On Passivity for Learning Dynamics in Games

Lacra Pavel
University of Toronto

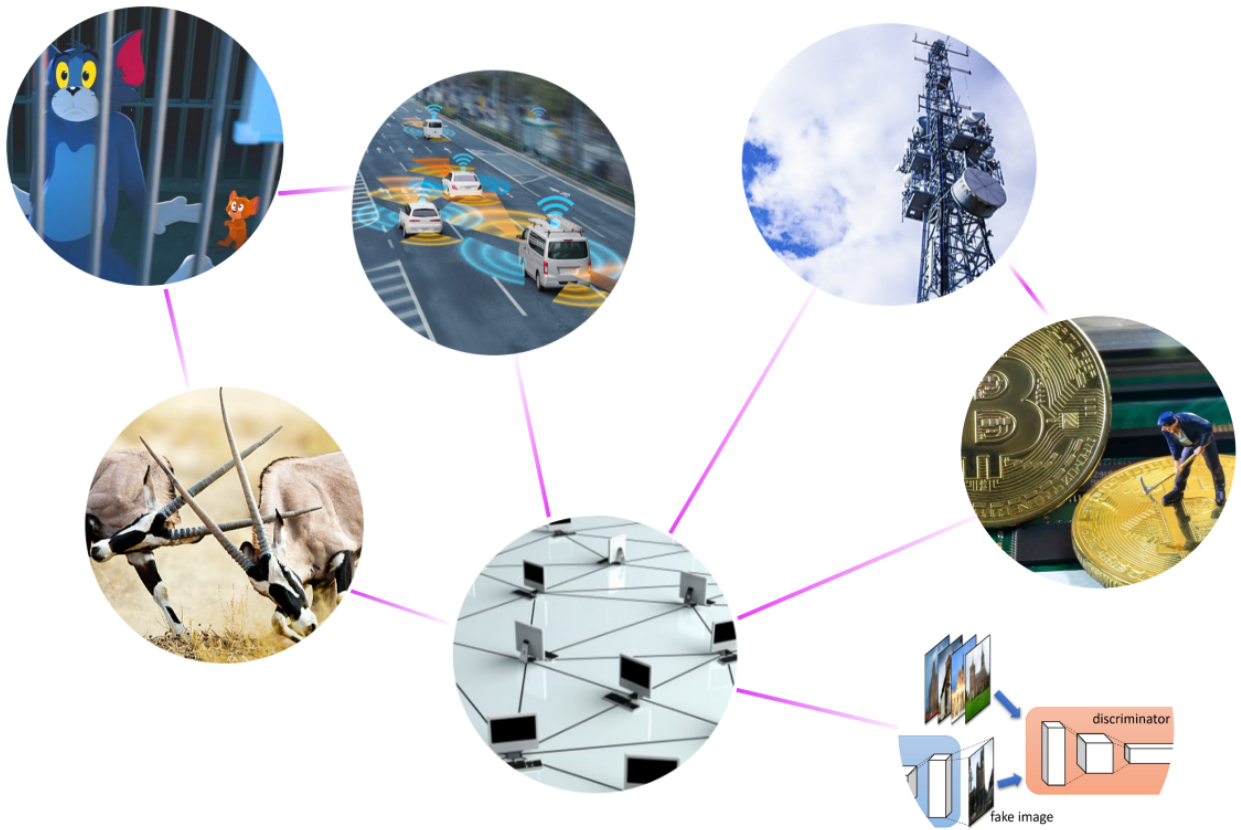
JST ASPIRE Kick-off Meeting, March 3, 2026

Credits

Bolin Gao

Dian Gadjov

Game Theory - Learning (Online) Setting



- A Nash equlibrium (NE) arises as the limiting point of *repeated play* in which (less than fully) rational players update their decision (strategy/action).
  - a collective strategic state where decisions are individual optimal for each of them (no reason to unilaterally deviate).

Learning in Games: An Incomplete List

- Dynamical systems and games

  Flam,"Equilibrium, evolutionary stability and gradient dynamics, " Int. Game Theory Review, 2002.

  Flam & Morgan,"Newtonian Mechanics and Nash Play," Int. Game Theory Review, 2004.

  Shamma & Arslan,"Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria," IEEE Trans. on Automatic Control (TAC), 2005.

- Passivity (and Games)

  Fox & Shamma,"Population Games, Stable Games, and Passivity," Games, 2013.

  Mabrok & J.S. Shamma,"Passivity analysis of higher-order evolutionary dynamics & population games," CDC2016.

- Higher-order learning

  Laraki & Mertikopoulos,"Higher order game dynamics," J. Econ. Theory, 2013.

  Arslan & Shamma,"Anticipatory learning in general evolutionary games," CDC2006.

Learning in Games: An Incomplete List

- Best-response and Fictitious-play

  Benaim, Hofbauer & Sorin,"Stochastic approximation and differential inclusions, " SIAM J. Control Optim, 2005.

  Hofbauer & Sorin,"Best-response dynamics for continuous zero-sum games, " Discrete & Cont. Dyn. Sys, 2006.

  Hofbauer & Hokpins,"Learning in perturbed asymmetric games, " Games & Ec. Behavior., 2005.

  Benaim, Hofbauer & Hokpins,"Learning in games with unstable equilibria, " J. Econ. Th., 2009.

- Q-learning and payoff-based learning

  Leslie & Collins,"Individual Q-learning in normal form games," SIAM J. Control Optim., 2005.

  Erev & Roth,"Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria," Amer. Econ. Rev., 1998.

- Reinforcement learning & Mirror Descent

  Coucheney, Gaujal & Mertikopoulos,"Penalty-regulated dynamics and robust learning procedures in games," Math. Oper Res, 2015.

  Mertikopoulos & Sandholm,"Learning in games via reinforcement and regularization," Math. Oper Res, 2016.

  Zhou, Mertikopoulos et al"Mirror descent learning in continuous games," CDC, 2017.

  Mertikopoulos, Zenati, Lecouat, Foo, Chandrasekhar & Piliouras,"Mirror descent in saddle-point problems: Going the extra (gradient) mile," ICLR 2019.

  Mertikopoulos, Papadimitriou & Piliouras,"Cycles in adversarial regularized learning," ACM-SIAM, 2018.

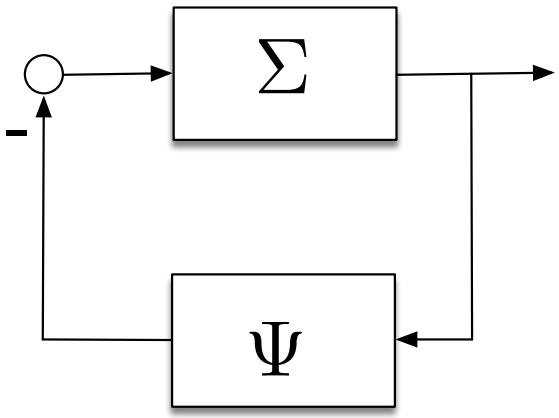
Learning in Games

- A plethora of algorithms and dynamics:
  - Best-response (BR) play.
  - Fictitious play (play optimally/BR against the empirical distribution of opponents' past plays).
  - Gradient-play ("better-response" play).
  - Projected-gradient play.
  - Reinforcement-learning (payoff-based play).
  - Mirror-descent play.
  - Q-learning play.
  - Bandit-learning play (zero-order play).
  - Proximal-point algorithm.
  - Forward-backward iteration.
  - ...

- Settings: infinite (continuous-kernel) games, finite-action games, evolutionary games, in discrete-time or continuous-time.
- Work for special classes of games: zero-sum games, two-player games, 2×2 games, potential games, strictly/strongly monotone games.

This Talk

- Questions:
  - Why some learning dynamics work in certain games and others don't?
  - How can we generalize them in a more systematic manner ?
- Passivity theory can help.



Focus on finite-action games.

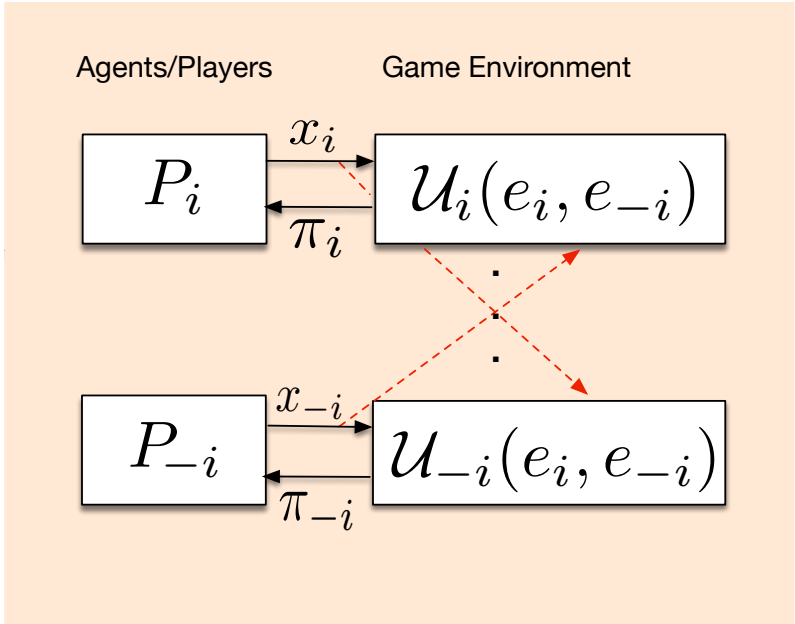
This Talk

- A class of multi-agent learning dynamics/ODEs through the lens of passivity.

  Show that the learning dynamics

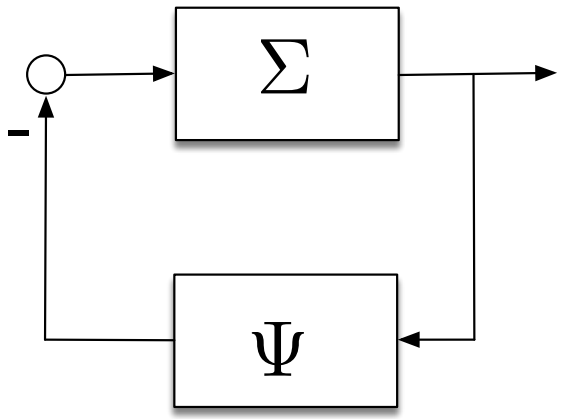
  - form a passive system.
  - solution trajectories converge to equilibria of a class of games characterized by monotonicity properties.
  - can be extended to a class of higher-order dynamics ($\ddot{z}$, $\dddot{z}$, $\ldots$, z^n) that improves upon the original dynamics ($\dot{z}$).

Learning in Games: A Generic Formulation



- Player i (red, left) strategy) is x_i (everybody else's is x_{-i}), individual payoff $\mathcal{U}_{,i}$, adjusts its decision using dynamics denoted P_i (right).
- Everyone else's algorithm/dynamics is denoted P_{-i}.
- The overall learning dynamics is feedback interconnected system $P := (P_i, P_{-i})$.

