A gradient-based model for learning in human/machine interaction

Ben Chasnov, PhD Student

Electrical and Computer Engineering

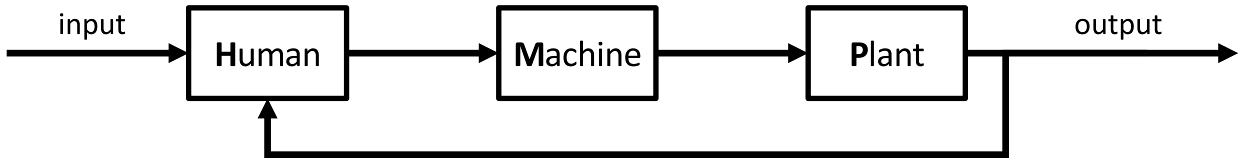
University of Washington, Seattle WA

Advisors: Dr. Sam Burden and Dr. Lillian Ratliff

Collaborators: Behcet Açıkmeşe, Dan Calderone, Amber Chou, Tanner Fiez, Maneeshika Madduri, Eric Mazumdar, Skye Mceowen, Joey Sullivan, Momona Yamagami, Liyuan Zheng

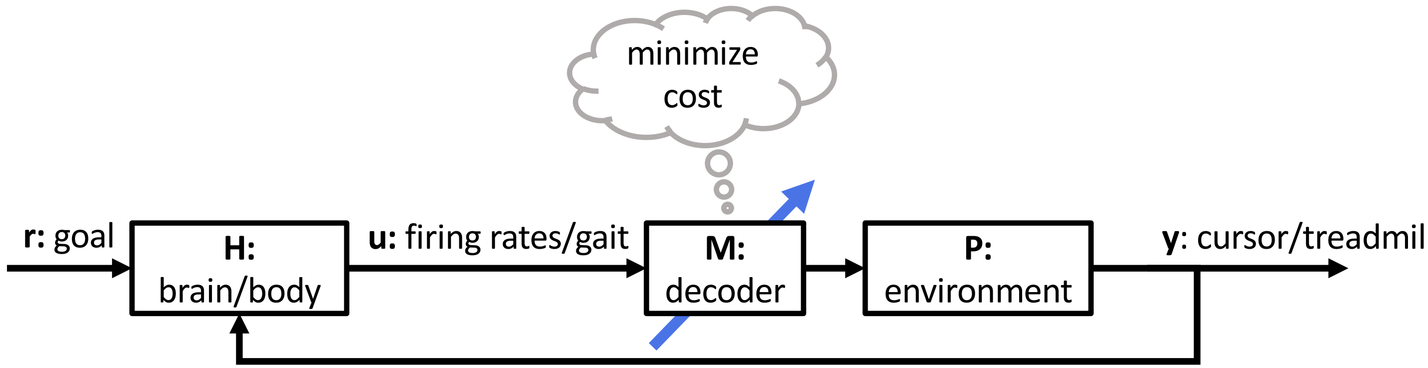
Neural Computation and Engineering Connection (NCEC) 2021

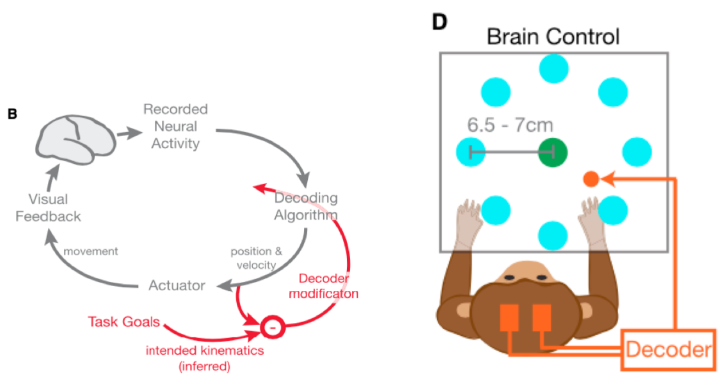
Human/machine interaction is a closed loop



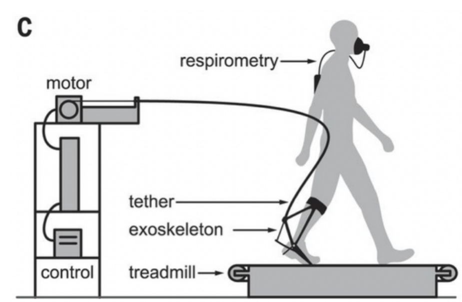
Zhou, Doyle, and Glover. *Robust and optimal control*. New Jersey: Prentice hall, 1996.
Åström and Murray. *Feedback systems.* Princeton university press. 2008.

Intervening in this loop with optimization



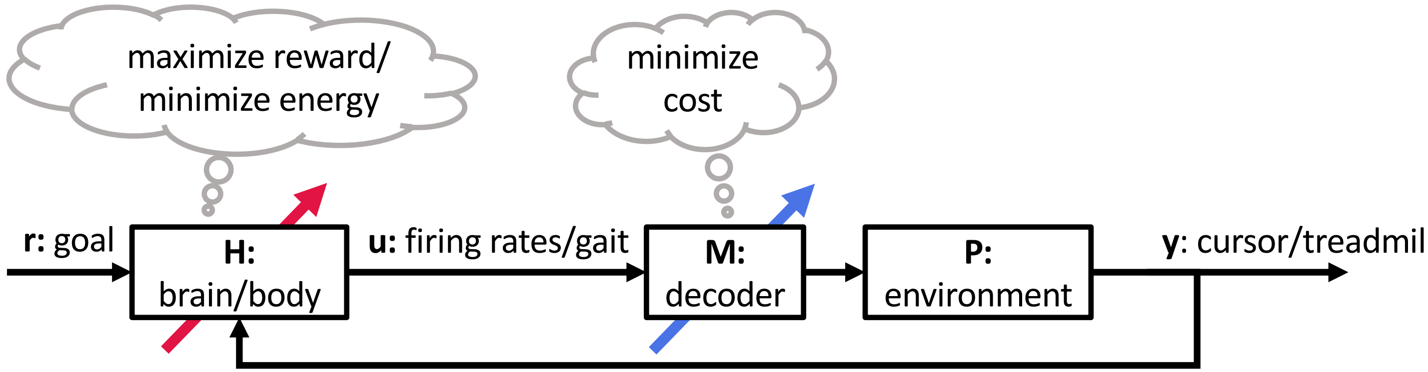


Orsborn, Dangi, Moorman, and Carmena. "Closed-loop decoder adaptation on intermediate time-scales facilitates rapid BMI performance improvements independent of decoder initialization conditions." *IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2012.*

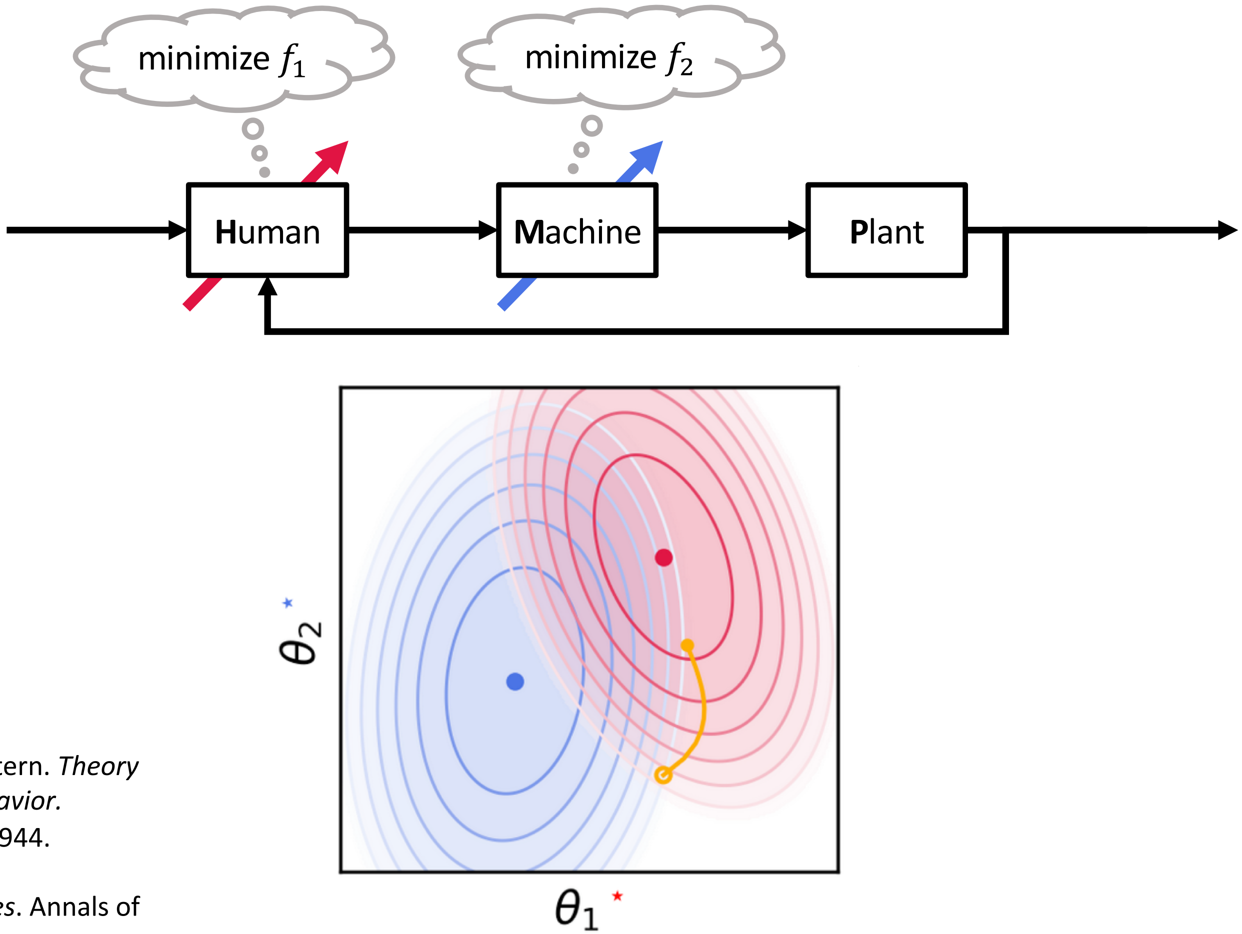


Zhang, Fiers, Witte, Jackson, Poggensee, Atkeson, Collins. *Human-in-the-loop optimization of exoskeleton assistance during walking.* Science. 2017

