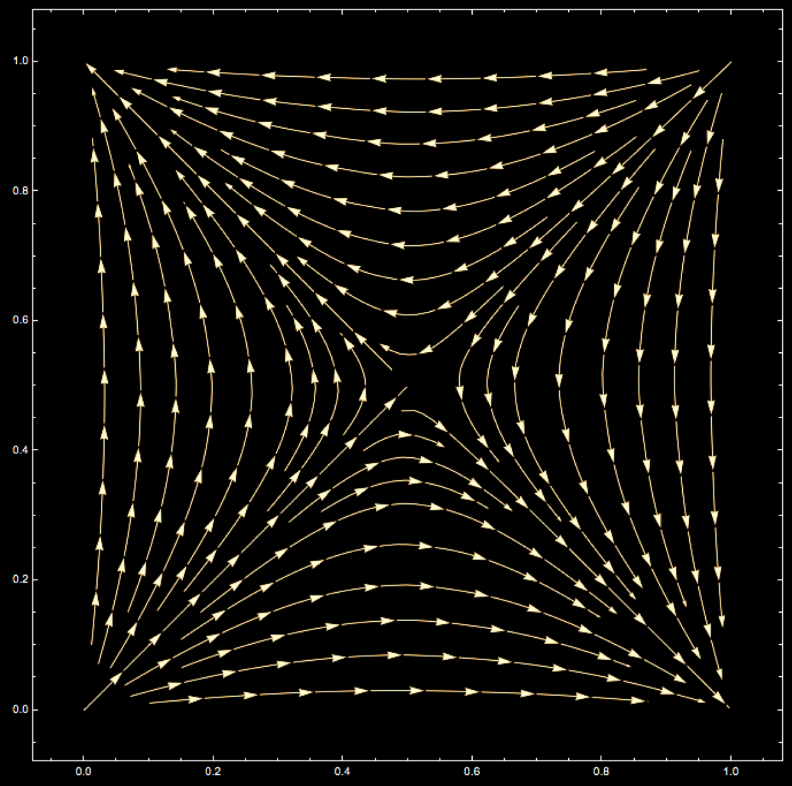
e.g. $U_1 \times U_2 = (k_1, k_2)$



Issue 2: not all Nash equilibria are attractors



Zero-sum game



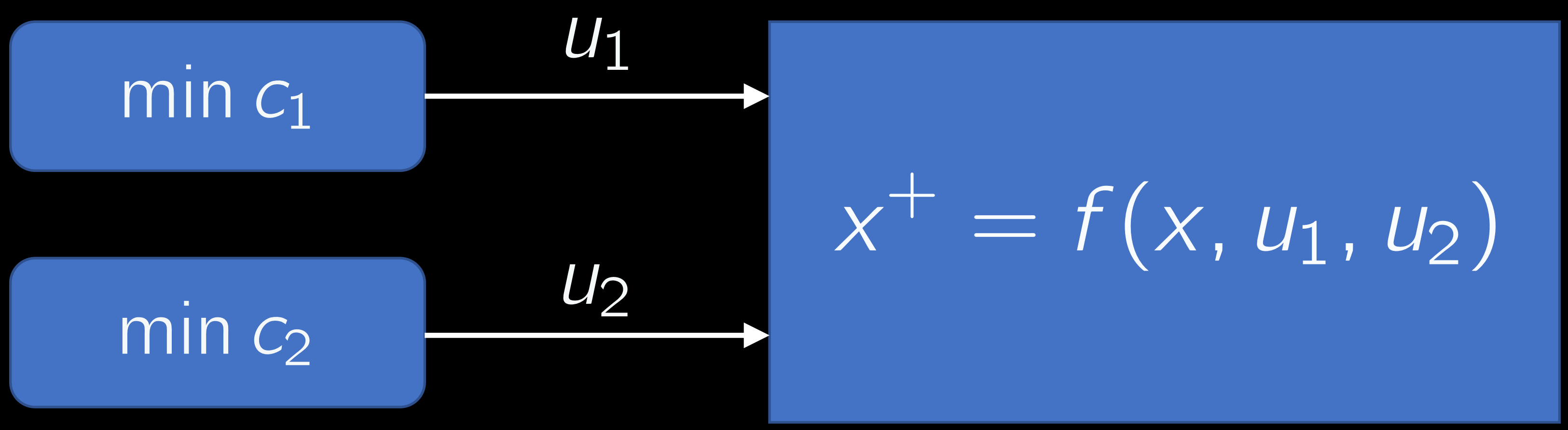
Partnership game

Part II: Towards application in dynamic games

$$x^+ = f(x, u_1, u_2)$$