Learning Dynamics via Passivity Framework



- Cast learning dynamics $P := (P_i, P_{-i})$ as instances of a feedback interconnection (Σ, Ψ) between
  - some passive dynamical system Σ.
  - some specific game mapping Ψ.
- Convergence to NE from standard passivity theory
  - simple, concise arguments.
  - ideas to design novel learning algorithms/dynamics.
- Framework explains convergence (analysis) and how to generalize (design).

A bit of overview

- Passive Dynamical Systems
- Monotone maps

Passive Dynamical Systems

Let

$$\Sigma : \begin{cases} \dot{x} = f(x,u), \\ y = h(x,u), \end{cases}$$

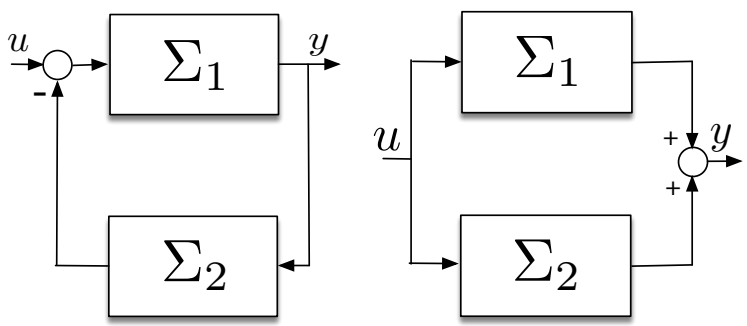
with $x \in \mathbb{R}^n, u \in \mathbb{R}^q, y \in \mathbb{R}^q, f$ locally Lipschitz $f(0,0) = 0, h(0) = 0$.

Definition (Willems'72)

System Σ is *passive* if $\exists$ diff., positive semi-def. storage fcn. $V : \mathbb{R}^n \to \mathbb{R}$, $V(0) = 0$ s.t.

$$\dot{V}(x) \leq y^T u, \ \forall u \in \mathbb{R}^q, \forall x \in \mathbb{R}^n. \tag{1}$$

- Feedback and parallel interconnections preserve these properties.
- Feedback system is stable if both are passive and asy. stable if:
  - (at least one) strict passive.



Passive Dynamical Systems

Let

$$\Sigma : \begin{cases} \dot{x} = f(x,u), \\ y = h(x,u), \end{cases}$$

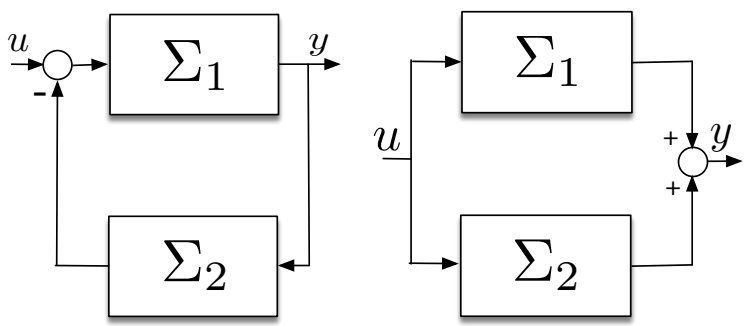
with $x \in \mathbb{R}^n, u \in \mathbb{R}^q, y \in \mathbb{R}^q, f$ locally Lipschitz $f(0,0) = 0, h(0) = 0$.

Definition (Willems'72)

System Σ is *passive* if $\exists$ diff., positive semi-def. storage fcn. $V : \mathbb{R}^n \to \mathbb{R}$, $V(0) = 0$ s.t.

$$\dot{V}(x) \leq y^T u, \ \forall u \in \mathbb{R}^q, \forall x \in \mathbb{R}^n. \tag{1}$$

- Feedback and parallel interconnections preserve these properties.
- Feedback system is stable if both are passive and asy. stable if:
  - (at least one) strict passive.



Passive Dynamical Systems

Let

$$\Sigma : \begin{cases} \dot{x} = f(x,u), \\ y = h(x,u), \end{cases}$$

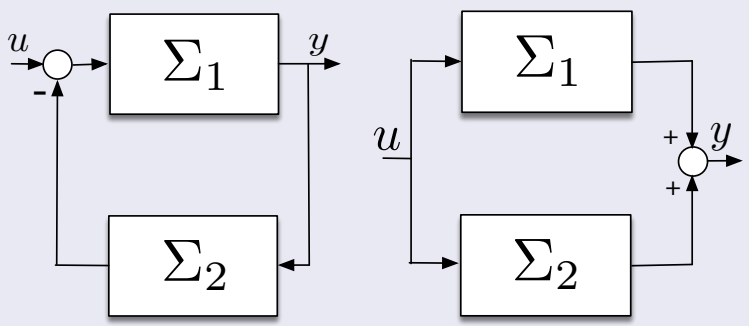
s.t. $0 = f(\overline{x},\overline{u}), \overline{y} = h(\overline{x},\overline{u})$, with $\overline{x}$ an equilibrium.

Definition (Hines, Arcak & Packard'11)

System Σ is *Equilibrium Independent Passive* (EIP) if $\exists$ storage function $V_{\overline{x}} : \mathbb{R}^n \to \mathbb{R}$ s.t. $V_{\overline{x}}(\overline{x}) = 0$

$$\dot{V}_{\overline{x}}(x) \leq (y-\overline{y})^\top (u-\overline{u}). \tag{2}$$

- Σ is *output-strict* EIP if for some $\beta > 0$
  $\dot{V}_{\overline{x}}(x) \leq (y-\overline{y})^\top (u-\overline{u}) - \beta \|y-\overline{y}\|^2$.
- Σ is *input-strict* EIP if for some $\mu > 0$
  $\dot{V}_{\overline{x}}(x) \leq (y-\overline{y})^\top (u-\overline{u}) - \mu \|u-\overline{u}\|^2$.



- EIP applies to any equilibrium point.
- Other versions: Incremental passivity, differential passivity, δ-passivity.
- LTI systems: all versions are equivalent.

Passive Dynamical Systems

Let

$$\Sigma : \begin{cases} \dot{x} = f(x,u), \\ y = h(x,u), \end{cases}$$

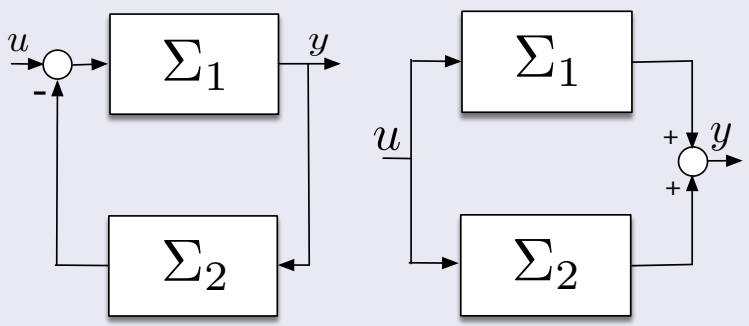
s.t. $0 = f(\overline{x},\overline{u}), \overline{y} = h(\overline{x},\overline{u})$, with $\overline{x}$ an equilibrium.

Definition (Hines, Arcak & Packard'11)

System Σ is *Equilibrium Independent Passive* (EIP) if $\exists$ storage function $V_{\overline{x}} : \mathbb{R}^n \to \mathbb{R}$ s.t. $V_{\overline{x}}(\overline{x}) = 0$

$$\dot{V}_{\overline{x}}(x) \leq (y-\overline{y})^\top (u-\overline{u}). \tag{2}$$

- Σ is *output-strict* EIP if for some $\beta > 0$
  $\dot{V}_{\overline{x}}(x) \leq (y-\overline{y})^\top (u-\overline{u}) - \beta \|y-\overline{y}\|^2$.
- Σ is *input-strict* EIP if for some $\mu > 0$
  $\dot{V}_{\overline{x}}(x) \leq (y-\overline{y})^\top (u-\overline{u}) - \mu \|u-\overline{u}\|^2$.



- EIP applies to any equilibrium point.
- Other versions: Incremental passivity, differential passivity, δ-passivity.
- LTI systems: all versions are equivalent.

Monotone Maps

Let $y = \Psi(u)$, $\Psi : \mathcal{D} \subseteq \mathbb{R}^n \to \mathbb{R}^n$

Definition

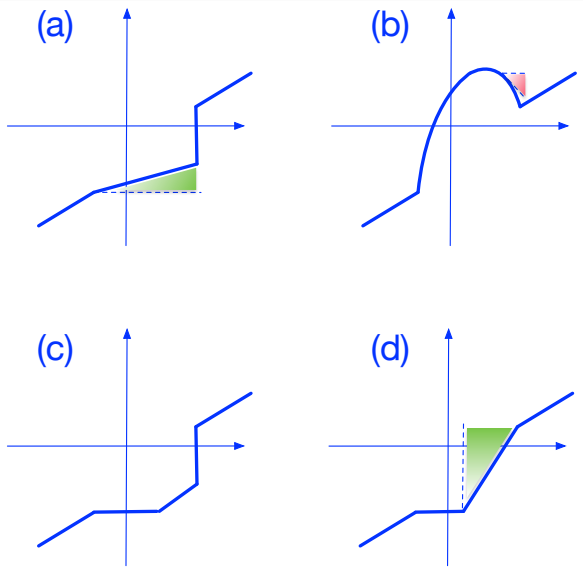
Map Ψ is *monotone* if

$$(\Psi(u) - \Psi(u'))^\top (u - u') \geq 0 \quad , \forall u, u' \in \mathcal{D}.$$

- Ψ is μ-*strongly monotone* $(\Psi(u) - \Psi(u'))^\top (u - u') \geq \mu \|u - u'\|^2, \ \mu > 0.$
- Ψ is μ-*hypo/weakly monotone*: the above inequality holds for $\mu < 0$.
- Ψ is β-*cocoercive* $(\Psi(u) - \Psi(u')^\top (u - u') \geq \beta \|\Psi(u) - \Psi(u)'\|^2, \beta > 0.$

Example of operators on the real line:

- (a) strongly monotone
- (b) not monotone
- (c) mere monotone
- (d) cocoercive operator



Monotone Maps and Passivity

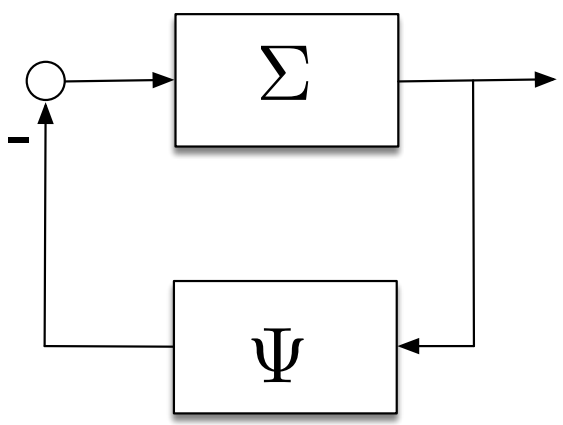
Let $y = \Psi(u)$, $\Psi : \mathcal{D} \subseteq \mathbb{R}^n \to \mathbb{R}^n$

Definition

Map Ψ is *monotone* $\Longleftrightarrow$ EIP with 0 storage fnc.