Humans and brains are constantly adapting too



Human/machine interaction is a *game*



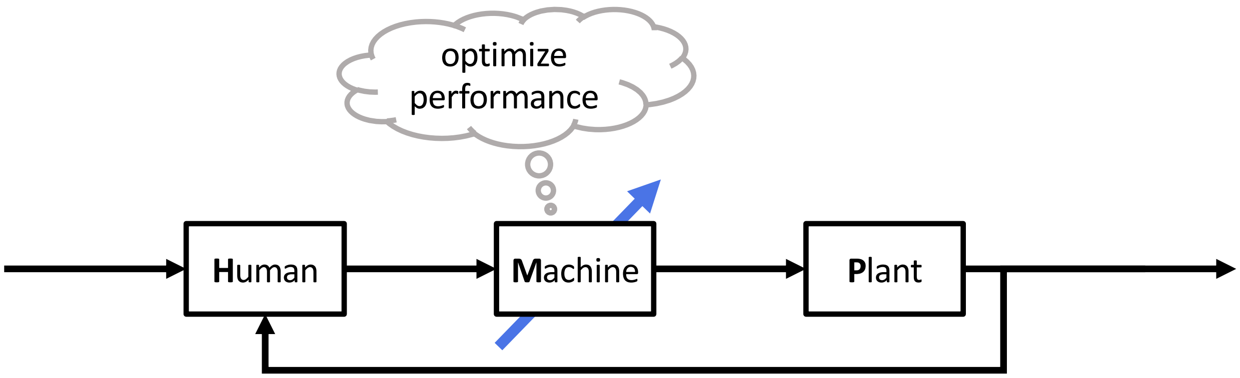
$$\min_{\theta_1} \; f_1(\theta_1, \theta_2), \quad \min_{\theta_2} \; f_2(\theta_1, \theta_2)$$

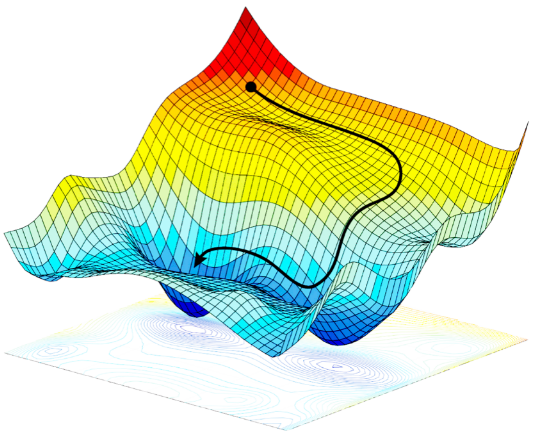
Von Neumann and Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, 1944.

Nash. *Non-cooperative games*. Annals of mathematics, 1951.

Başar and Olsder. *Dynamic noncooperative game theory*. Society for Industrial and Applied Mathematics, 1998.

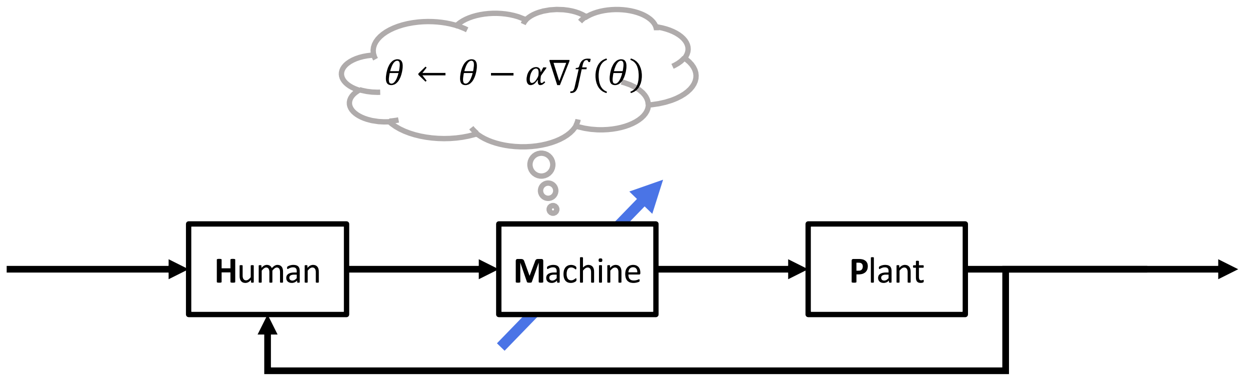
Machines optimize

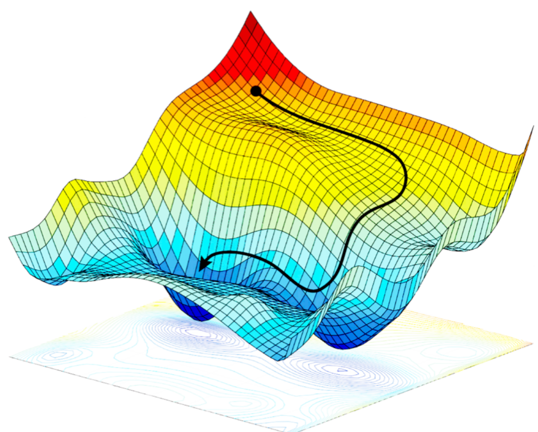




$$\min_{\theta} \; f(\theta)$$

Machines optimize using gradients





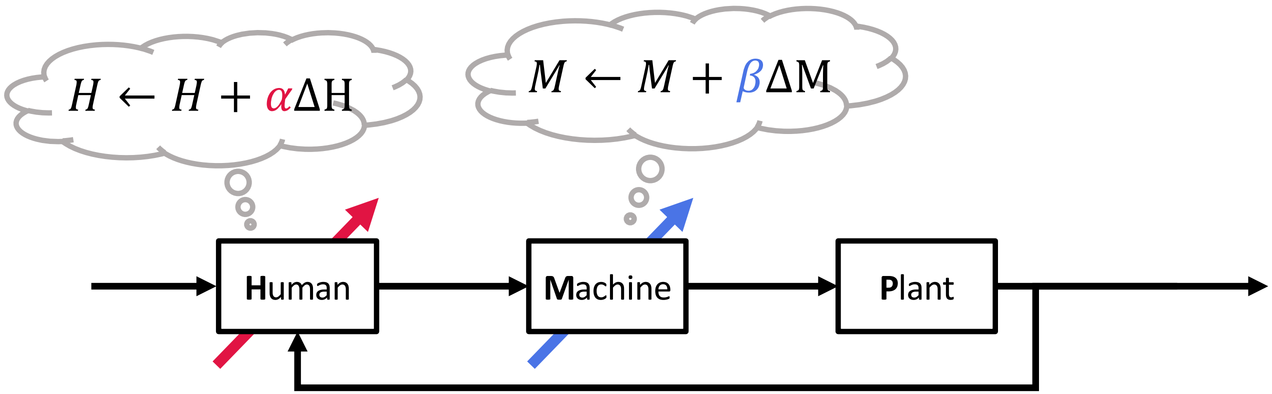
$$\min_{\theta} \; f(\theta)$$

Bertsekas and Tsitsiklis. *Neuro-dynamic programming*. Athena Scientific, 1996.

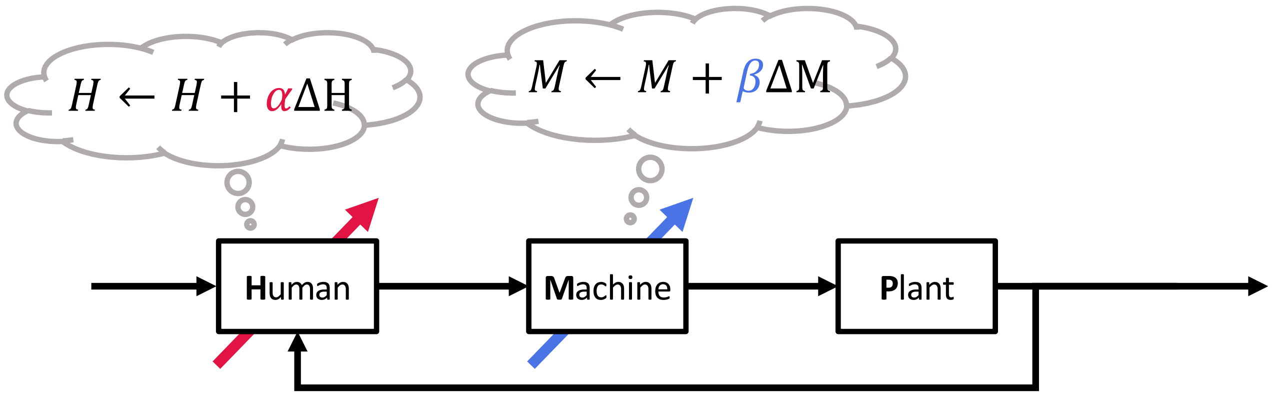
Sutton and Barto. *Introduction to reinforcement learning*. Cambridge: MIT press, 1998.

Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Hypothesis: humans respond analogously



Gradient-based learning dynamics in games



Discrete-time dynamics

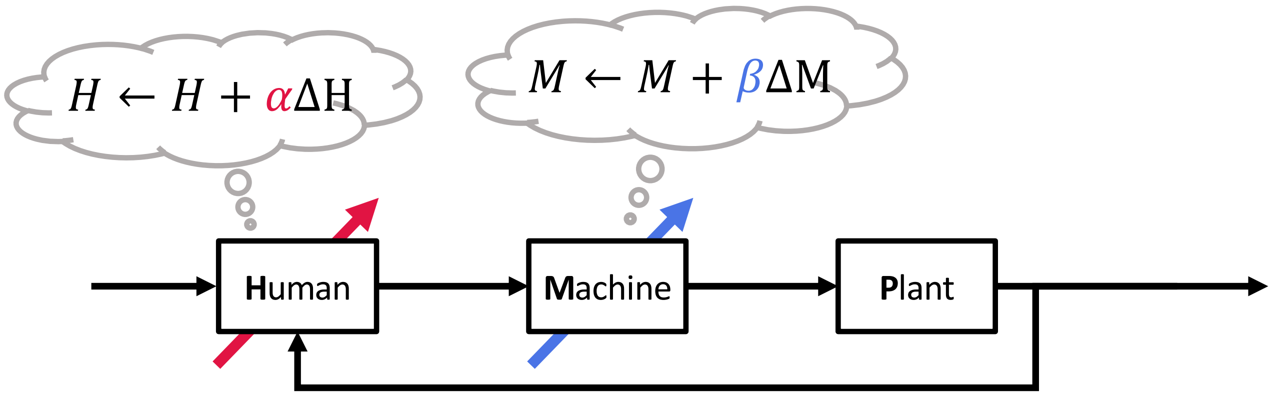
$$H \leftarrow H + \alpha \Delta \mathrm{H}$$
$$M \leftarrow M + \beta \Delta \mathrm{M}$$

Fudenberg and Levine. *The theory of learning in games*. Vol. 2. MIT press, 1998.