$$\min_{u_1} \; c_1(x, u), \; \min_{u_2} \; c_2(x, u)$$

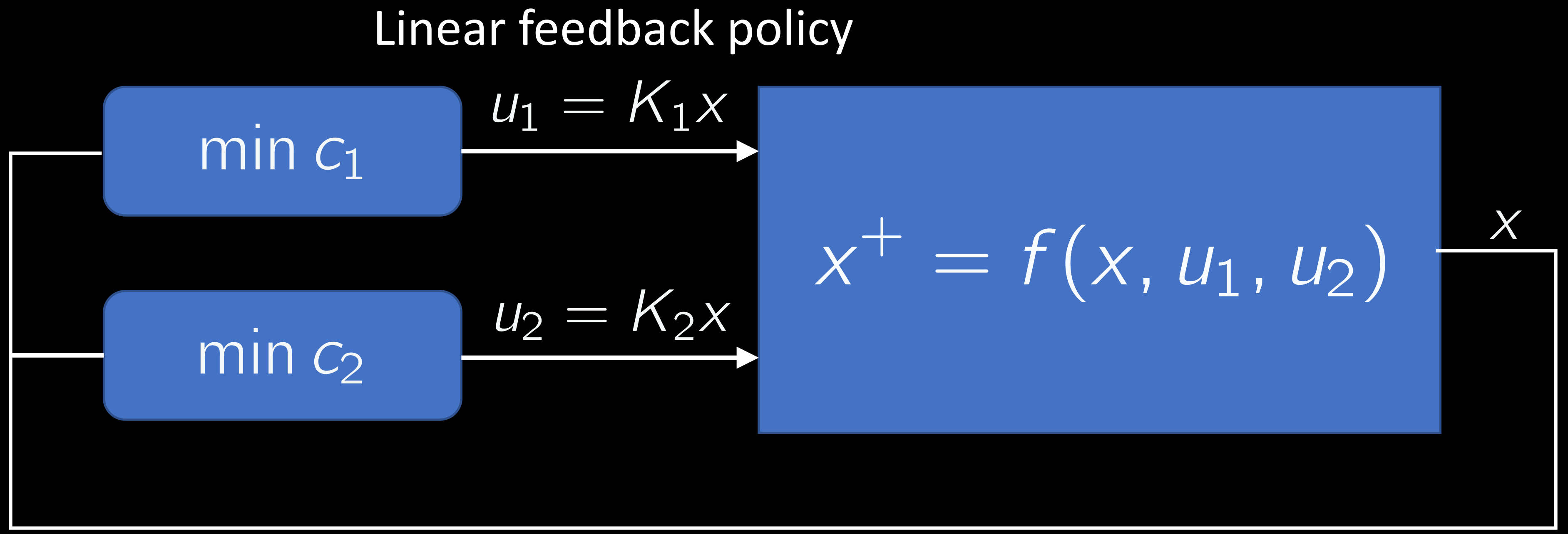
Open loop dynamic games



$$\frac{\partial}{\partial u_1} c_1(x_0, u)$$

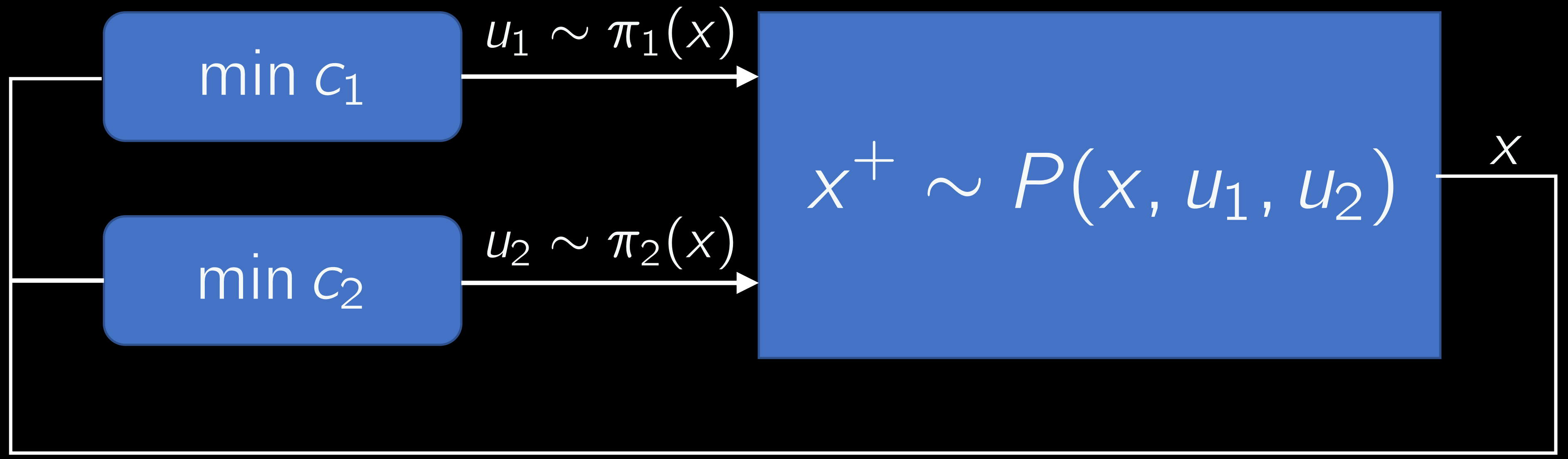
$$\frac{\partial}{\partial u_2} c_2(x_0, u)$$