$$(\Psi(u) - \Psi(u'))^\top (u - u') \geq 0.$$

- Ψ is μ-*strongly monotone* $\Longleftrightarrow$ Ψ is input-strict EIP
  $(\Psi(u) - \Psi(u'))^\top (u - u') \geq \mu \|u - u'\|^2, \ \mu > 0.$
- Ψ is μ-*hypo/weakly monotone*: $\Longleftrightarrow$ Ψ is input EIP-short
  the above inequality holds for $\mu < 0$.
- Ψ is β-*cocoercive* $\Longleftrightarrow$ Ψ is output-strict EIP
  $(\Psi(u) - \Psi(u')^\top (u - u') \geq \beta \|\Psi(u) - \Psi(u)'\|^2, \beta > 0.$



Monotone Maps and Passivity

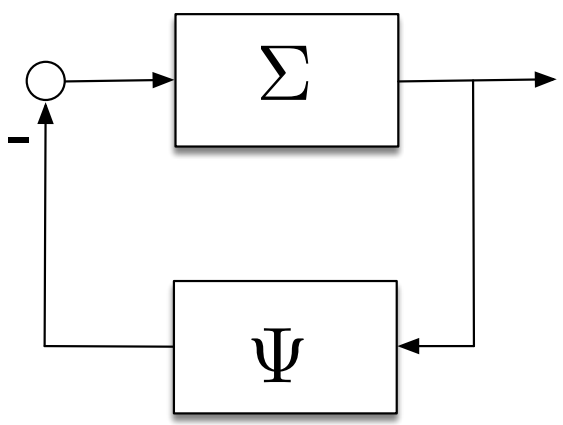
Let $y = \Psi(u)$, $\Psi : \mathcal{D} \subseteq \mathbb{R}^n \to \mathbb{R}^n$

Definition

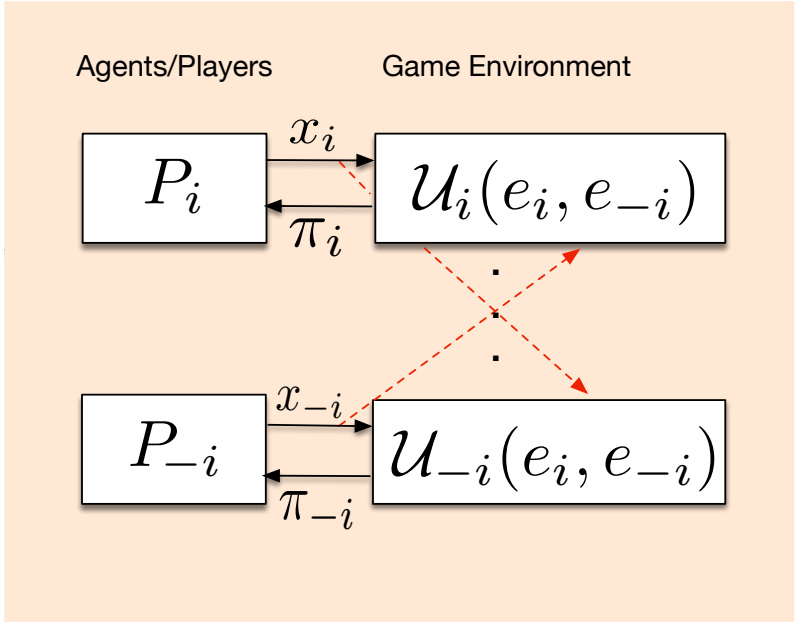
Map Ψ is *monotone* $\Longleftrightarrow$ EIP with 0 storage fnc.

$$(\Psi(u) - \Psi(u'))^\top (u - u') \geq 0.$$

- Ψ is μ-*strongly monotone* $\Longleftrightarrow$ Ψ is input-strict EIP
  $(\Psi(u) - \Psi(u'))^\top (u - u') \geq \mu \|u - u'\|^2, \ \mu > 0.$
- Ψ is μ-*hypo/weakly monotone*: $\Longleftrightarrow$ Ψ is input EIP-short
  the above inequality holds for $\mu < 0$.
- Ψ is β-*cocoercive* $\Longleftrightarrow$ Ψ is output-strict EIP
  $(\Psi(u) - \Psi(u')^\top (u - u') \geq \beta \|\Psi(u) - \Psi(u)'\|^2, \beta > 0.$



Learning: A Generic Formulation



- Player i (red, left) strategy is x_i (everybody else's is x_{-i}), individual payoff $\mathcal{U}_i$, adjusts its decision using dynamics denoted P_i (right).
- Everyone else's algorithm/dynamics is denoted P_{-i}.
- The overall learning dynamics is feedback interconnected $P := (P_i, P_{-i})$.

Two properties are essential:

1. equilibria of P are (related to) NE x^* of the game.
2. asymptotically stable equilibria, i.e., (global) convergence to NE.

Finite-Action Games

Game $\mathcal{G}$. Player $i \in \mathcal{N}$ has

- $\mathcal{A}_i = \{1, \ldots, n_i\}$ finite action set. x_{ij} probability of playing j^{th} action, $j \in \mathcal{A}_i$. $e_j = [0, \ldots 1 \ldots 0]^T$ associated to $j \in \mathcal{A}_i$.
- $x_i = \sum_{j \in \mathcal{A}_i} x_{ij} e_j$, mixed-strategy, $x_i \in \Delta_i := \{x_i \in \mathbb{R}^{n_i}_{\geq 0} | \sum_{j \in \mathcal{A}_i} x_{ij} = 1\}$.
- expected payoff: $\mathcal{U}_i(x_i, x_{-i}) = \sum_{j \in \mathcal{A}_i} \mathcal{U}_i(e_j, x_{-i}) x_{ij} := x_i{}^T U_i(x_{-i})$, with $\mathcal{U}_i(e_j, x_{-i}) = \mathbb{E}[\mathcal{U}_i(e_i, e_{-i}) | e_i = e_j]$ and $U_i(x_{-i}) = (\mathcal{U}_i(e_j, x_{-i}))_{j \in \mathcal{A}_i}$.
- A mixed-strategy profile $x^* = (x_i^*, x_{-i}^*) \in \Delta$ is NE if $\forall i \in \mathcal{N}$

$$\mathcal{U}_i(x_i^*, x_{-i}^*) \geq \mathcal{U}_i(x_i, x_{-i}^*), \ \forall x_i \in \Delta_i, \quad \text{or} \quad x_i^* \in BR_i(x_{-i}^*)$$

$$BR_i(x_{-i}) = \text{argmax}_{x_i \in \Delta_i} x_i^T U_i(x_{-i}).$$

Learning Algorithms

- BR-play, needs x_{-i} mixed strategies of others and expected payoff $U_i(x_{-i})$.
- Fictitious-play uses estimate $\widehat{x}_{-i}$ of x_{-i}, based on history of past actions e_{-i}.
- Payoff-based (Reinforcement Learning) needs only realized payoff $\pi_i = \mathcal{U}_i(e_i, e_{-i})$.

Finite-Action Games

Game $\mathcal{G}$. Player $i \in \mathcal{N}$ has

- $\mathcal{A}_i = \{1, \ldots, n_i\}$ finite action set. x_{ij} probability of playing j^{th} action, $j \in \mathcal{A}_i$. $e_j = [0, \ldots 1 \ldots 0]^T$ associated to $j \in \mathcal{A}_i$.
- $x_i = \sum_{j \in \mathcal{A}_i} x_{ij} e_j$, mixed-strategy, $x_i \in \Delta_i := \{x_i \in \mathbb{R}^{n_i}_{\geq 0} | \sum_{j \in \mathcal{A}_i} x_{ij} = 1\}$.
- expected payoff: $\mathcal{U}_i(x_i, x_{-i}) = \sum_{j \in \mathcal{A}_i} \mathcal{U}_i(e_j, x_{-i}) x_{ij} := x_i{}^T U_i(x_{-i})$, with $\mathcal{U}_i(e_j, x_{-i}) = \mathbb{E}[\mathcal{U}_i(e_i, e_{-i}) | e_i = e_j]$ and $U_i(x_{-i}) = (\mathcal{U}_i(e_j, x_{-i}))_{j \in \mathcal{A}_i}$.
- A mixed-strategy profile $x^* = (x_i^*, x_{-i}^*) \in \Delta$ is NE if $\forall i \in \mathcal{N}$

$$\mathcal{U}_i(x_i^*, x_{-i}^*) \geq \mathcal{U}_i(x_i, x_{-i}^*), \ \forall x_i \in \Delta_i, \quad \text{or} \quad x_i^* \in BR_i(x_{-i}^*)$$

$$BR_i(x_{-i}) = \text{argmax}_{x_i \in \Delta_i} x_i^T U_i(x_{-i}).$$

Learning Algorithms

- BR-play, needs x_{-i} mixed strategies of others and expected payoff $U_i(x_{-i})$.
- Fictitious-play uses estimate $\widehat{x}_{-i}$ of x_{-i}, based on history of past actions e_{-i}.
- Payoff-based (Reinforcement Learning) needs only realized payoff $\pi_i = \mathcal{U}_i(e_i, e_{-i})$.

Softmax map σ_ϵ and Nash Distribution

- Softmax map $\sigma_\epsilon : \mathbb{R}^{n_i} \to \Delta_i$, defined by

$$[\sigma_\epsilon(z_i)]_j = \frac{\exp(\epsilon^{-1} z_{ij})}{\sum_{j' \in \mathcal{A}_i} \exp(\epsilon^{-1} z_{ij'})} \quad \forall j \in \mathcal{A}_i,$$

  where $\epsilon > 0$ is temperature parameter.

- Softmax is generated via regularization from BR as

$$[\sigma_\epsilon(z_i)] = \text{argmax}_{x_i \in \Delta_i} \left\{ x_i^T z_i - \epsilon x_i^T \mathbf{log}(x_i) \right\},$$

  as a perturbed (smooth) BR.