Singh, Kearns, and Mansour. "Nash Convergence of Gradient Dynamics in General-Sum Games." In *Uncertainty in Artificial Intelligence*, 2000.

Papadimitriou and Piliouras. "Game dynamics as the meaning of a game." *ACM SIGecom Exchanges*, 2019.

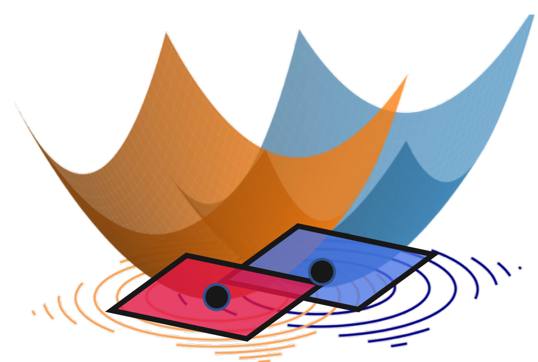
A solution concept of *continuous games*



Discrete-time dynamics

$$H \leftarrow H + \alpha \Delta \mathrm{H}$$
$$M \leftarrow M + \beta \Delta \mathrm{M}$$

Sufficient condition for a *differential Nash equilibrium*



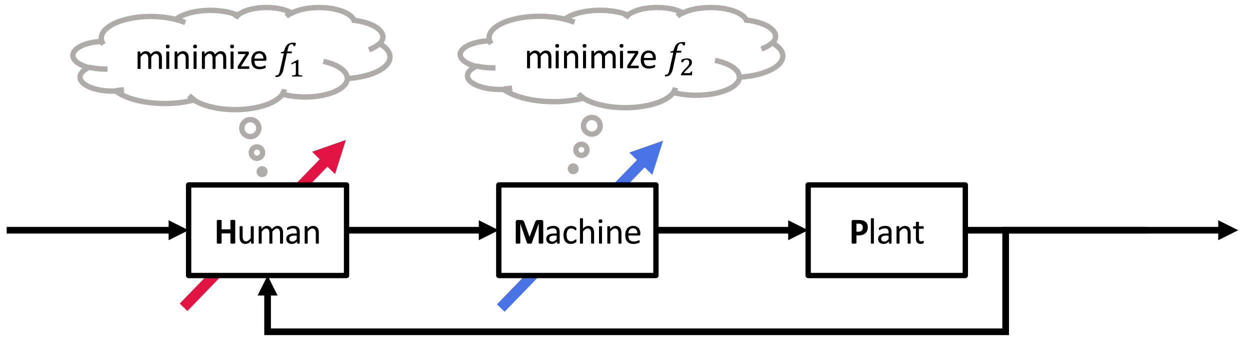
Derivatives:

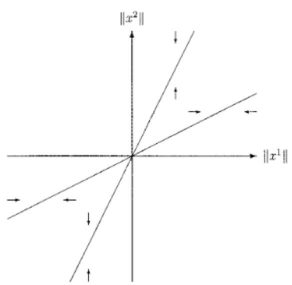
1st order:	flat
2nd order:	positive

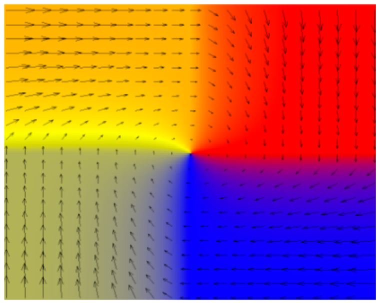
Chasnov, Ratliff, Mazumdar, and Burden. "Convergence analysis of gradient-based learning in continuous games." In *Uncertainty in Artificial Intelligence,* PMLR, 2020.

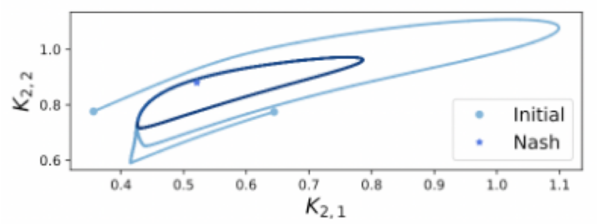
Ratliff, Burden, and Sastry. "Characterization and computation of local Nash equilibria in continuous games." In *Allerton*. IEEE, 2013.

Fundamental issues with learning in games









Hart and Mas-Colell. "Uncoupled dynamics do not lead to Nash equilibrium." *American Economic Review*, 2003.

Daskalakis, Goldberg, and Papadimitriou. "The complexity of computing a Nash equilibrium." Electronic Colloquium on Computational Complexity, 2005.

Mazumdar, Ratliff, Jordan, and Sastry. "Policy-gradient algorithms have no guarantees of convergence in linear quadratic games." In *AAMAS,* 2020.

Theoretical questions on game dynamics

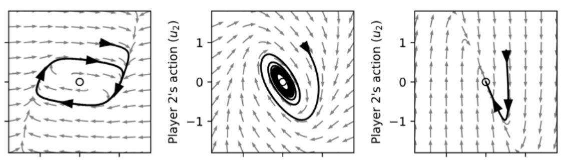
Discrete-time dynamics

$$H_{t+1} \leftarrow H_t + \alpha \Delta \mathrm{H_t}$$
$$M_{t+1} \leftarrow M_t + \beta \Delta \mathrm{M_t}$$

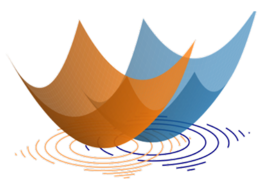
Gradients: $\Delta \mathrm{M_t}, \Delta H_t$

Learning rates: α, β

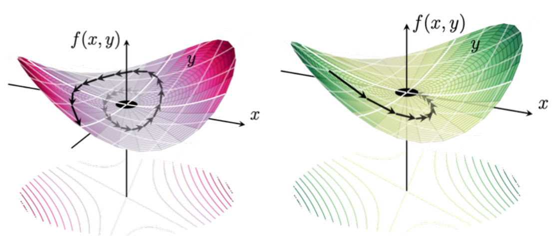
- **Q1** (*convergence*): as time t → ∞, does $\Delta H_t \to 0, \Delta M_t \to 0$?



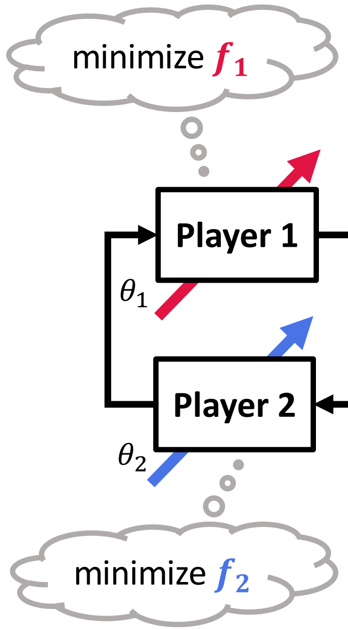
- **Q2** (*equilibrium*): when $\Delta H = 0, \Delta M = 0$, are the costs minimized?

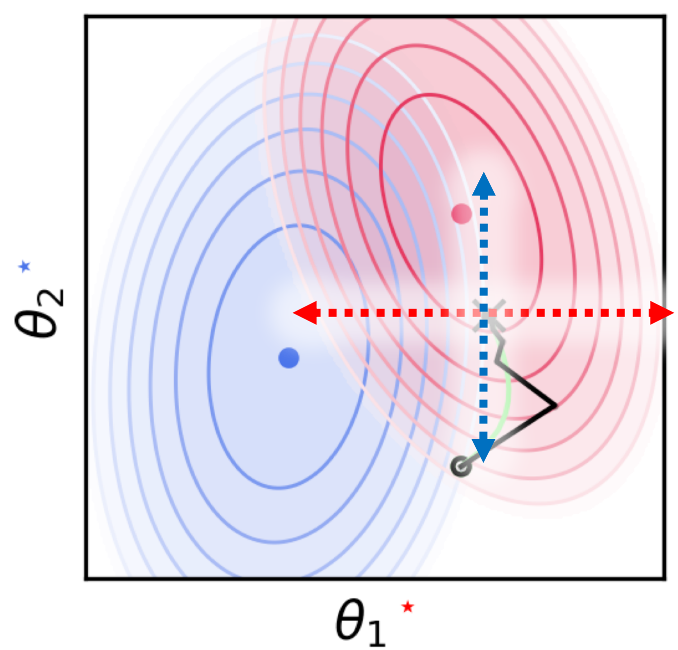


- **Q3** (*learning rates*): as $\tau := \beta/\alpha \to \infty$, what happens to the equilibrium?



Example: simultaneous gradient descent



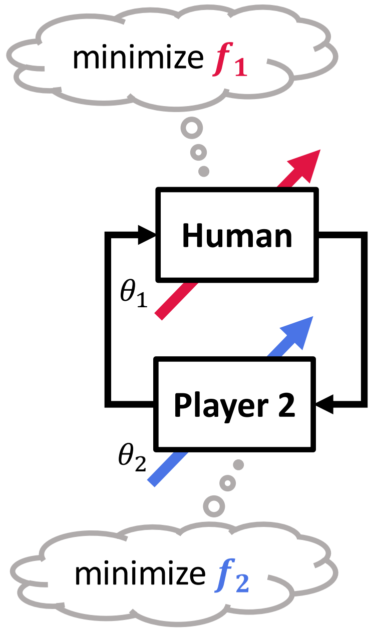


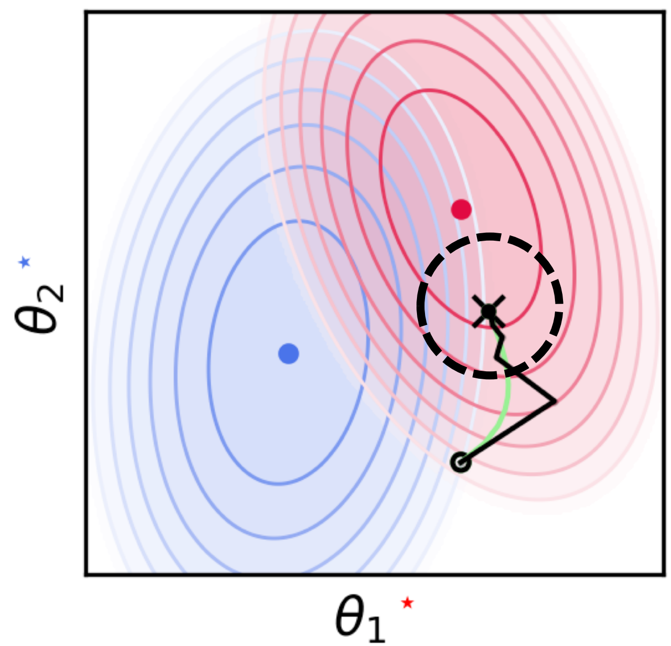
$$\theta_1 \leftarrow \theta_1 - \alpha \frac{\partial}{\partial \theta_1} f_1(\theta_1, \theta_2)$$

$$\theta_2 \leftarrow \theta_2 - \beta \frac{\partial}{\partial \theta_2} f_2(\theta_1, \theta_2)$$