Closed loop dynamic games



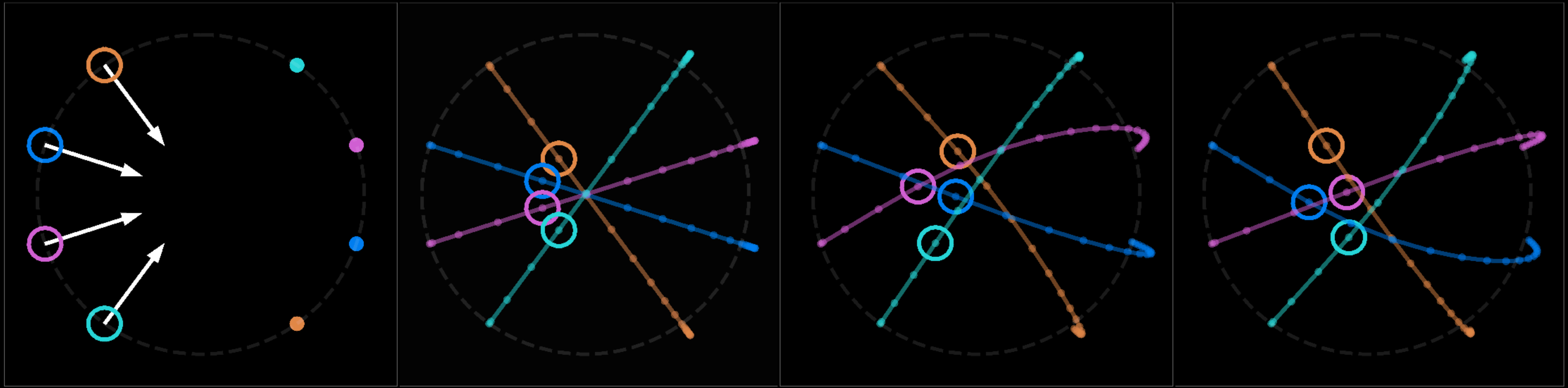
$$\frac{\partial}{\partial K_i} c_i(x, K)$$

Stochastic games



$$\widehat{\frac{\partial}{\partial \theta_i} c_i(\theta)}$$

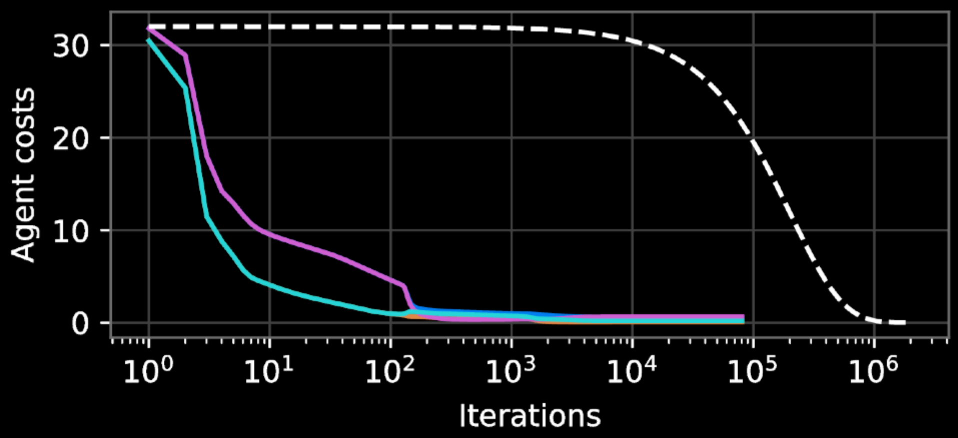
Open loop dynamic game



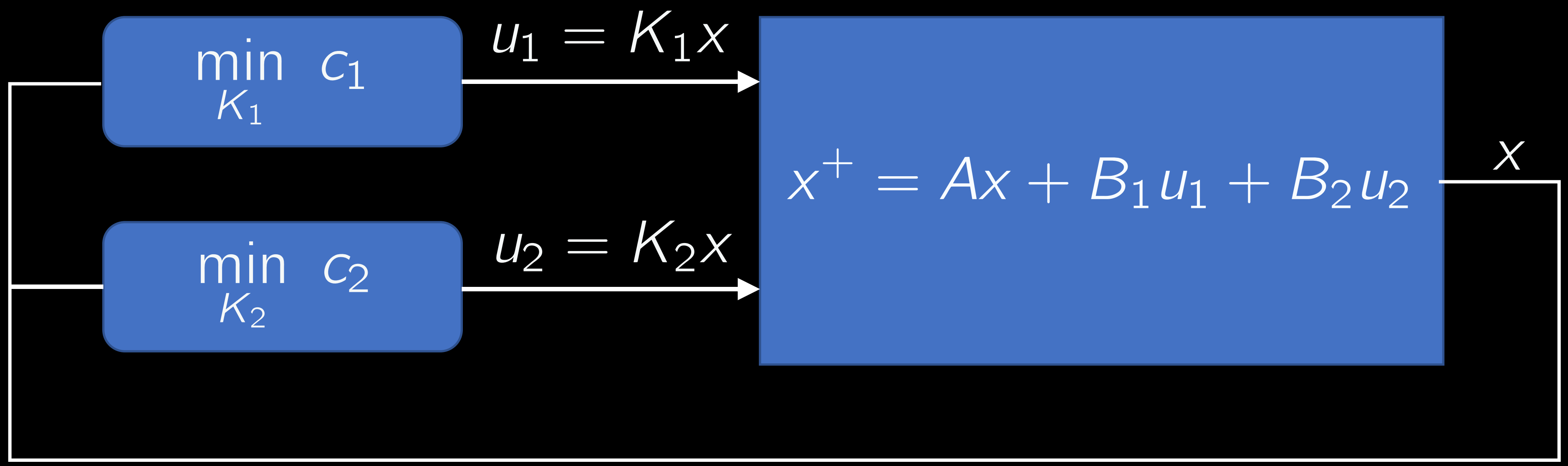
Initialization

Nash equilibrium (1)

Nash equilibrium (2)



Linear Quadratic games (infinite horizon)