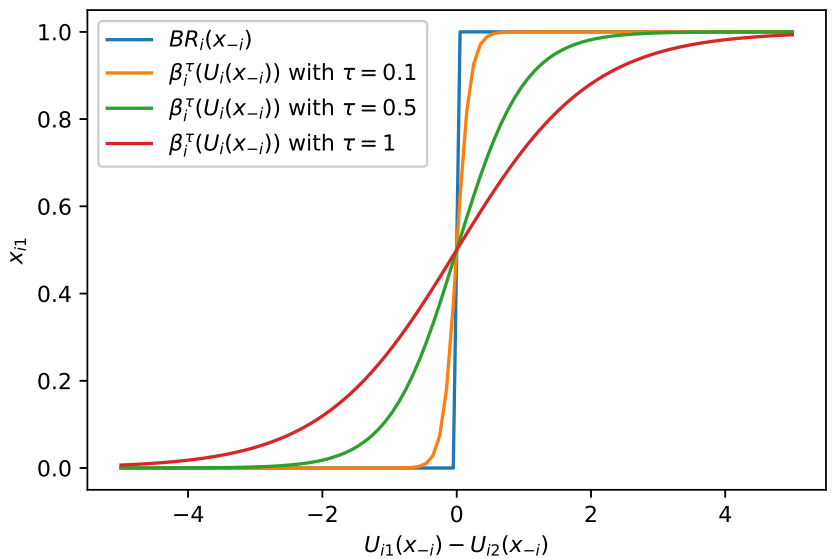
- **Nash distribution** (Logit equilibrium),

$$x_i^\epsilon = \sigma_\epsilon(U_i(x_{-i}^\epsilon)) = \widetilde{BR}_i(x_{-i}^\epsilon), \forall i \in \mathcal{N},$$

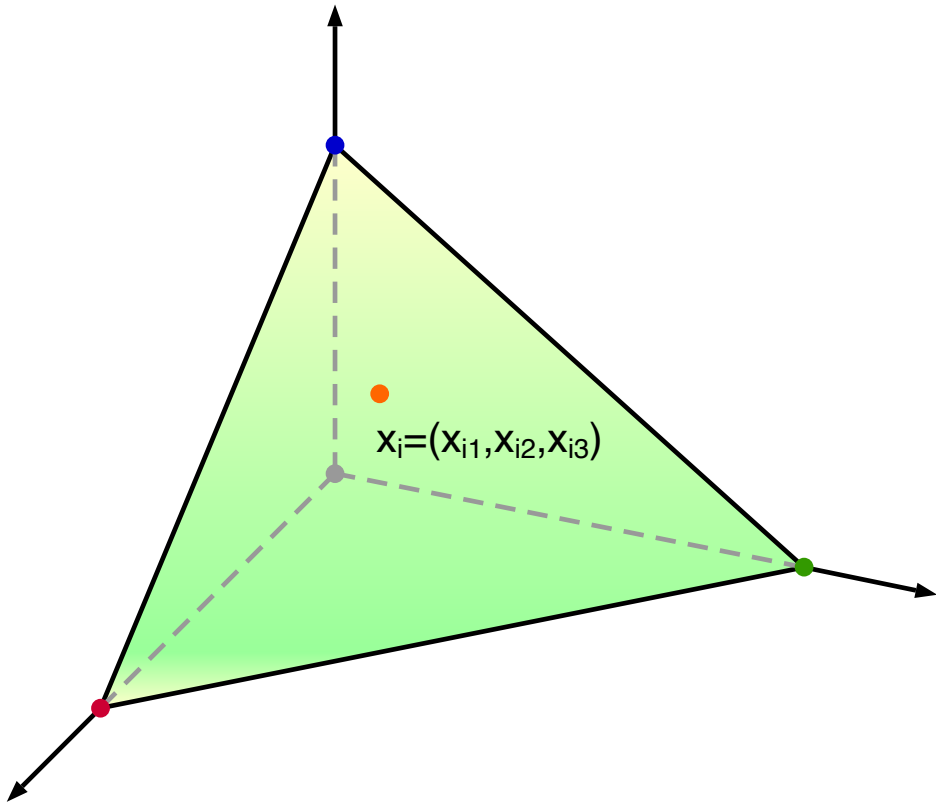
  is NE of game with the perturbed payoff,

$$\widetilde{\mathcal{U}}_i(x_i, x_{-i}) = \mathcal{U}_i(x_i, x_{-i}) - \epsilon\, x_i^T \mathbf{log}(x_i).$$

- For $\epsilon \to 0$, $x^\epsilon \to x^*$.



Softmax map σ_ϵ



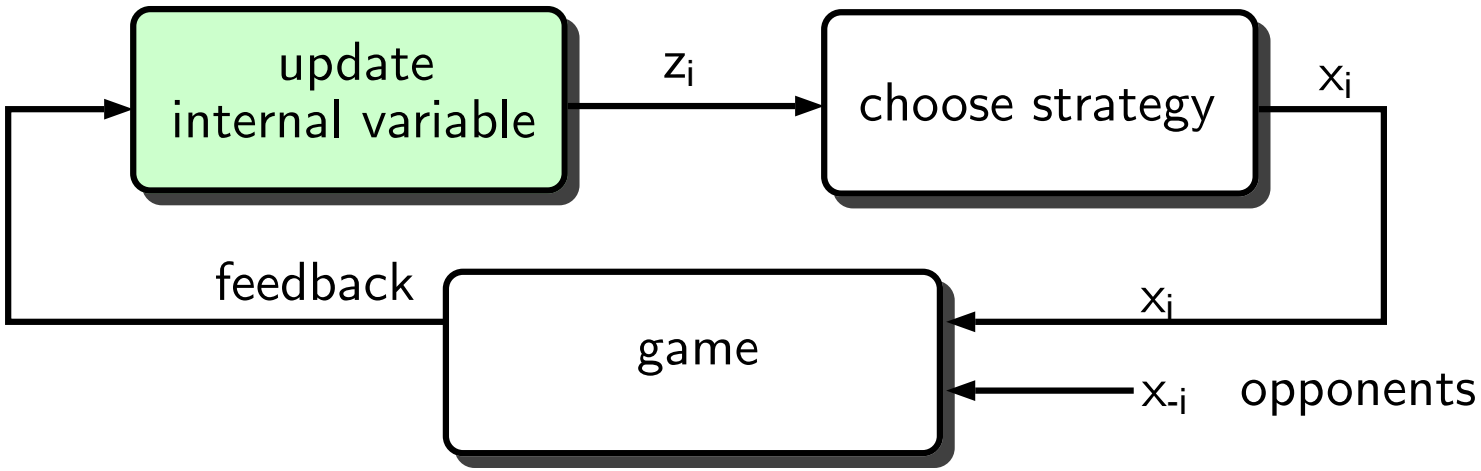
Generates a vector inside of the probability simplex Δ_i

$$\Delta_i = \{x_i \in \mathbb{R}^n_{\geq 0} | \sum_{j=1}^{n_i} x_{ij} = 1\}.$$

e.g., $x_i = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}), (\frac{1}{2}, \frac{1}{3}, \frac{1}{6}), (\frac{1}{2}, 0, \frac{1}{2})$

Suitable for matrix games.

How Players Learn



A group of players learn by individually:

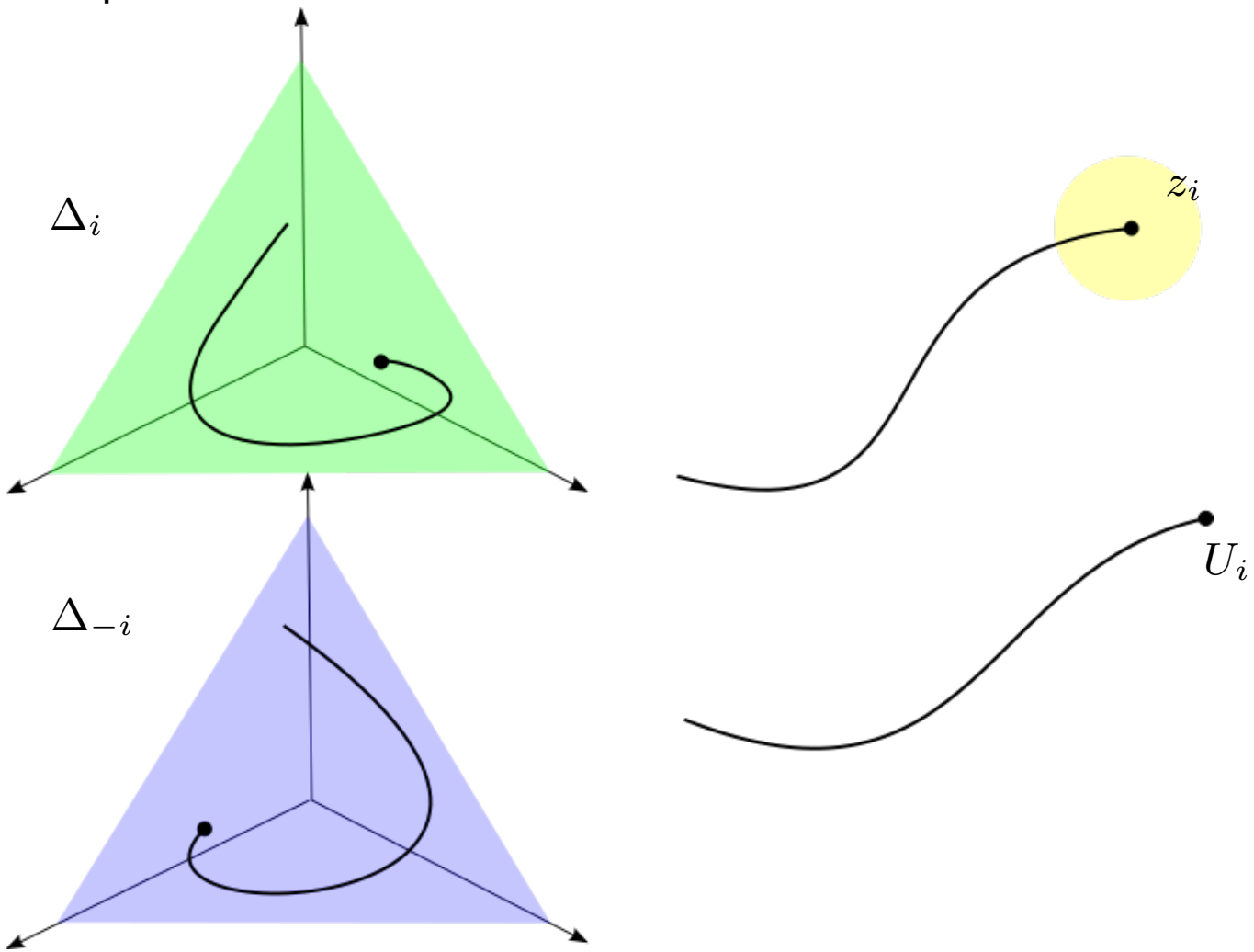
- choosing strategies via an internal variable[1]
- receiving information (feedback) from the game
- updating own internal variable

E.g. mirror descent, dual averaging, follow-the-regularized-leader ...

[1] score, perception, dual aggregate, Q-value, model weights, etc.