Static webpage:
https://dynam.space/demo/

Example: simultaneous gradient descent





$\theta_1 \leftarrow$ horizontal mouse position

$\theta_2 \leftarrow \theta_2 - \beta \frac{\partial}{\partial \theta_2} f_2(\theta_1, \theta_2)$

Static webpage:
https://dynam.space/demo/

Game Jacobian of the learning dynamics

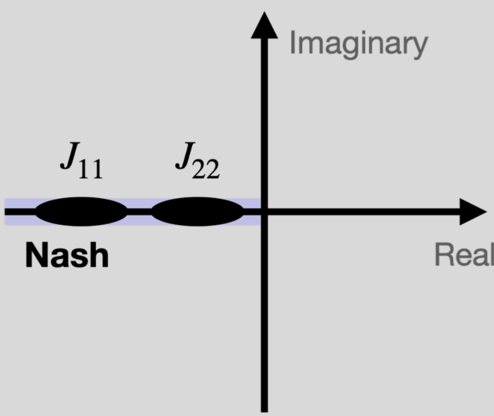
Linearized dynamics about an equilibrium:

$$\dot{\theta} = J(\tilde{\theta})\theta, \qquad g(\tilde{\theta}) = 0, \qquad J(\tilde{\theta}) = \begin{bmatrix} J_{11} & \boxed{J_{12}} \\ \boxed{J_{21}} & J_{22} \end{bmatrix}$$

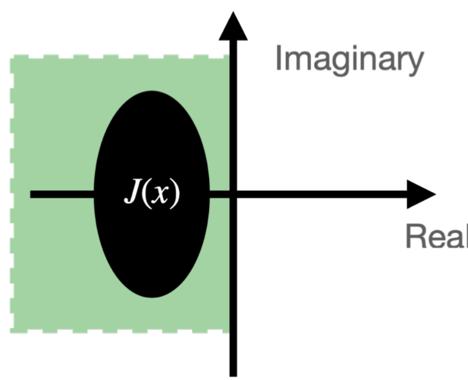
1st order derivative

2nd order derivatives

Individual **terms:**
Nash



Interaction **terms:**
stability



Decomposition:
class

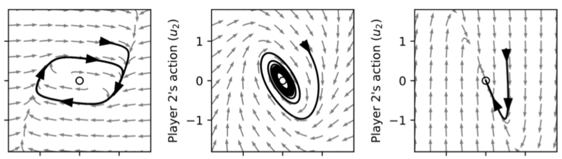
$$J = \begin{bmatrix} A & P \\ P^\top & D \end{bmatrix} + \begin{bmatrix} 0 & Z \\ -Z^\top & 0 \end{bmatrix}$$

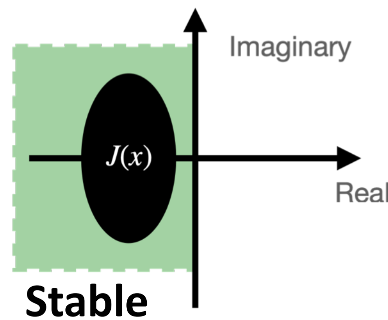
Chasnov, Calderone, Açıkmeşe, Burden, and Ratliff. "Stability of Gradient Learning Dynamics in Continuous Games: Vector Action Spaces." *arXiv: 2011.05562*, 2021.

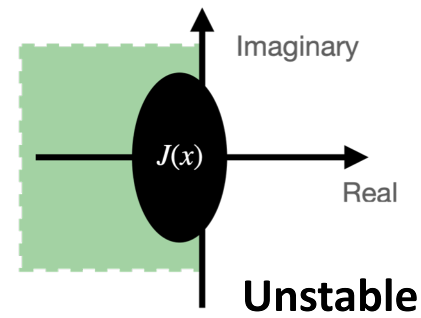
arXiv: 2011.05562

Theoretical results using the game Jacobian

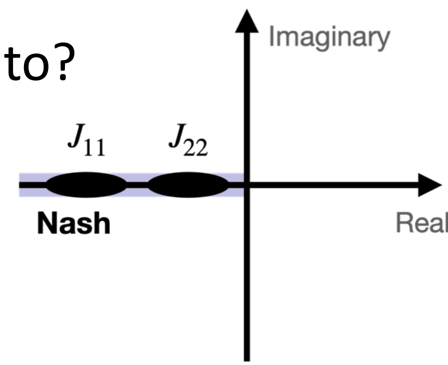
- **Q1:** does learning converge?

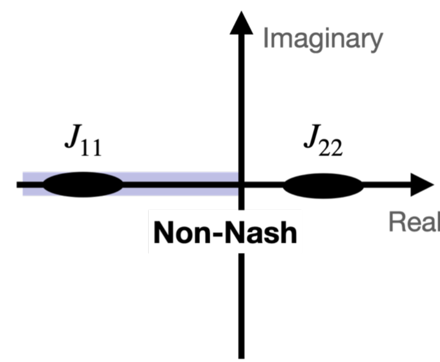




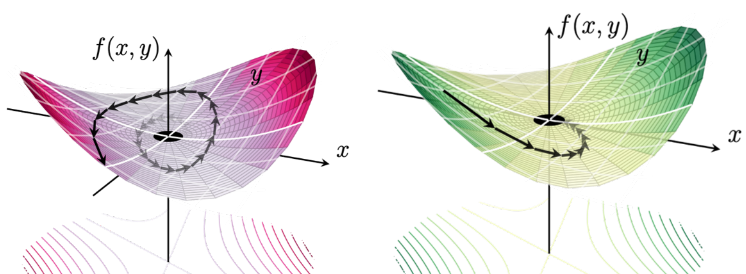


- **Q2:** what does it converge to?





- **Q3**: how do learning rates affect the dynamics?



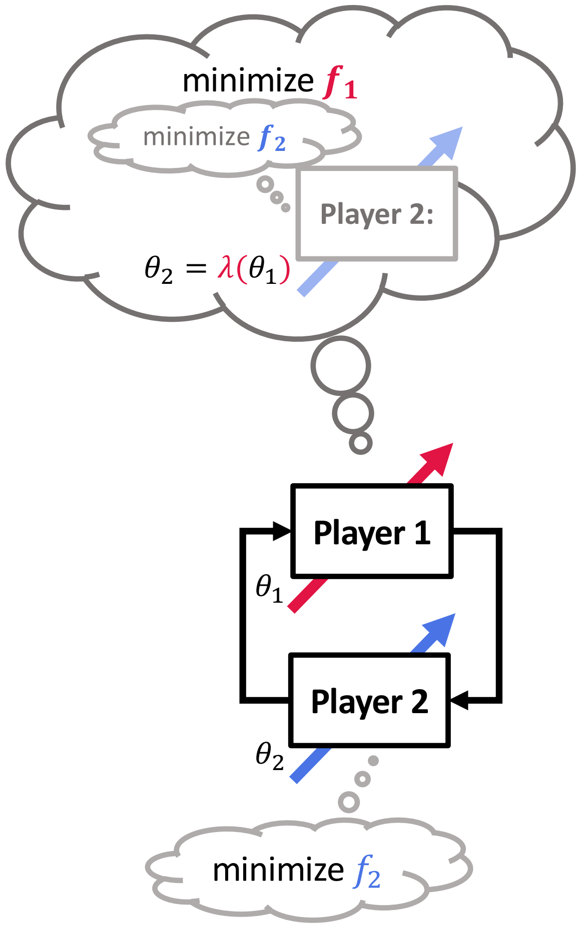
$$J = \begin{bmatrix} J_{11} & J_{12} \\ \tau J_{21} & \tau J_{22} \end{bmatrix}$$

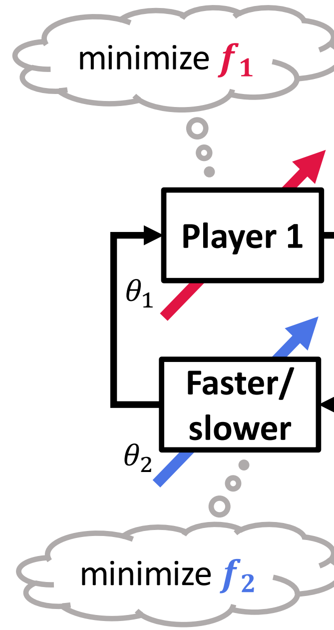
$$S_1 = J_{11} - J_{12} J_{22}^{-1} J_{21}$$

Chasnov, Calderone, Açıkmeşe, Burden, and Ratliff. "Stability of Gradient Learning Dynamics in Continuous Games: Vector Action Spaces." *arXiv: 2011.05562*, 2021.

arXiv: 2011.05562

Modeling each other and time-scale separation





$$\theta_1 \leftarrow \theta_1 - \alpha \frac{d}{d\theta_1} f_1(\theta_1, \lambda(\theta_1))$$

$$\lambda(\theta_1) = \arg\min_{\theta} f_2(\theta_1, \theta)$$

$$\theta_2 \leftarrow \theta_2 - \beta \frac{\partial}{\partial\theta_2} f_2(\theta_1, \theta_2)$$

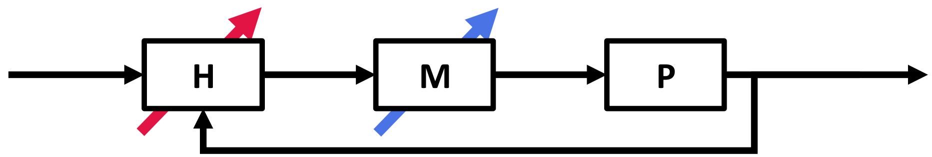
$$\theta_1 \leftarrow \theta_1 - \alpha \frac{\partial}{\partial\theta_1} f_1(\theta_1, \theta_2)$$

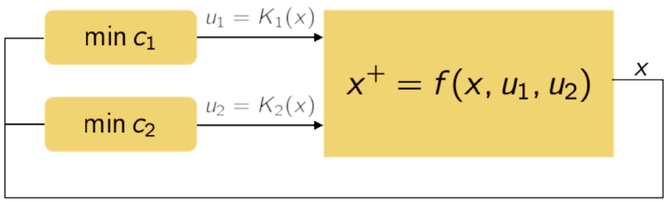
$$\theta_2 \leftarrow \theta_2 - \alpha\tau \frac{d}{d\theta_2} f_2(\theta_1, \theta_2)$$

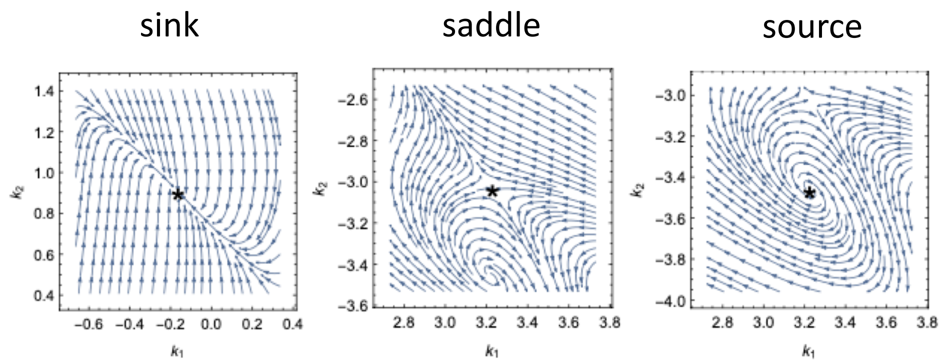
Fiez, Chasnov, and Ratliff. "Implicit learning dynamics in Stackelberg games: Equilibria characterization, convergence analysis, and empirical study." In *ICML*. PMLR, 2020.