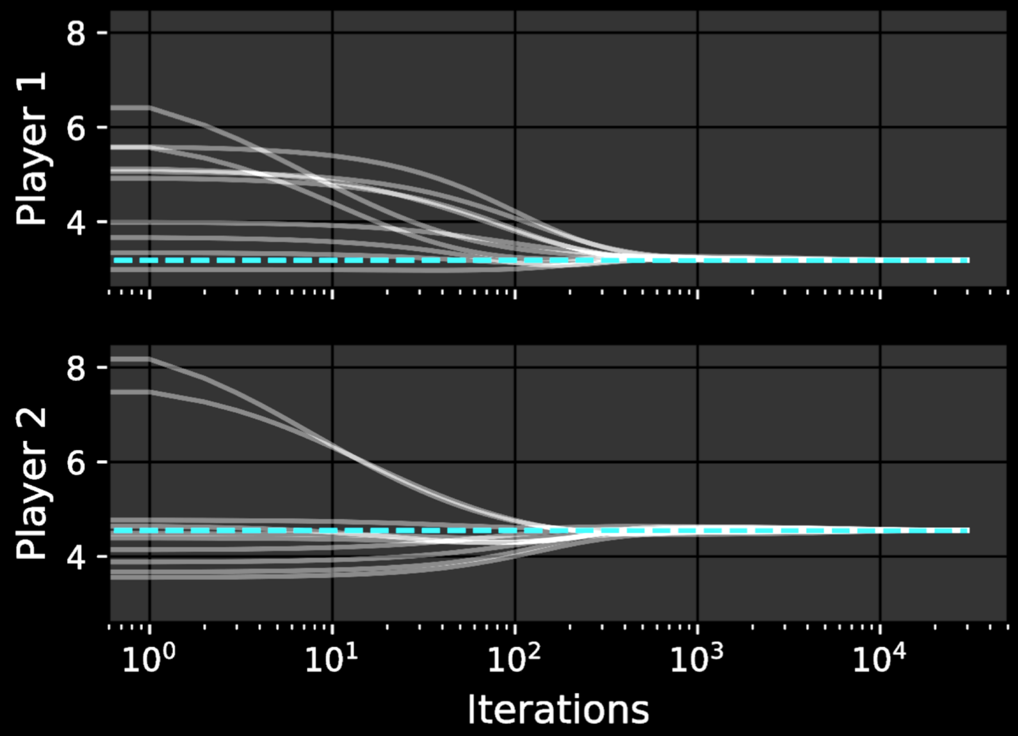


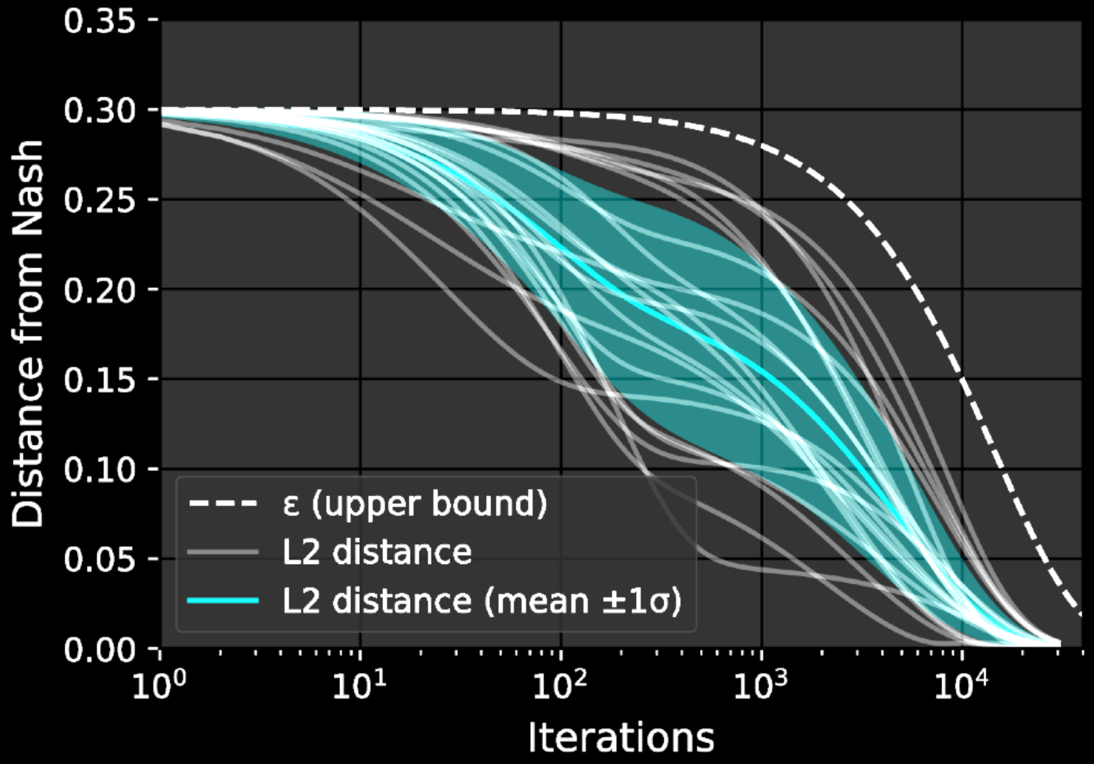
$$c_1(x_0, K_1, K_2) = \sum_{t=0}^{\infty} x^T Q_1 x + u_1^T R_{11} u_1 + u_2^T R_{12} u_2$$

$$c_2(x_0, K_1, K_2) = \sum_{t=0}^{\infty} x^T Q_2 x + u_1^T R_{21} u_1 + u_2^T R_{22} u_2$$