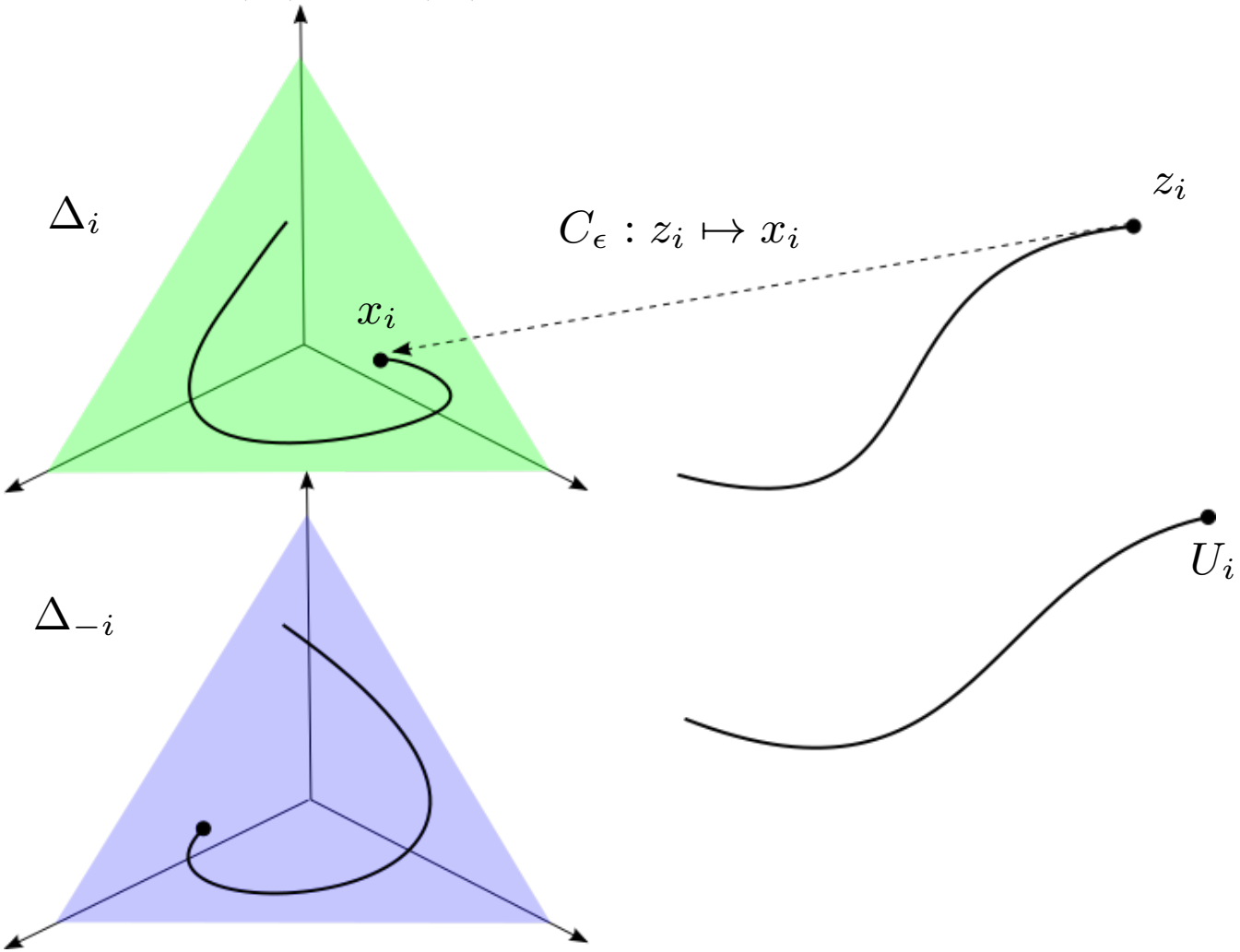
How Players Learn: A Continuous Learning Scheme

Player i keeps an *internal variable* $z_i \in \mathbb{R}^{n_i}$



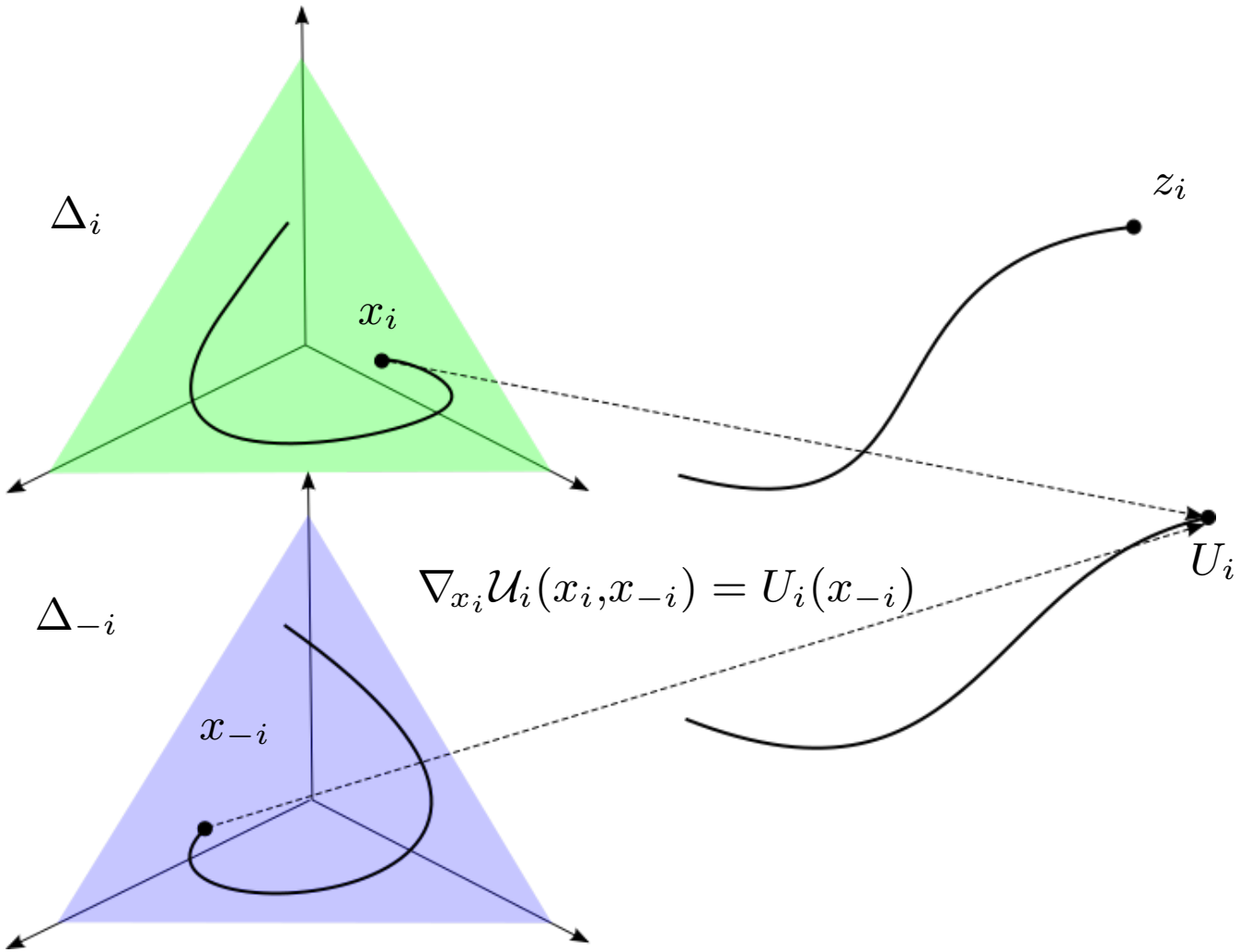
How Players Learn: A Continuous Learning Scheme

Player i maps own variable z_i into a strategy $x_i \in \Delta_i$ through a *mirror (choice) map* $C_\epsilon : \mathbb{R}^{n_i} \to \Delta_i$, $C_\epsilon(z_i) = \sigma_\epsilon(z_i)$ softmax, which selects its next strategy



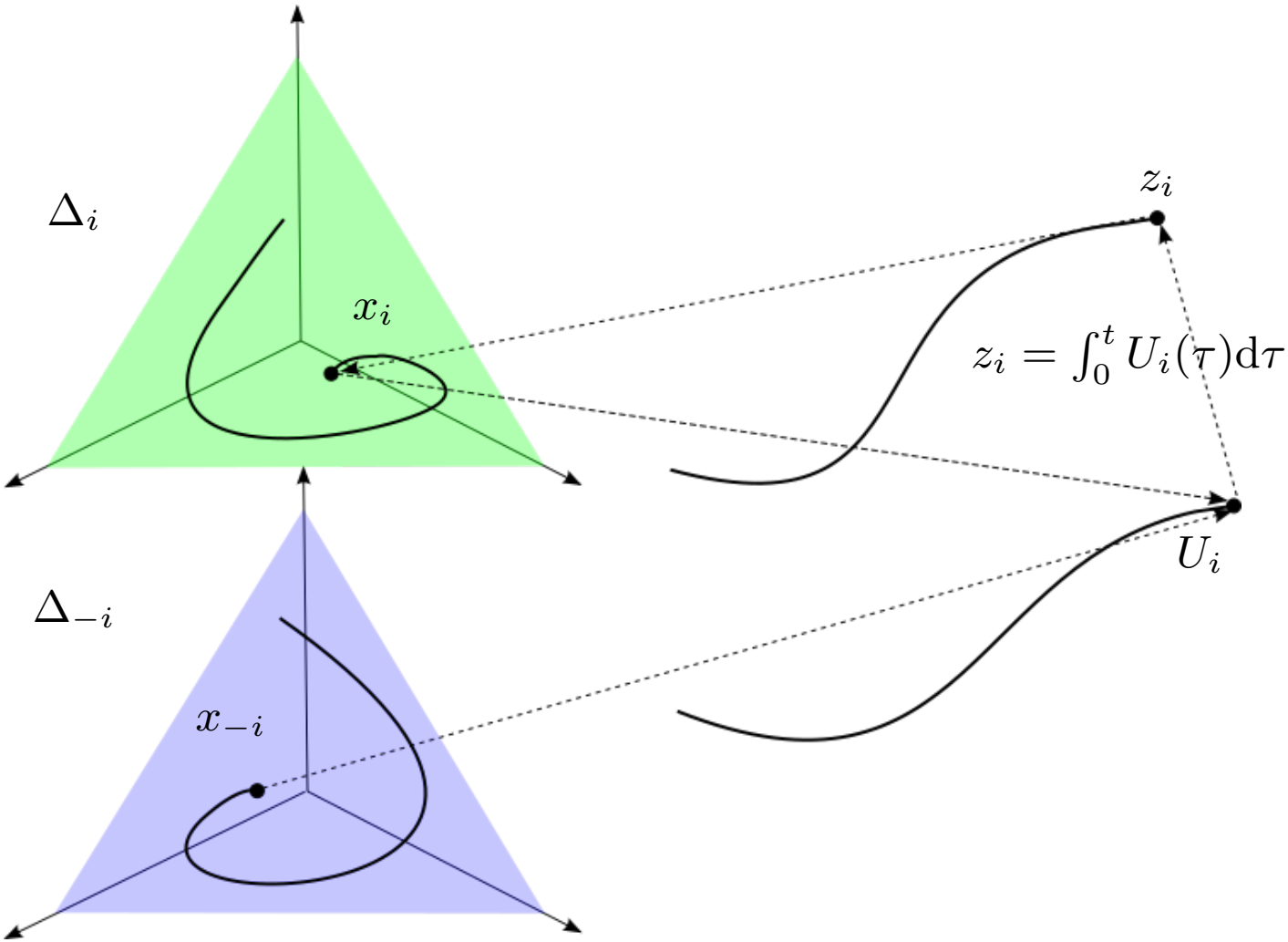
How Players Learn: A Continuous Learning Scheme

Player i plays the game using strategy x_i (opponents play $x_{-i} \in \Delta_{-i}$) and obtains its expected payoff vector $U_i(x_{-i}) = (\mathcal{U}_i(e_j, x_{-i}))_{j \in \mathcal{A}_i} = \nabla_{x_i} \mathcal{U}_i(x_i, x_{-i})$.