Chasnov, Calderone, Açıkmeşe, Burden, and Ratliff. "Stability of Gradient Learning Dynamics in Continuous Games: Scalar Action Spaces." In *CDC*. IEEE, 2020.

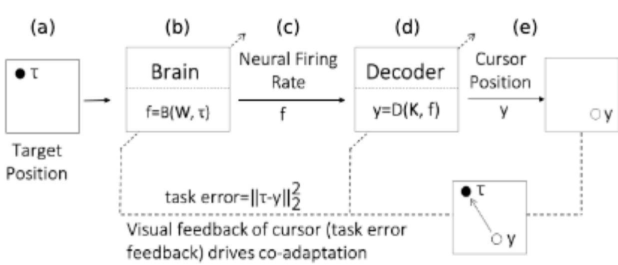
Towards the design of mechanisms

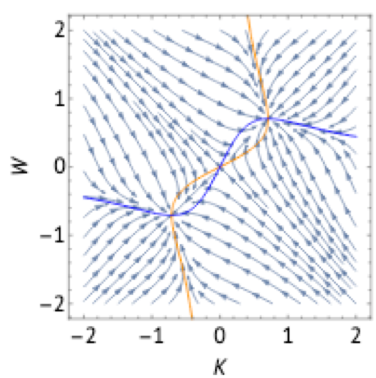


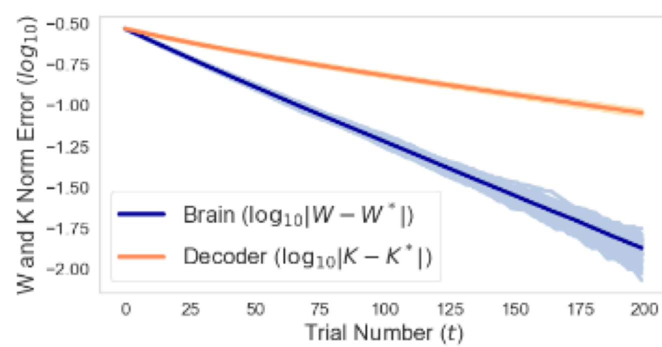




Chasnov, Yamagami, Parsa, Ratliff, and Burden. "Experiments with sensorimotor games in dynamic human/machine interaction." In *Micro-and Nanotechnology Sensors, Systems, and Applications XI*, 2019.







Madduri, Burden, and Orsborn. "A Game-Theoretic Model for Co-Adaptive Brain-Machine Interfaces." *bioRxiv:2020.12.11.421800*, 2020.

***bioRxiv*:2020.12.11.421800**

Thank you!

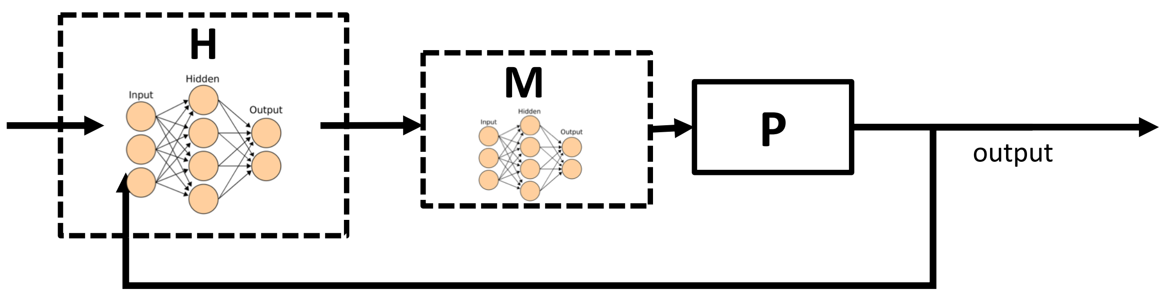
Advisors: Dr. Sam Burden and Dr. Lillian Ratliff

Collaborators: Behcet Açıkmeşe, Dan Calderone, Amber Chou, Tanner Fiez, Maneeshika Madduri, Eric Mazumdar, Skye Mceowen, Joey Sullivan, Momona Yamagami, Liyuan Zheng

COMPUTATIONAL NEUROSCIENCE CENTER

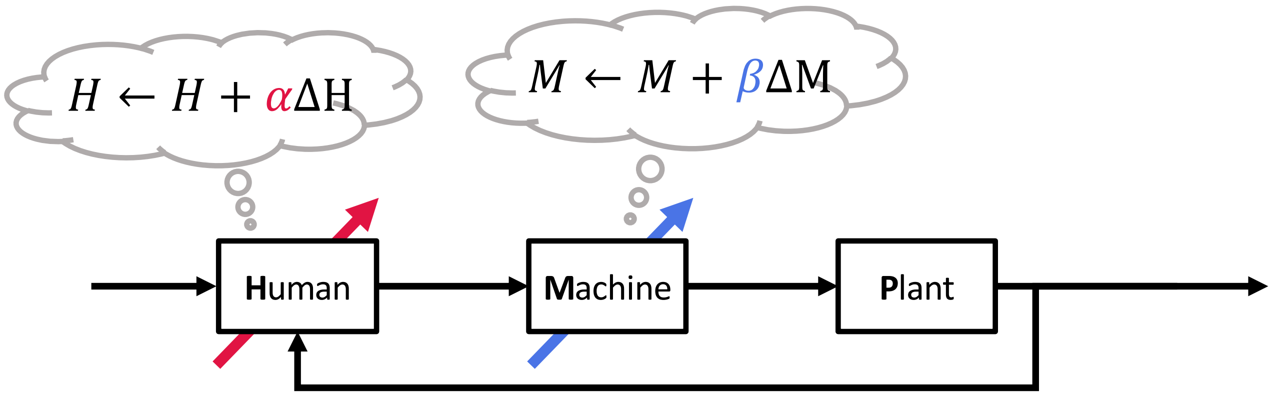
NIH 5T90DA032436-09 (MPI):
Computational Neuroscience Graduate Training Program

NSF CPS Medium #1836819:
Certifiable reinforcement learning for cyber-physical systems



Backup Slides

Continuous time approximation



Discrete-time dynamics

$$H \leftarrow H + \alpha \Delta \mathrm{H}$$
$$M \leftarrow M + \beta \Delta \mathrm{M}$$

for small α, β

$$\Longrightarrow$$

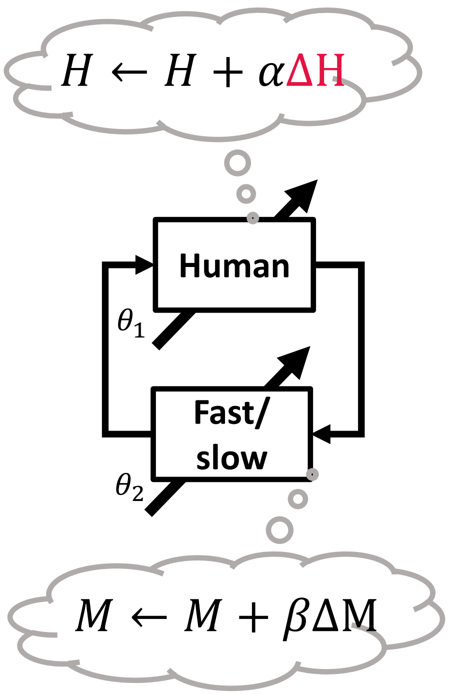
Continuous-time dynamics

$$\dot{H} = \Delta \mathrm{H}$$
$$\dot{M} = \tau \, \Delta \mathrm{M}$$

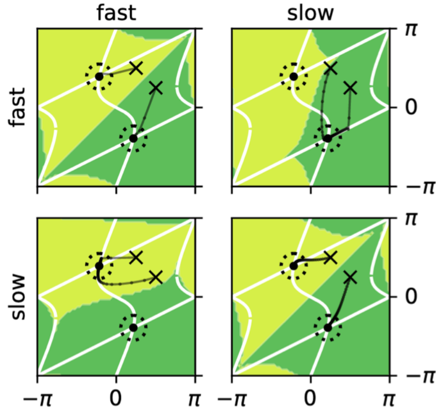
Chasnov, Ratliff, Mazumdar, and Burden. "Convergence analysis of gradient-based learning in continuous games." In *Uncertainty in Artificial Intelligence,* PMLR, 2020.