Linear Quadratic game: convergence of gradient method

$$K_1^+ = K_1 - \gamma \nabla_{K_1} c_1(x_0, K_1, K_2)$$
$$K_2^+ = K_2 - \gamma \nabla_{K_2} c_2(x_0, K_1, K_2)$$





Extensions and applications

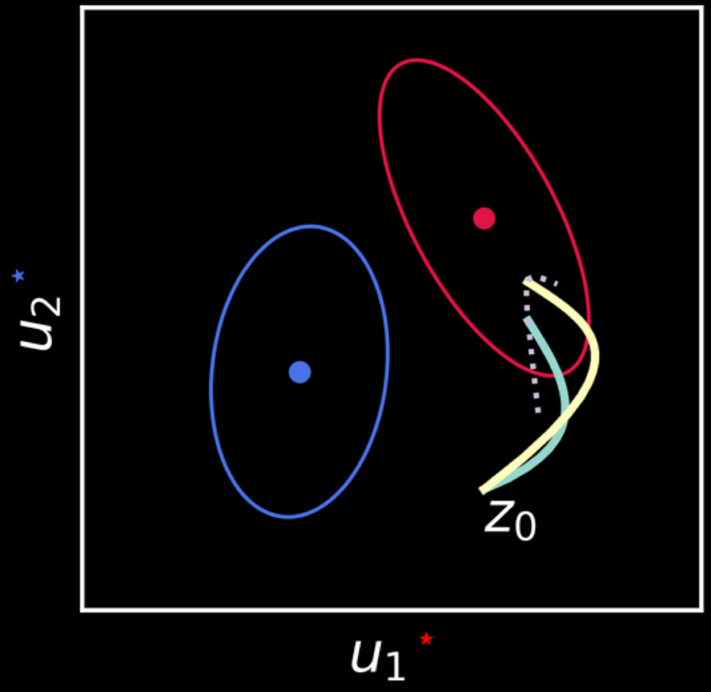
- Stochastic gradients
  - For unbiased estimates, we provide concentration bounds
- Non-uniform learning rates (UAI Mar 2019, in submission)
  - Scaling of agents' learning rates

- Reinforcement learning in games (AAAI Feb 2019 *RL in games* workshop)
- Human-machine sensorimotor games (SPIE Apr 2019)
- Modeling neuron interaction dynamics (NCEC Jan 2019)

Future extensions

- Constrained action space
  - projected descent
- Strategic learning for faster convergence
  - recursive model of agents' learning

- Real world robotic systems
  - dynamically coupled quadcopters
- Human/machine games
  - teleoperation via optimization



Thank you

Timeline

Spectrum of the Jacobian

Proof:

$$
\begin{aligned}
\|I - \gamma J\|_2^2 &= (I - \gamma J)^T (I - \gamma J) \\
&= I - \gamma(J + J^T) + \gamma^2 J^T J
\end{aligned}
$$