How Players Learn: A Continuous Learning Scheme

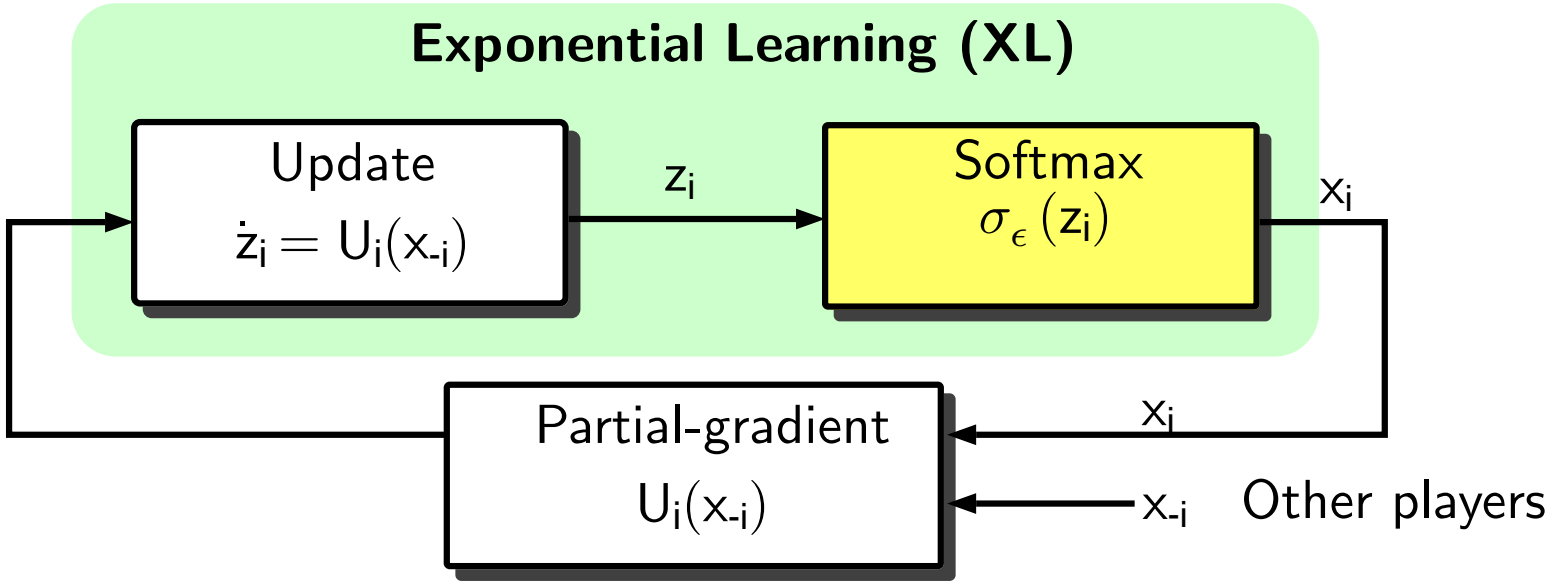
Player i maps U_i back into z_i via aggregation $z_i = \int_0^t U_i(\tau)\mathrm{d}\tau$ and the process repeats indefinitely with an infinitesimal time-step between each stage.



This process called **Exponential Learning (XL)** a form of mirror descent (MD).

Exponential Learning (XL)

- Feedback perspective of Player i



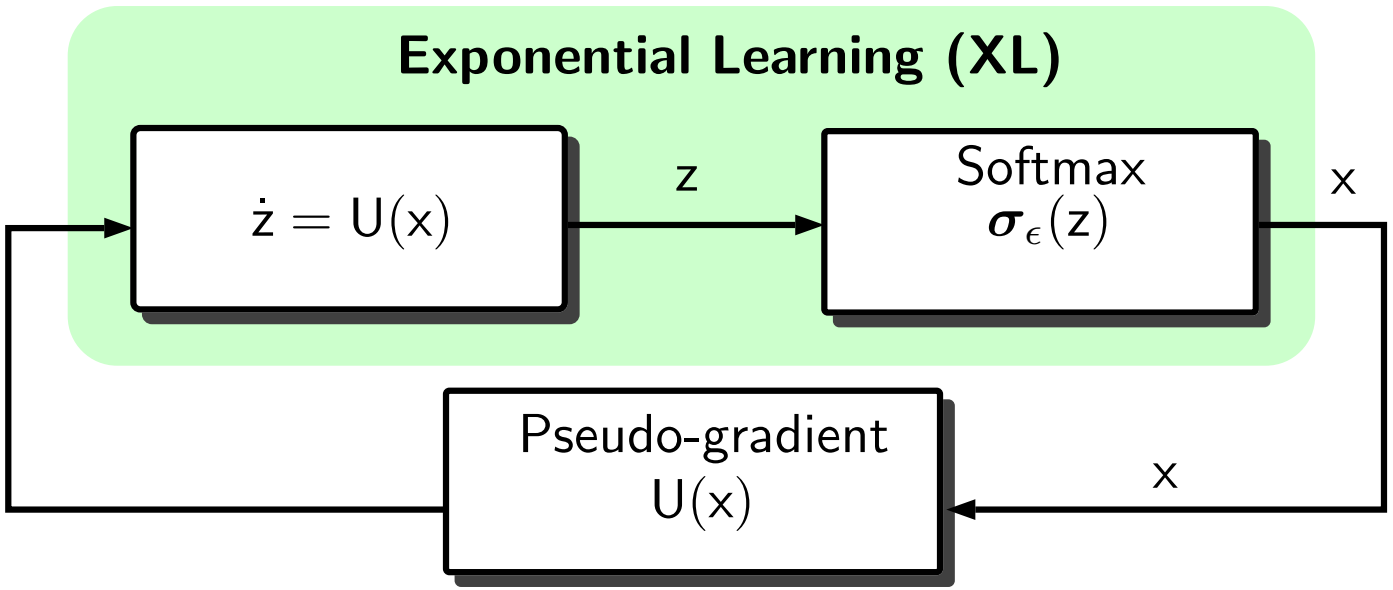
- Player i dynamics P_i is an ODE (first-order dynamics) in dual space

$$\text{XL} \quad P_i : \begin{cases} \dot{z}_i = U_i(x_{-i}) \\ x_i = \sigma_\epsilon(z_i) \end{cases} \qquad \text{where} \qquad \sigma_\epsilon(z_i) = \frac{(\exp(\epsilon^{-1} z_{ij}))_{j \in \mathcal{A}_i}}{\sum_{j \in \mathcal{A}_i} \exp(\epsilon^{-1} z_{ij})}$$

- XL converges to a strict NE.
- XL fails to converge in zero-sum games with interior NE, e.g. RPS game.
- ODE induced for x_i is Replicator Dynamics (RD) in evolutionary game theory. [Mertikopoulos & Sandholm'16]

Exponential Learning (XL)

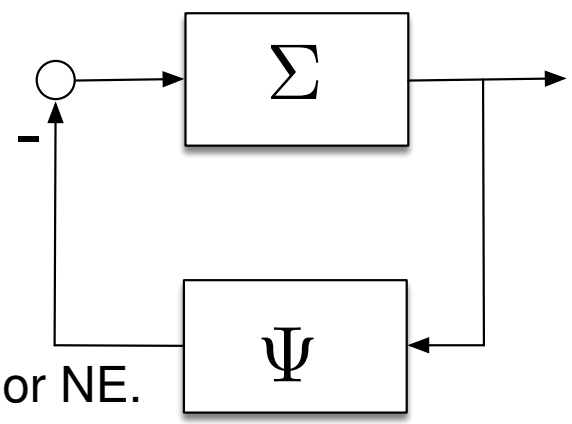
- Stack for all players $i \in \mathcal{N} = \{1, \dots, N\}$. Let $z = (z_i)_{i \in \mathcal{N}}$, $x = (x_i)_{i \in \mathcal{N}}$, $\boldsymbol{\sigma}_\epsilon = (\sigma_\epsilon)_{i \in \mathcal{N}}$, $U(x) = (U_i(x_{-i}))_{i \in \mathcal{N}}$, called a **pseudo-gradient**.



- Overall dynamics is $P = (P_i, P_{-i})$

$$\text{XL} \quad P : \begin{cases} \dot{z} = U(x) \\ x = \boldsymbol{\sigma}_\epsilon(z) \end{cases}$$

as feedback interconnection (Σ, Ψ).



- At rest, $0 = U(x) \Longrightarrow$ equilibrium is x^* an interior NE.

Two-player Rock-Paper-Scissor (RPS) game

- Player 1 and Player 2 payoff matrices

$$(A, B) = \begin{bmatrix} (0,0) & (-l,1) & (1,-l) \\ (1,-l) & (0,0) & (-l,1) \\ (-l,1) & (1,-l) & (0,0) \end{bmatrix} \quad l: \text{ loss } > 0,$$

where $A = \begin{bmatrix} 0 & -l & 1 \\ 1 & 0 & -l \\ -l & 1 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 & -l \\ -l & 0 & 1 \\ 1 & -l & 0 \end{bmatrix}$. $B = A^\top$.

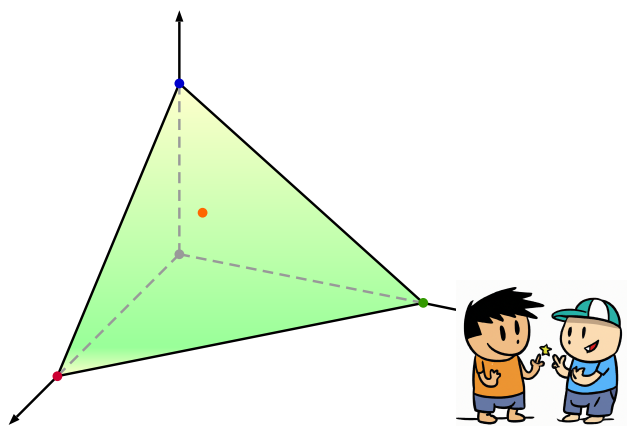
- Expected Payoffs: $\mathcal{U}_1(x_1, x_2) = {x_1}^\top A x_2, \quad \mathcal{U}_2(x_1, x_2) = {x_1}^\top B x_2.$
- Partial gradients:

$$\nabla_{x_1} \mathcal{U}_1(x_1, x_2) = A x_2 := U_1(x_2),$$
$$\nabla_{x_2} \mathcal{U}_2(x_1, x_2) = B^\top x_1 := U_2(x_1).$$

- Pseudo-gradient:

$$U(x) := \begin{bmatrix} U_1(x_2) \\ U_2(x_1) \end{bmatrix} = \Phi x,$$

where $\Phi = \begin{bmatrix} 0 & A \\ B^\top & 0 \end{bmatrix}$.