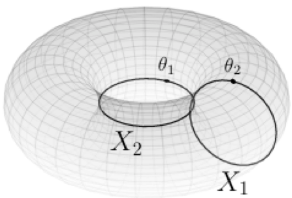
https://dynam.space/demo/torus.html

Timescale separation affects dynamics and equilibrium

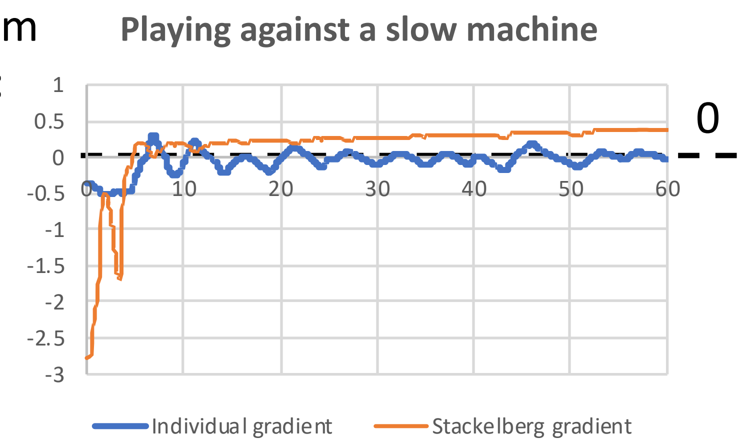


Region of attraction:

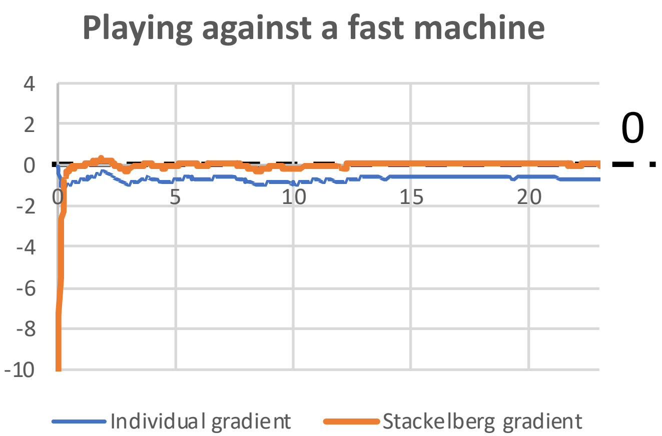




Equilibrium selection:



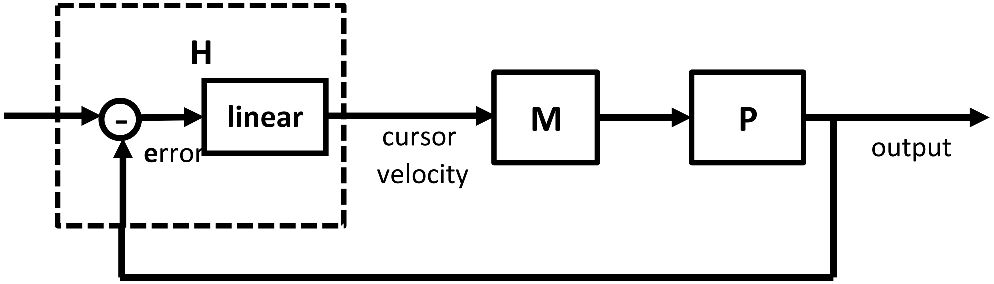
$$\Delta\mathrm{H} \approx \frac{\partial}{\partial H} f_1(H, M) \to 0$$



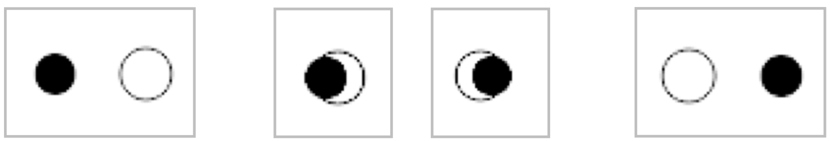
$$\Delta\mathrm{H} \approx \frac{d}{dH} f_1\big(H, \lambda(H)\big) \to 0$$

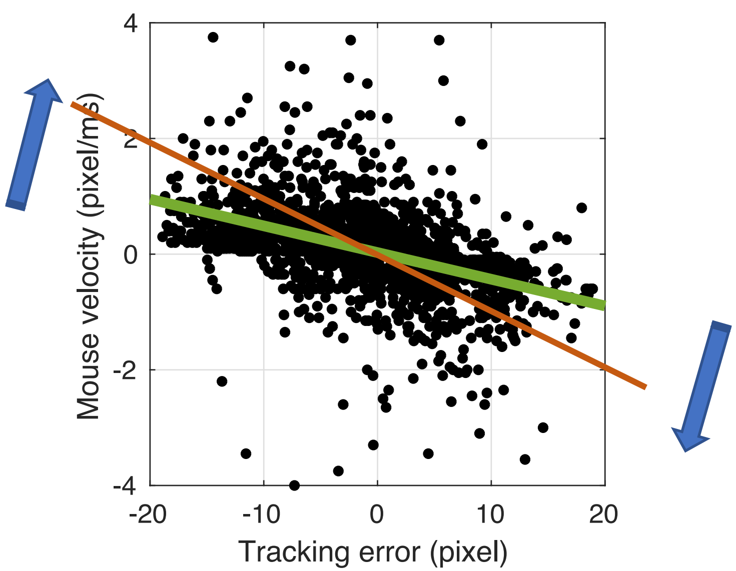
https://dynam.space/demo/tracking.html

Learning a linear policy



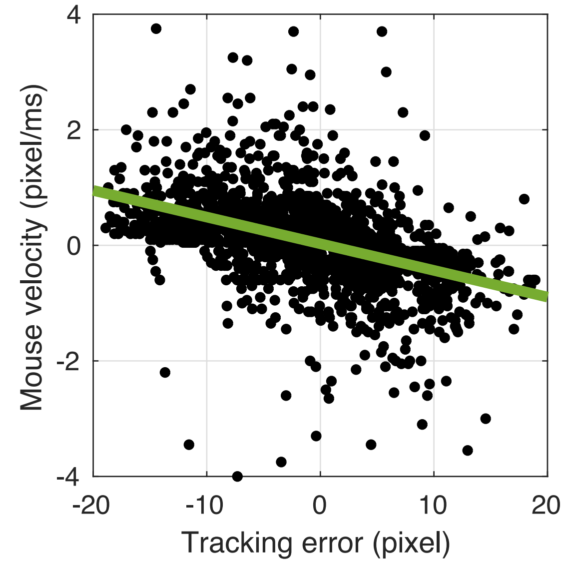
error<0 error≈0 error>0



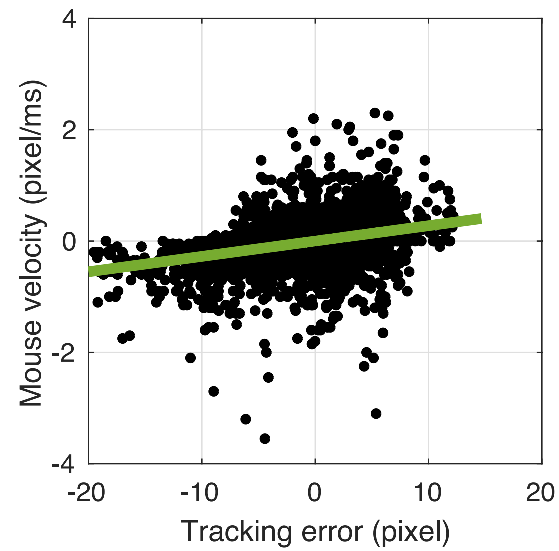


Learning...?

$M = 1$

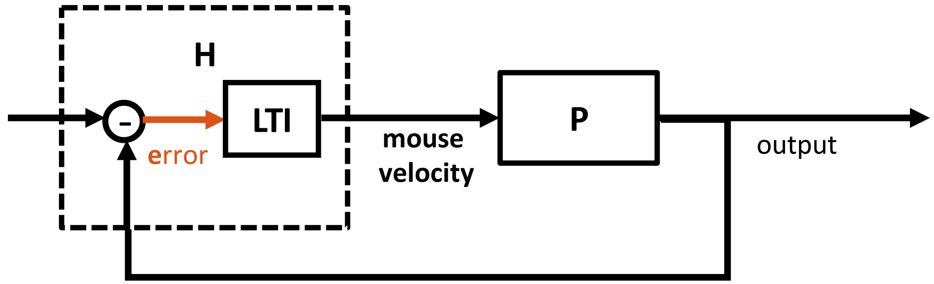


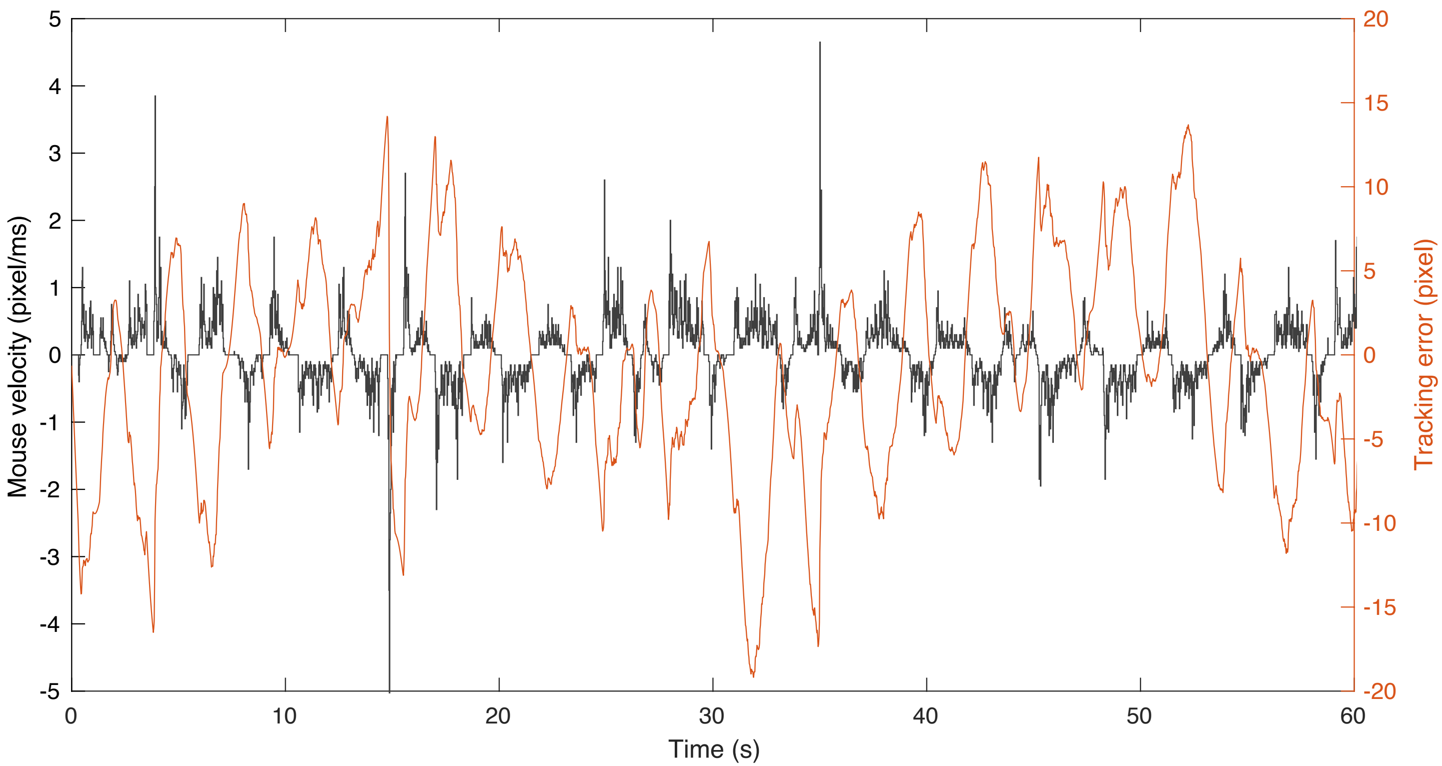
$M = -1$



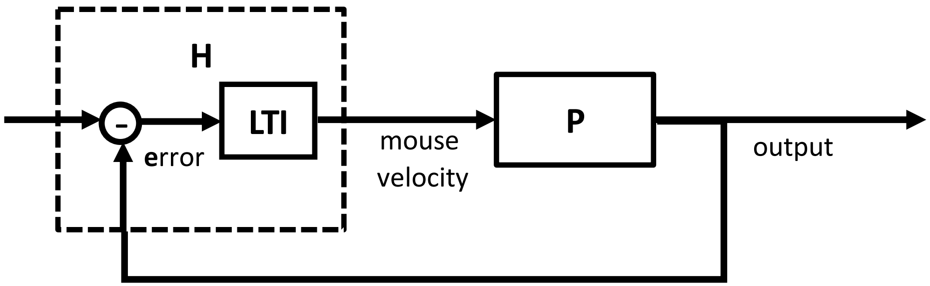
https://dynam.space/demo/tracking.html?mode=velocity