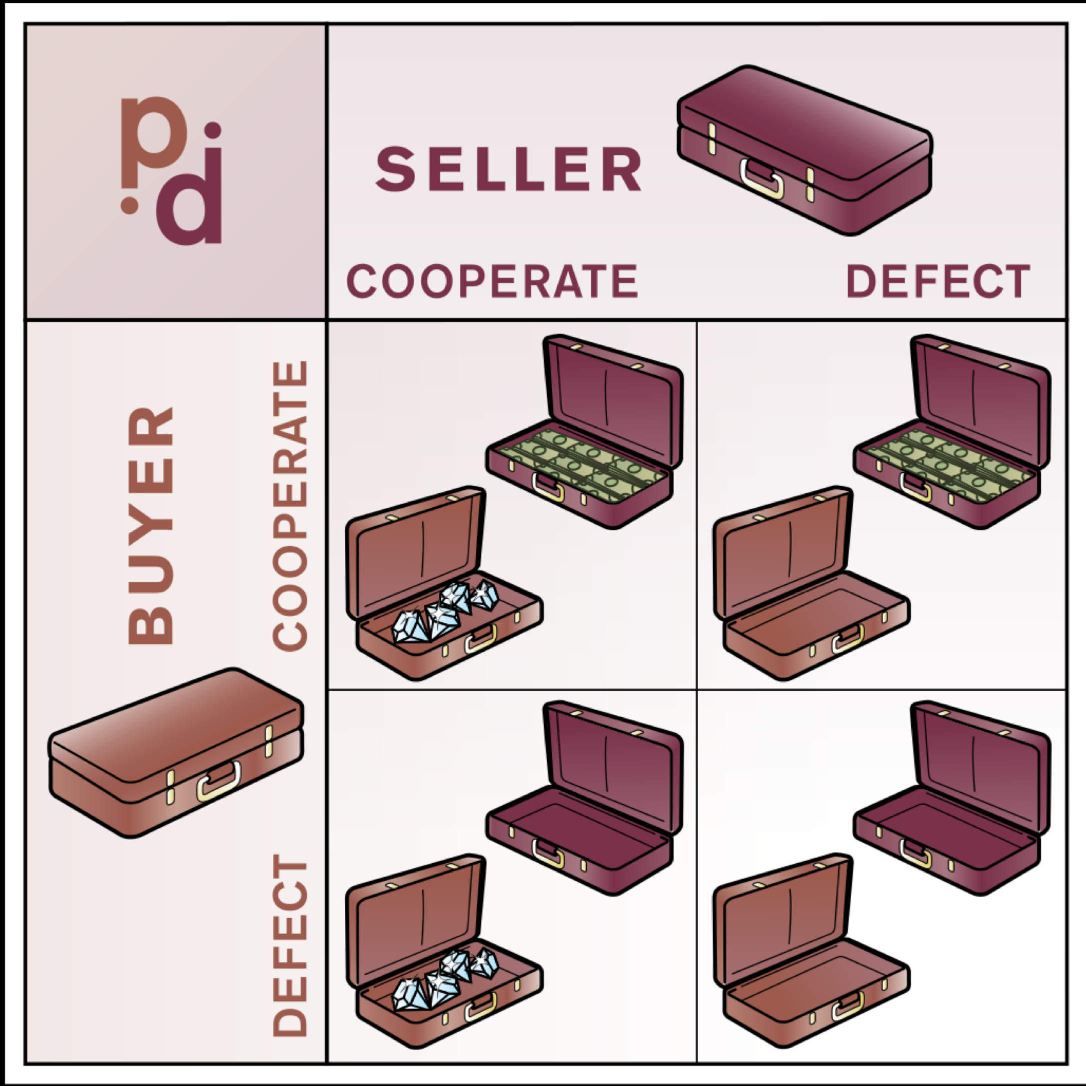
Asymmetric Jacobian

$$J = D\omega = \begin{bmatrix} D_{11}c_1 & D_{12}c_1 \\ D_{21}c_2 & D_{22}c_2 \end{bmatrix}$$

$$J = S + A,\ A \neq 0$$

$$D_{12}c_1 \neq D_{21}c_2^T$$

Prisoner’s dilemma



Local convergence analysis: gradient-play vs. gradient descent

Gradient-play

$$x_1^+ = x_1 - \gamma D_1 f_1(x_1, x_2)$$
$$x_2^+ = x_2 - \gamma D_2 f_2(x_1, x_2)$$

Main theorem (informal):

$$\alpha = \min_{x \in B_r(x)} \sigma_{\min} \overbrace{(D\omega(x)^\top + D\omega(x))/2}^{\text{symmetric part of } D\omega}$$

$$\beta = \max_{x \in B_r(x)} \sigma_{\max} \left(D\omega(x)\right)$$

With learning rate $\gamma = \alpha/\beta^2$ ….

$$\boxed{\|x^{(T)} - x^\star\| \leq \exp\left(-\frac{\alpha^2}{2\beta^2} T\right) \|x^{(1)} - x^\star\|}$$

Gradient descent

$$x^+ = x - \gamma Df(x)$$

Classical result:

μ-strongly convex and L-smooth

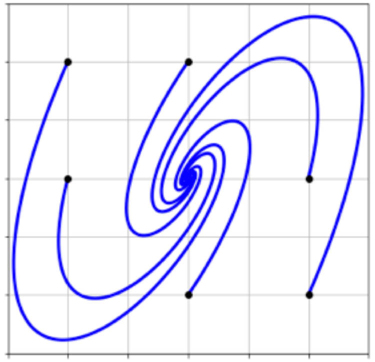
$$\mu \preceq D^2 f(x) \preceq L\,.$$

With learning rate $\gamma = 1/L$
$x^{(T)}$ approaches $x^\star$ in T iterations:

$$\boxed{\|x^{(T)} - x^\star\| \leq \exp\left(-\frac{\mu}{L} T\right) \|x^{(1)} - x^\star\|}$$

L.J. Ratliff, B. Chasnov, D. Calderone, E. Mazumdar, S. Burden. Convergence Guarantees for Gradient-Based Learning in Continuous Games, under review 2018.

Non-Nash stable equilibria: saddle point



$$\dot{x} = -\omega(x)$$

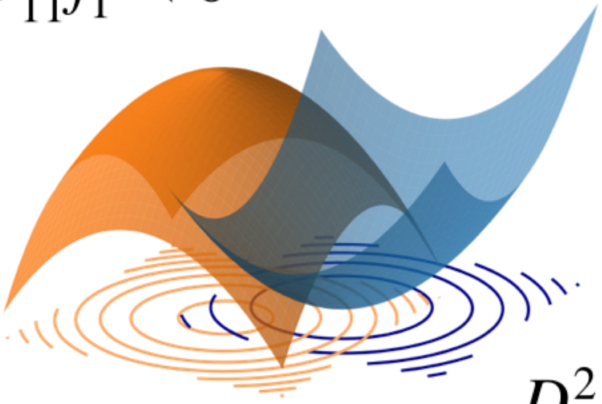
$$D\omega = \begin{bmatrix} - & \\ & + \end{bmatrix}, \quad \text{spec}(D\omega) \subset \mathbb{C}_+^\circ$$

Example:

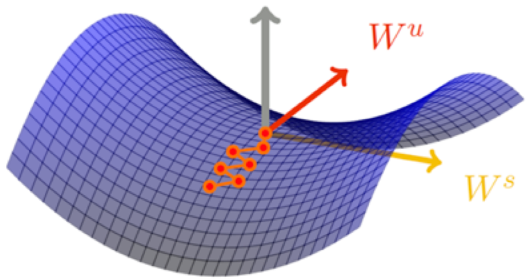
$$f_1(x_1, x_2) = -x_1^2 + 4x_1x_2$$
$$f_2(x_1, x_2) = 6x_2^2 - 8x_1x_2$$

$$D\omega = \begin{bmatrix} -2 & 4 \\ -8 & 12 \end{bmatrix} \qquad \text{spec}(D\omega) = \{2 \pm 4i\}$$