Rock-Paper-Scissor (RPS) game

The NE of the RPS game is,

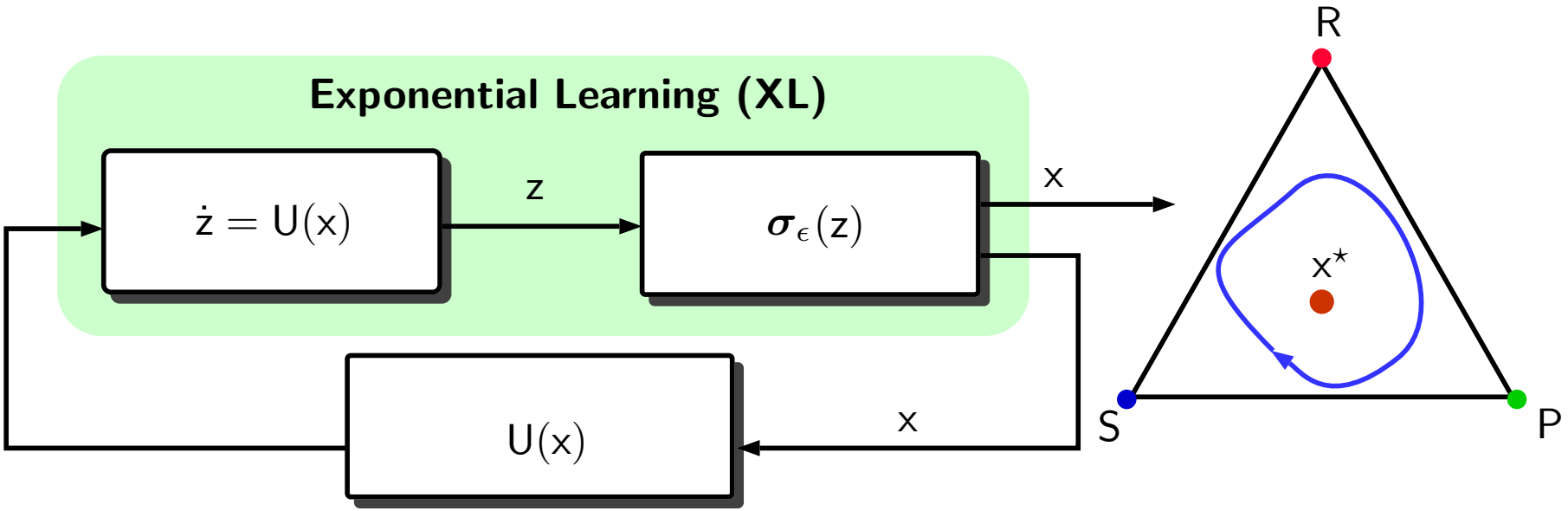
$$x^* = (x_1^*, x_2^*) = \left(\begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} \right). \tag{3}$$

Interior (mixed NE). Each player chooses actions with equal likelihood.

Does XL converge to the NE x^* for RPS game?

Standard Zero-sum RPS game

$$A = \begin{bmatrix} 0 & -l & 1 \\ 1 & 0 & -l \\ -l & 1 & 0 \end{bmatrix} \qquad l = 1$$



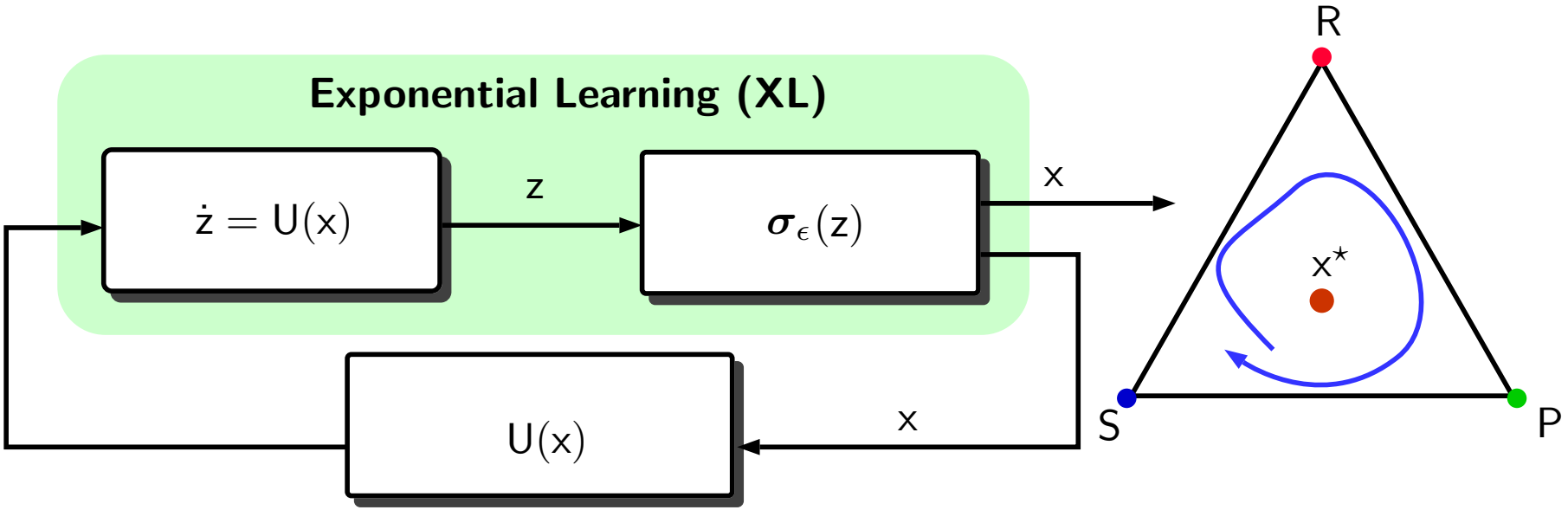
- This is an example of a **merely monotone** game:

$$(U(x)-U(x'))^\top(x-x') \leq 0, \forall x, x' \in \Delta. \tag{4}$$

- Mapping U is linear, $U(x) = \Phi x$, (4) $\Leftrightarrow$ $\frac{1}{2}y^\top(\Phi + \Phi^\top)y \leq 0$, $\forall y \in \mathsf{T}_\Delta$.
- Mapping $-U$ is (null) monotone. $\Phi = \begin{bmatrix} 0 & A \\ -A & 0 \end{bmatrix}$, $\Phi + \Phi^\top = 0$.

General RPS game

$$A = \begin{bmatrix} 0 & -l & 1 \\ 1 & 0 & -l \\ -l & 1 & 0 \end{bmatrix} \qquad l \neq 1$$



- This is an example of a μ-**weakly/hypo monotone** game:

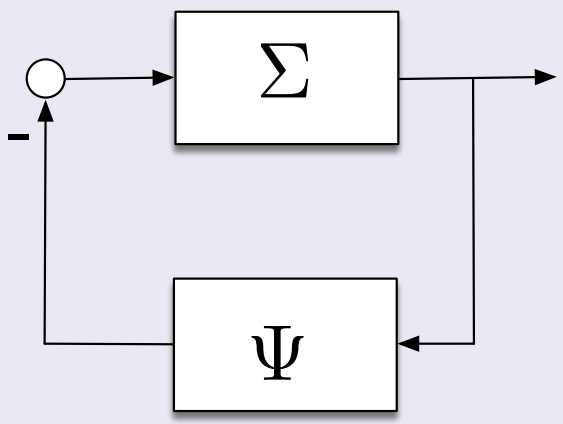
$$(U(x)-U(x'))^\top(x-x') \leq \mu\|x-x'\|_2^2, \forall x, x' \in \Delta. \tag{5}$$

- Mapping U is linear, $U(x) = \Phi x$, **(5)** $\Leftrightarrow$ $\frac{1}{2}y^\top(\Phi+\Phi^\top)y \leq \mu\|y\|^2$, $\forall y \in \mathsf{T}_\Delta$, $\mu = \frac{1}{2}\lambda_{\max}(\Phi+\Phi^\top)|_{\mathsf{T}_\Delta}$.
- Mapping $-U$ is weakly/hypo monotone. $\mu = \frac{1}{2}|l-1|$
  Higher $\mu \geq 0$, “harder" the game, “worse" the behavior.

Overcoming non-convergence of XL

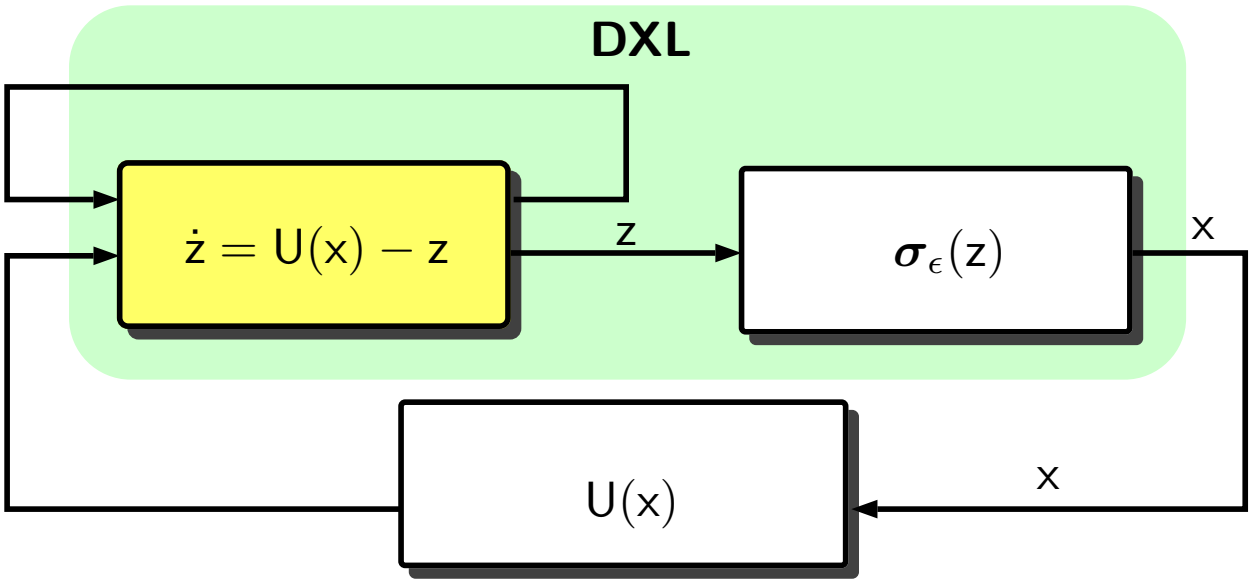
Observation

- (Weak) monotonicity of game = (lack of) energy dissipation, **passivity short** feedback map $-U = \Psi$.



- Idea: add **passivity excess** to the forward system.