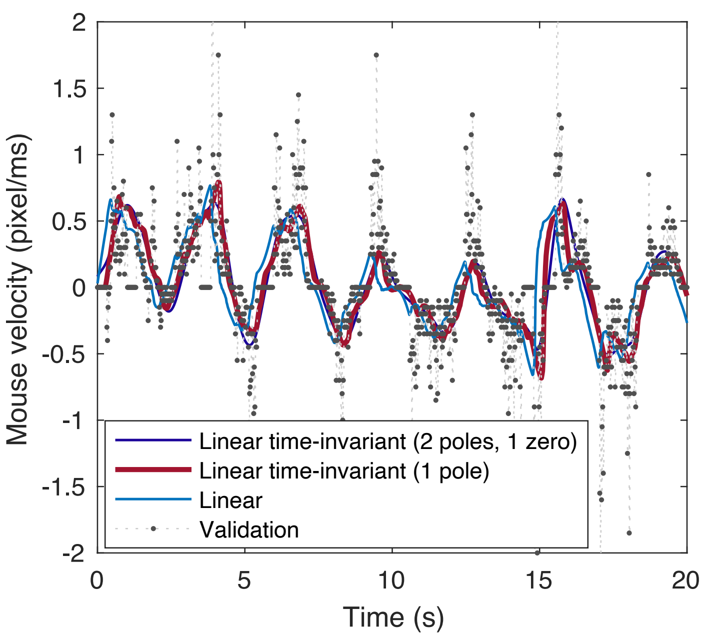
Learning a linear time-invariant policy





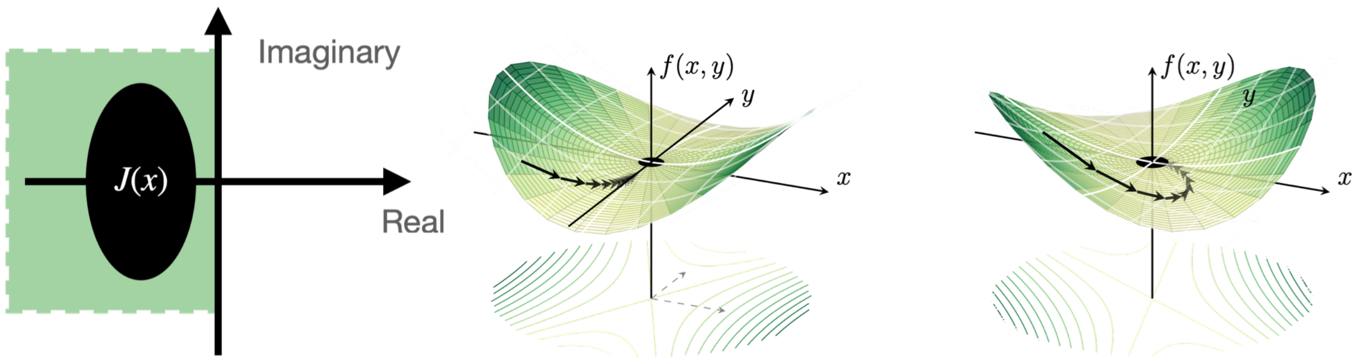
Learning a linear time-invariant policy



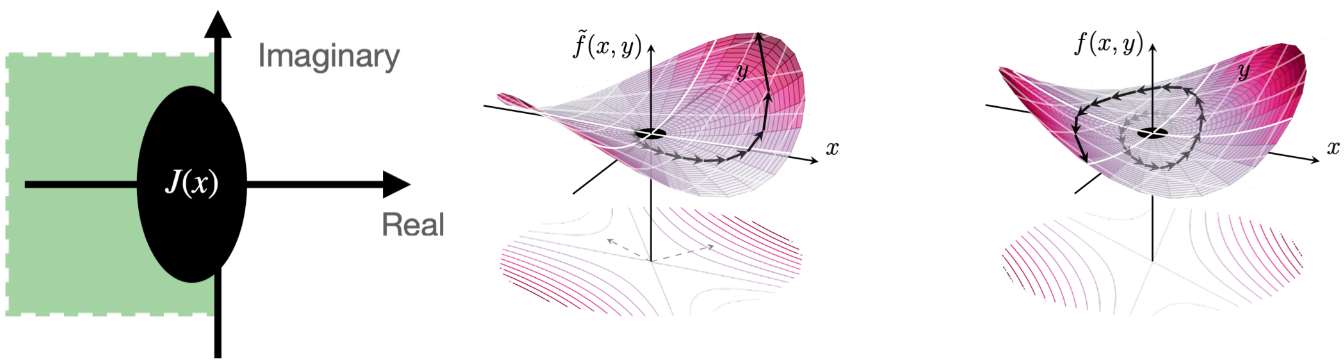


Stability of game dynamics near fixed points

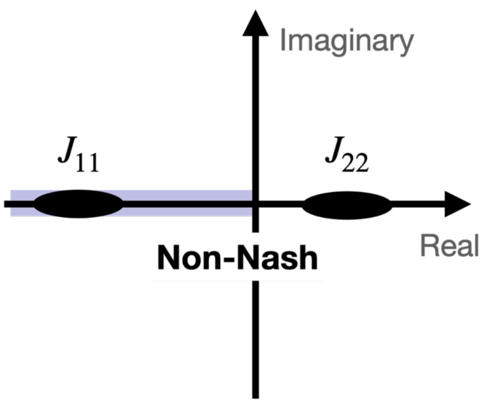
Stable:

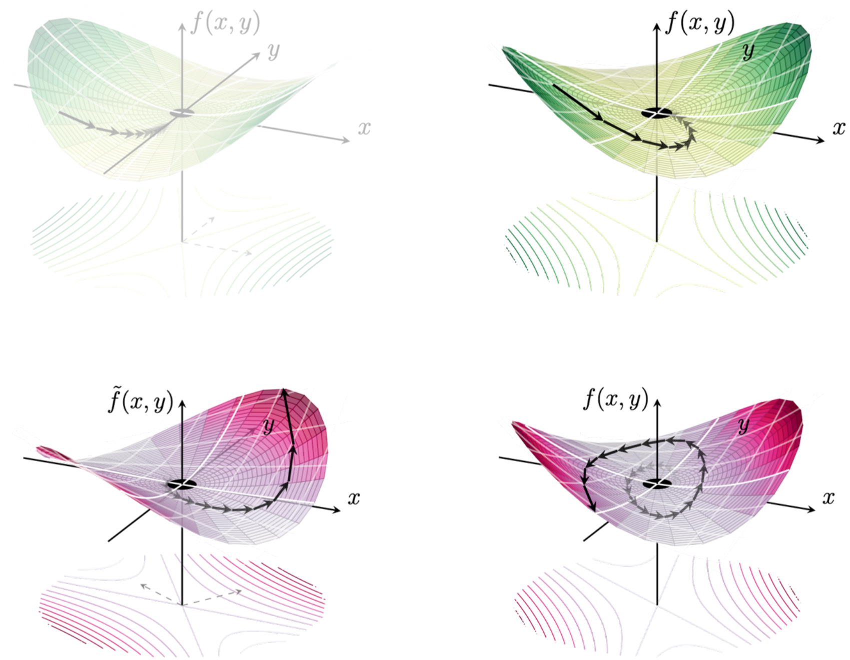


Unstable:

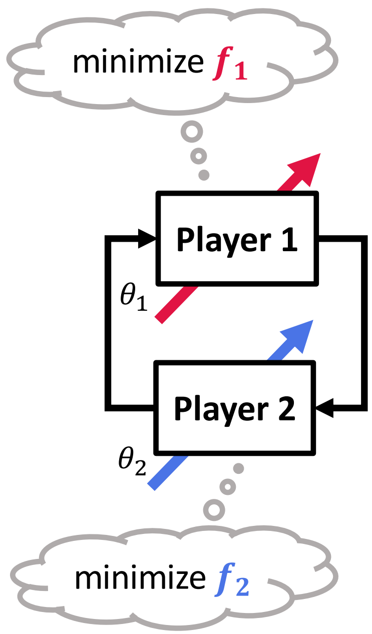


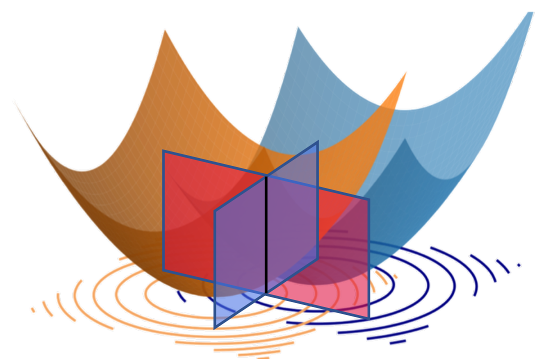
Non-Nash equilibria can be stable





Freezing the parameters of one player





Timescale separation is related to a leader-follower structure

The *Schur complement*:
timescale and order of play

$$\begin{bmatrix} \boxed{J_{11}} & \boxed{J_{12}} \\ \boxed{J_{21}} & \boxed{J_{22}} \end{bmatrix} \qquad S_1 = \boxed{J_{11}} - \boxed{J_{12}}\,\boxed{J_{22}^{-1}}\,\boxed{J_{21}}$$

$$\theta_1^+ = \theta_1 - \gamma_1 \tfrac{d}{d\theta} f_1(\theta_1, \lambda(\theta_1))$$

$$\lambda(\theta_1) = \arg\min_{\theta} f_2(\theta_1, \theta)$$

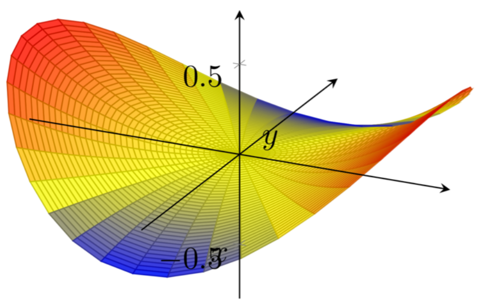
Fiez, Chasnov, and Ratliff. "Implicit learning dynamics in Stackelberg games: Equilibria characterization, convergence analysis, and empirical study." In *ICML.* PMLR, 2020.