Agent 1 is at a maximum! $D_{11}^2 f_1 \prec 0$



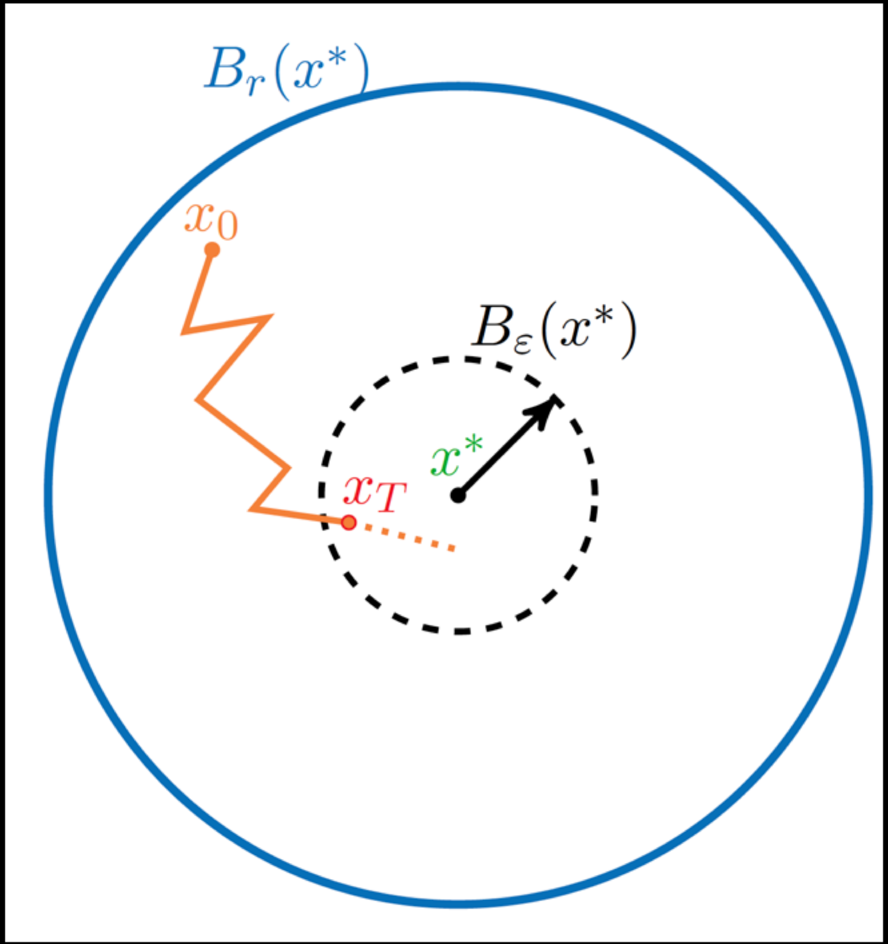
$D_{22}^2 f_2 \succ 0$



Theorem: (x^*: stable differential Nash)
suppose $x_0 \in B_r(x^*)$, ω is Lipschitz, and $\gamma_i = \sqrt{\alpha}/(k\beta)$ for each $i \in [n]$ with $\alpha < k\beta$. Gradient based learning obtains an ε-differential Nash in finite time $T \geq \lceil 2k\frac{\beta}{\alpha}\log(r/\varepsilon)\rceil$

$$\alpha = \min_{x \in B_r(x)} \sigma^2_{\min}(\underbrace{D\omega(x) + D\omega(x)^T}_{\text{symmetric part of } D\omega}),$$

$$\beta = \max_{x \in B_r(x)} \sigma^2_{\max}(D\omega(x))$$



Conclusion

References

Papers

- AAAI 2019 oral presentation
- SPIE 2019
- UAI 2019

Posters and presentations

- AMP fellow
- NCEC

Notation (two players)

- Partial derivatives

$$D_j c_i(u) \equiv \frac{\partial c_i(u)}{\partial u_j} \in \mathbb{R}^{d_j}$$

$$D_{jk} c_i(u) \equiv \frac{\partial^2 c_i(u)}{\partial u_j \partial u_k} \in \mathbb{R}^{d_j} \times \mathbb{R}^{d_k}$$

- Remarks

$$D_{jj} c_i(u)$$

True multi-agent interactions (i.e. society, evolution) has multiple decision-makers with multiple objectives.

- Natural formulation is a non-cooperative game
  - Games with discrete actions (Von Neuman 1944, Nash 1951)
  - Games with MDP-like state transitions (Shapely 1953)
  - Games with linear dynamics and quadratic costs (Basar 1976)

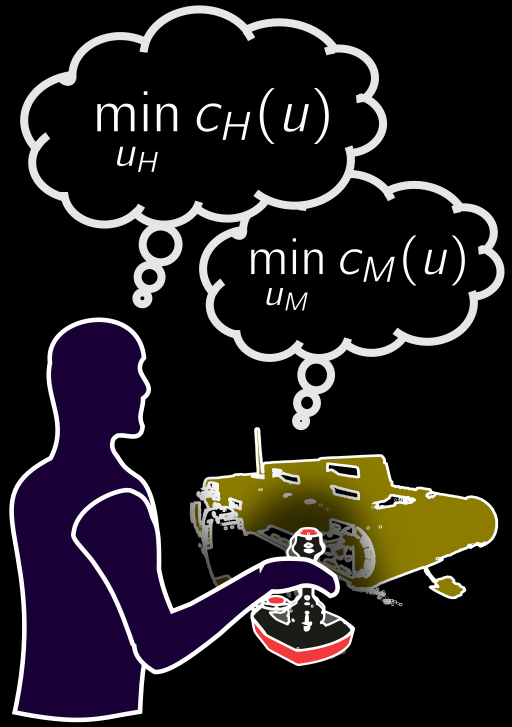
Theorem

[1] Chasnov, Ratliff, Calderone, Mazumdar, Burden, *"Finite-Time Convergence of Gradient-Based Learning in Continuous Games."* AAAI Workshop on Reinforcement Learning in Games (2019).
Workshop paper and 20 min oral presentation.

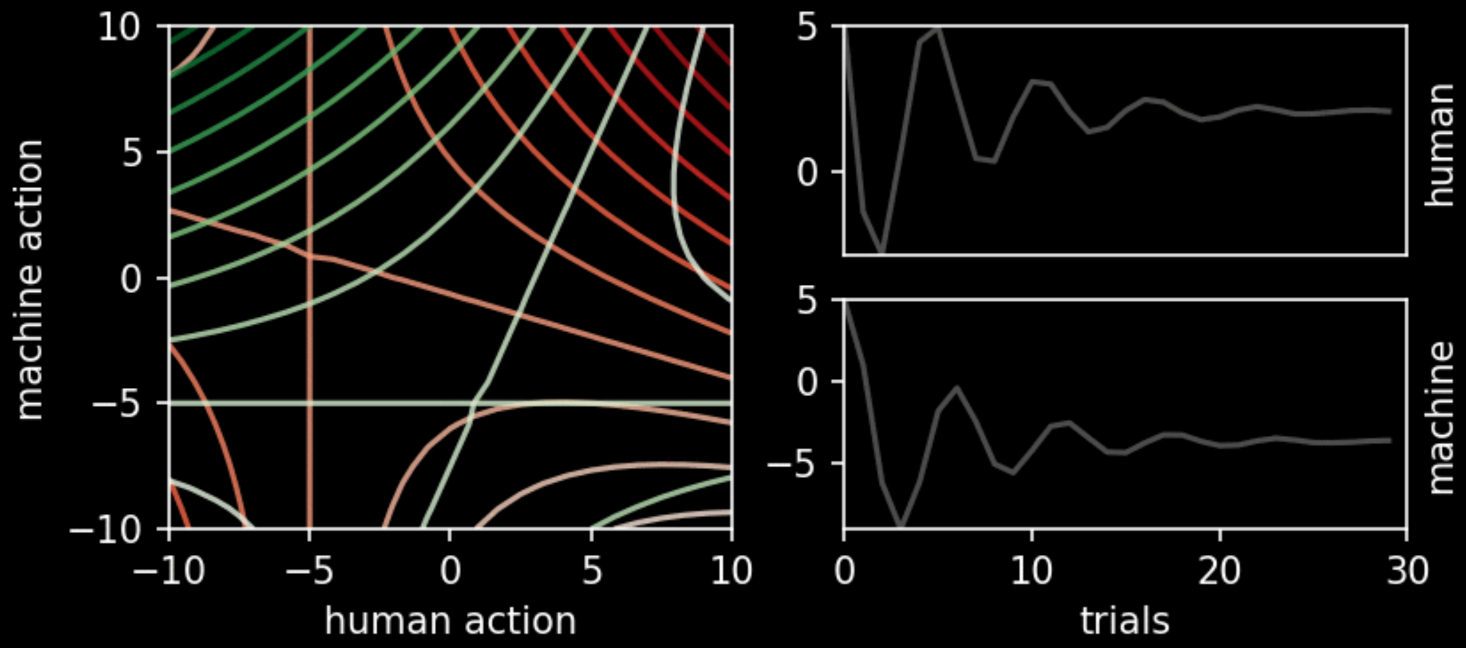
Human-machine sensorimotor games

$$u = (u_H, u_M)$$

$$u_H^+ = u_H - \gamma D_{u_H} c_H(u)$$
$$u_M^+ = u_M - \gamma D_{u_M} c_M(u)$$



Stable attractor:



- Analysis of coupled optimization problems is crucial for developing safe, reliable connected systems

Current paradigm

- A **single decision-maker** (centralized planner)

- Multiple agents carry out actions (distributed agents)
- *Trust & communication* is fully assumed

- \min_u={u_1, \dots, u_n}\ c(u)

Need for understanding

Next frontier

- **Multiple** decision-makers
- Actions carried out affect the decision-making
- Trustless and robust to limited communication
- The decision-making and actions are coupled

“Multi-agent” learning and control under this paradigm is similar to single mind with multiple bodies

- AlphaGo: two player game, but it is playing a clone of itself
- Multi-agent swarms: achieves a single objective with multiple bodies

Natural formulation of the problem is a continuous game

- n agents
- u_i: agent i's action
- $c_i(u)$: agent i's cost, twice continuously-differentiable, maps from joint action $u=(u_1, u_n)$ to R
- Goal: agents at a minimum of its own cost
- Definition: $u^*=(u_1^*, \dots u_n^*)$ differential Nash equilibrium if $D_ic_i(u^*)=0$ and $D_{ii}c_i(u^*) > 0$ for all $i = 1 \dots n$