Chasnov, Burden, and Ratliff. ”Gradient-based learning with a finite timescale separation in continuous games " 2021. In progress.

Classes of games with strong properties

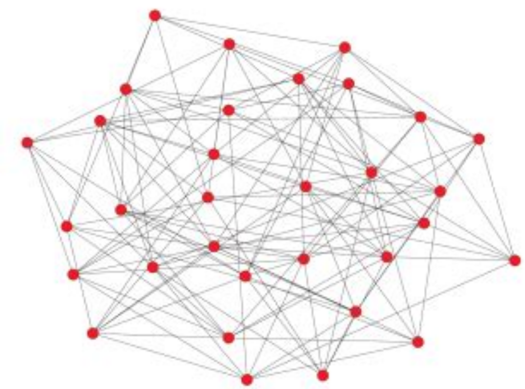
Adversarial / zero-sum

Potential game



$$\min_{\theta_1} \; f(\theta_1, \theta_2)$$

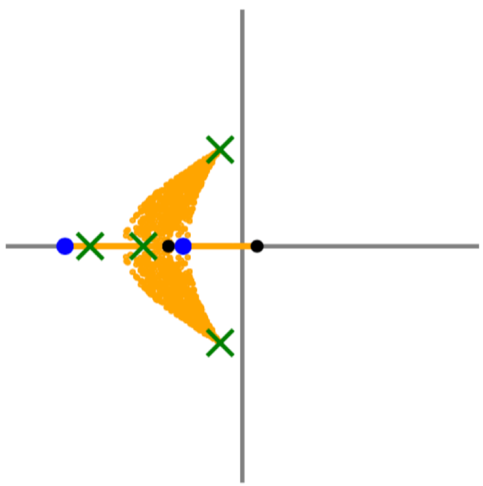
$$\max_{\theta_2} \; f(\theta_1, \theta_2)$$



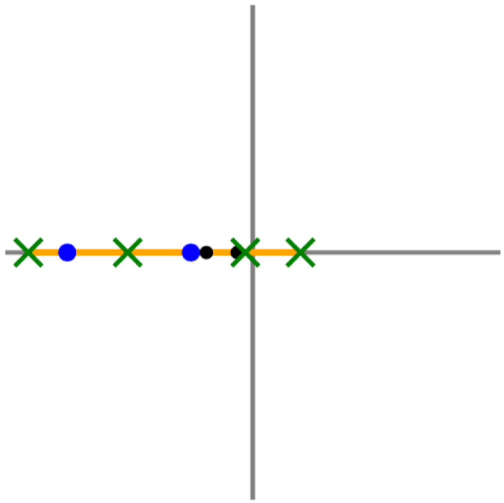
Von Neumann, Morgenstern. Theory of games and economic behavior. Princeton university press; 1944.

Monderer, Shapley. Potential Games. 1996

Spectrum of zero-sum and potential games

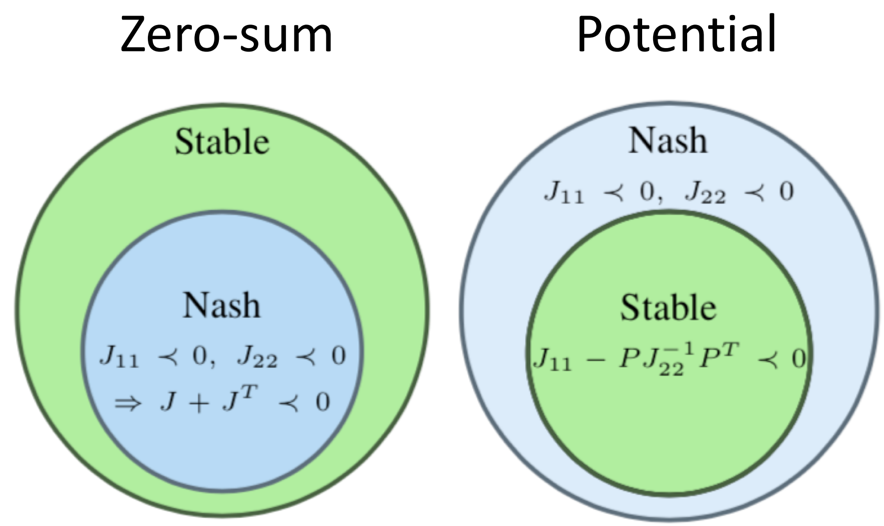


$$\begin{bmatrix} A & B \\ -B^\top & C \end{bmatrix}$$



$$\begin{bmatrix} A & B \\ B^\top & D \end{bmatrix}$$

Why zero-sum Nash are stable and stable equilibria of potential games are Nash



$$\begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix}$$

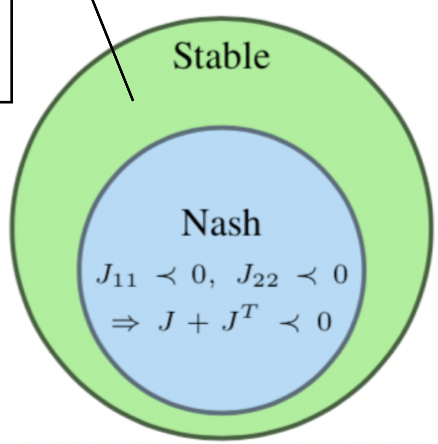
Implications of different classes of games

- Zero-sum game

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

$$\begin{bmatrix} A & B \\ -\tau B^\top & \tau D \end{bmatrix}$$

Spurious attractors (at least one agents is not optimal)

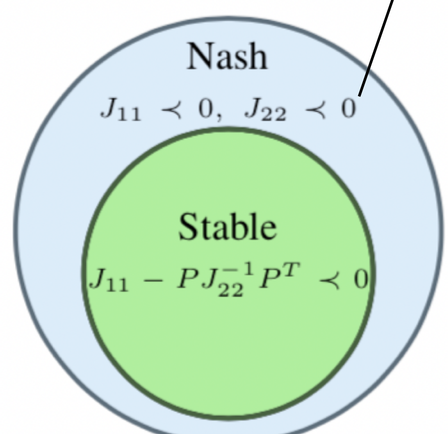


Potential game

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

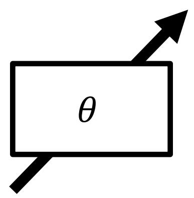
$$\begin{bmatrix} A & B \\ \tau B^\top & \tau D \end{bmatrix}$$

Not computable (using gradient dynamics)

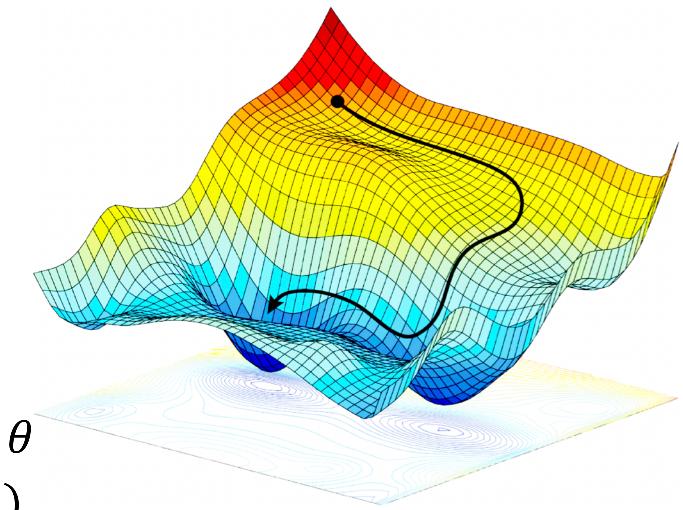


Single objective optimization with gradients

- Optimization abstraction
  - Learning modules



$$\theta \leftarrow \theta + \alpha\Delta\theta$$



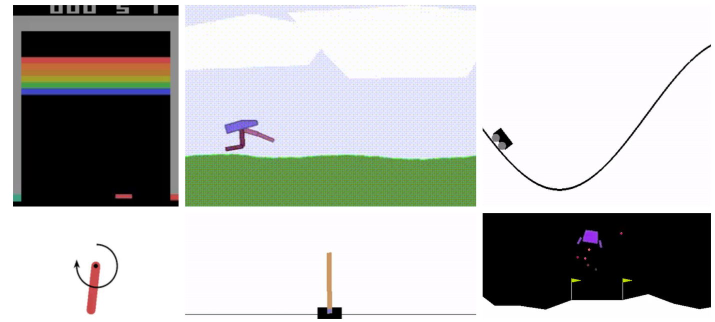
  - Gradient of reward: $\Delta\theta = \left.\frac{dR}{d\theta}\right|_{\theta=\theta}$
  - Converges to maximizer of reward. For $t \to \infty$, θ

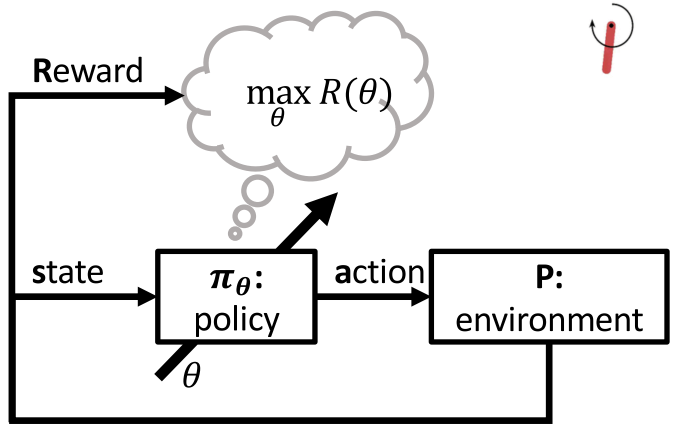
$$\theta^* = \arg\max R(\theta)$$

  - Local maximum ⇔ stable equilibrium:

$$J = \left.\frac{d^2R}{d\theta^2}\right|_{\theta=\theta^*} < 0$$

Reinforcement learning is optimization





Policy gradient: $\theta \leftarrow \theta + \alpha\Delta\theta$ $\tau = (s_0, a_0, s_1, a_1, \dots, s_T, a_T)$

$$\Delta\theta = \mathrm{E}_{\tau\sim\pi_\theta}\left[\int_t \log \pi_\theta(a_t|s_t) R(\tau) dt\right]$$