Discounted exponential learning (DXL)

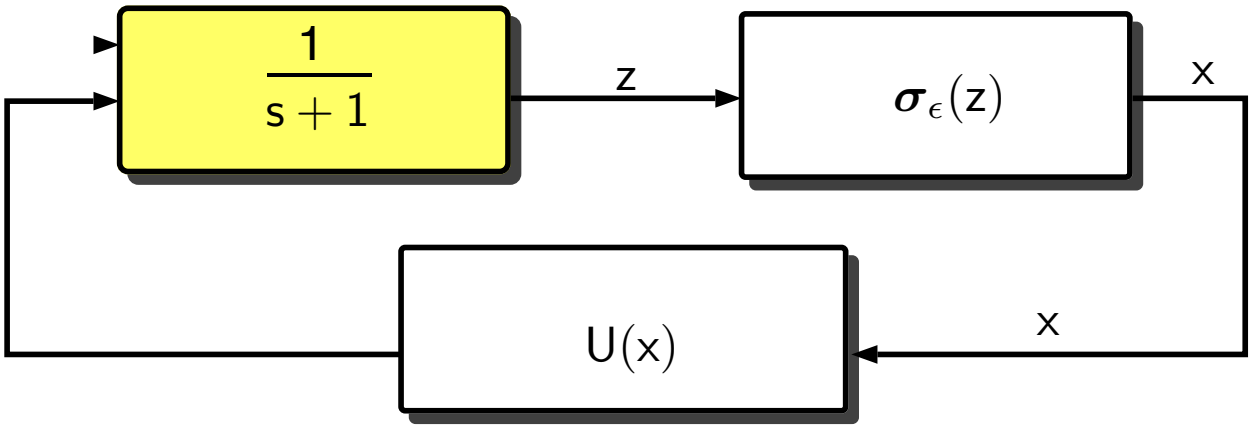


Idea

- Let players re-evaluate internal variable z during play.
- Updates scores based on exponentially-discounted aggregation of payoff.

$$\text{DXL} \quad P: \begin{cases} \dot{z} = U(x) - z \\ x = \boldsymbol{\sigma}_\epsilon(z) \end{cases} \quad \Rightarrow \quad \begin{matrix} \text{At Rest} \\ \begin{cases} \overline{z} = U(\overline{x}) \\ \overline{x} = \boldsymbol{\sigma}_\epsilon(\overline{z}), \quad \overline{x} = x^\epsilon \text{(logit eq.)} \end{cases} \end{matrix}$$

Discounted exponential learning (DXL)



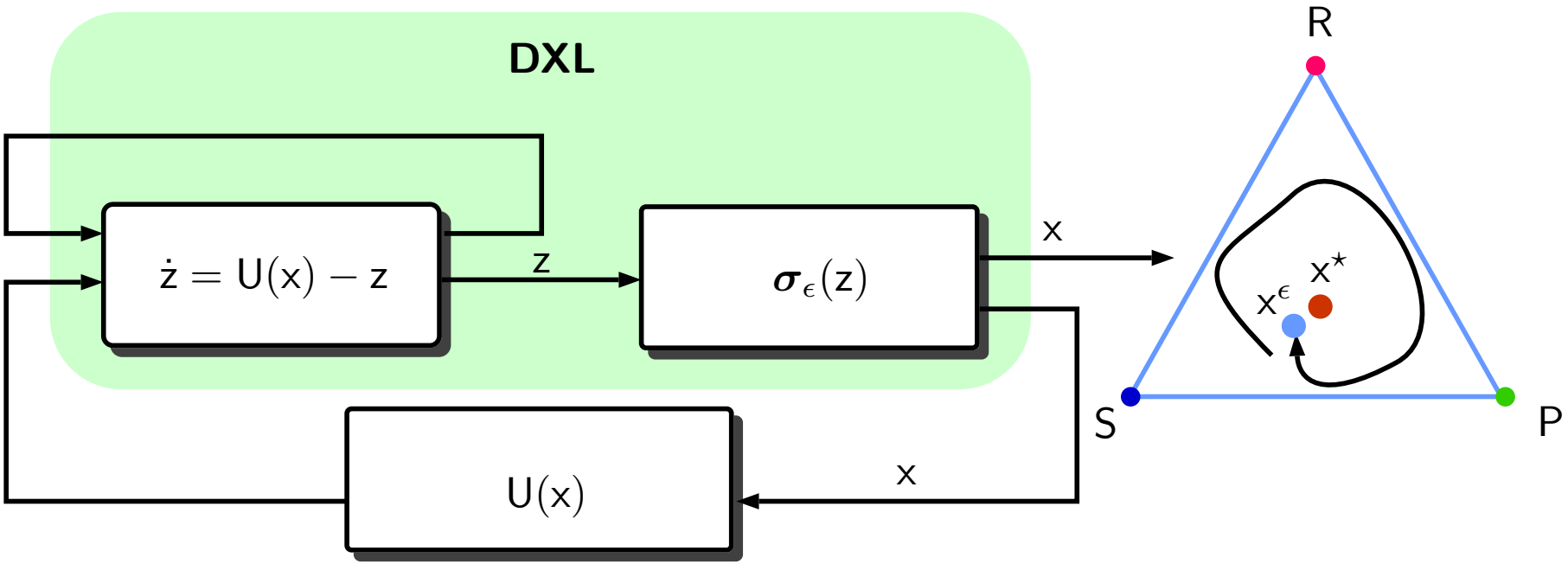
Forward system is output strict passive

- σ_ϵ is the gradient of $\mathsf{lse}_i(z_i) = \epsilon \log(\sum_{j\in\mathcal{A}_i} \exp(\epsilon^{-1} z_{ij}))$, $\nabla \mathsf{lse} = \sigma_\epsilon$.
- $\boldsymbol{\sigma}_\epsilon$ is ϵ-cocoercive, $(\boldsymbol{\sigma}_\epsilon(z) - \boldsymbol{\sigma}_\epsilon(z'))^\top (z - z^\top) \geq \epsilon \|\boldsymbol{\sigma}_\epsilon(z) - \boldsymbol{\sigma}_\epsilon(z')\|_2^2$
- $\exists$ storage func. $V_{\bar{z}}$ s.t. $\dot{V}_{\bar{z}} \leq (x - \overline{x})^\top (U(x) - U(\overline{x})) - \epsilon \|x - \overline{x}\|_2^2$, where $V_{\bar{z}}$ is Bregman divergence of lse,

$$V_{\bar{z}}(z) = \mathbf{1}^\top (\mathsf{lse}(z) - \mathsf{lse}(\bar{z})) - \nabla^\top \mathsf{lse}(\bar{z})(z - \bar{z})$$

[Gao & Pavel, TAC'21a].

Convergence of DXL

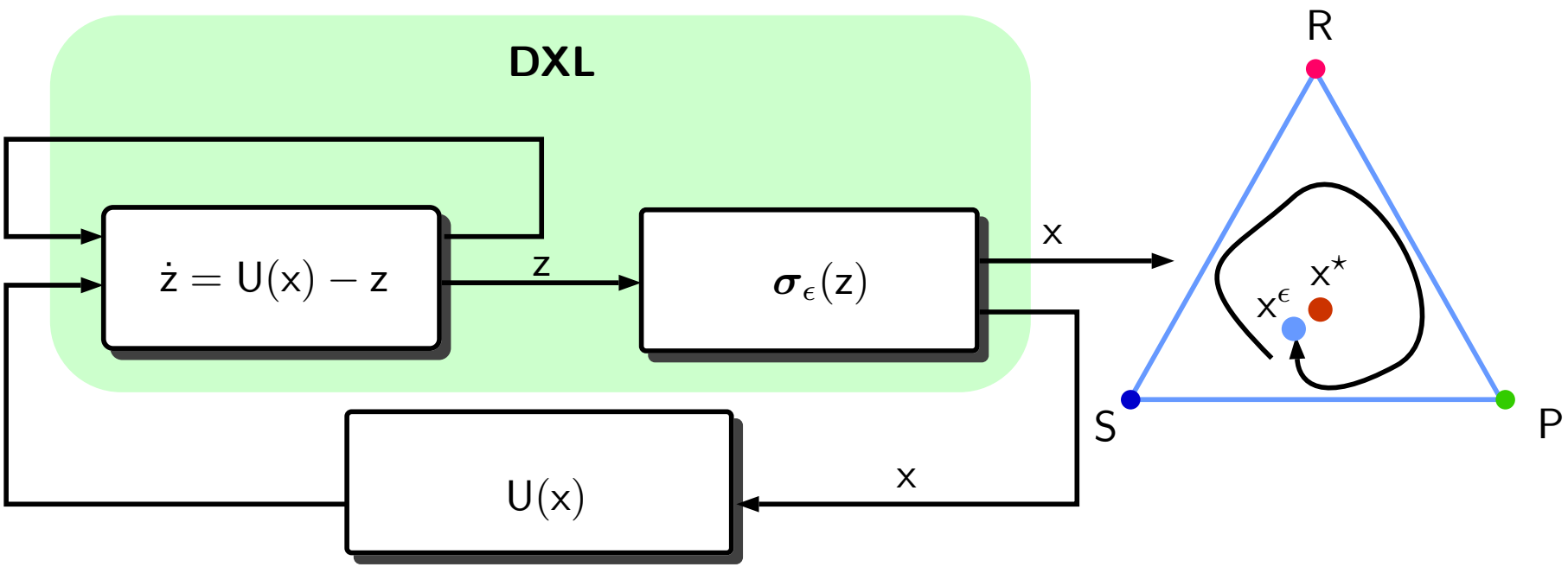


Theorem

$\mathcal{G}$ merely monotone, $x(t) \to x^\epsilon \approx$ NE x^* (x^ϵ **logit equilibrium**).

- Intuition: balancing of passivity.
- Note: temperature $\epsilon > 0$ of $\boldsymbol{\sigma}_\epsilon$ controls **rationality**, $\epsilon \to 0$, $x^\epsilon \to x^*$.

Convergence of DXL

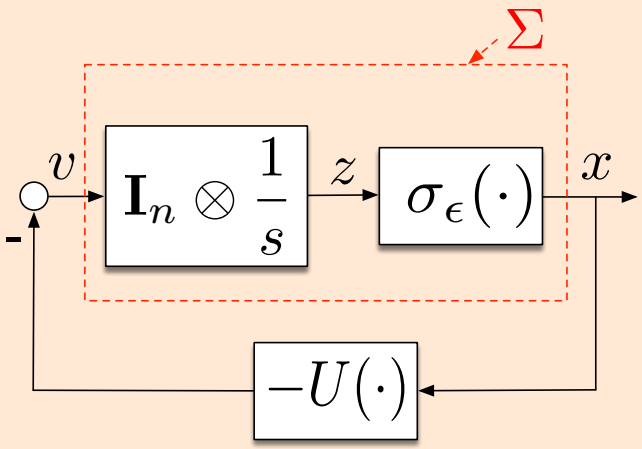


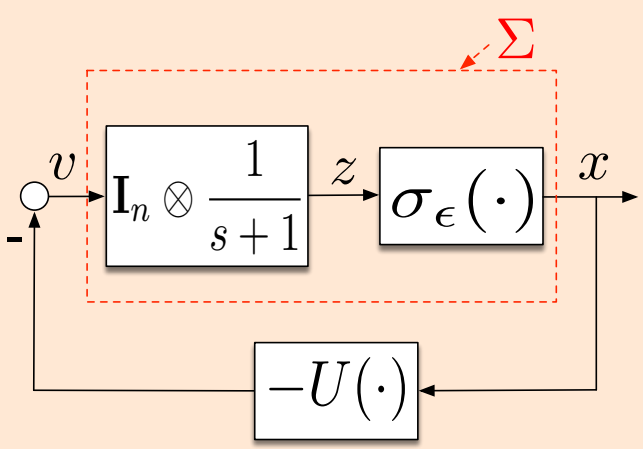
Theorem

$\mathcal{G}$ μ-weakly monotone, same conclusions hold for $\forall \epsilon > \mu$.

- Balance lack of passivity μ by excess passivity (ϵ-cocoercivity).
- Convergence to $x^\epsilon \neq x^*$.
- feedback term $-z$ is non-vanishing at rest, perturbing the solution away from x^*.

Snapshot: XL versus Discounted XL





XL $P: \begin{cases} \dot{z} = U(x) \\ x = \boldsymbol{\sigma}_\epsilon(z) \end{cases}$

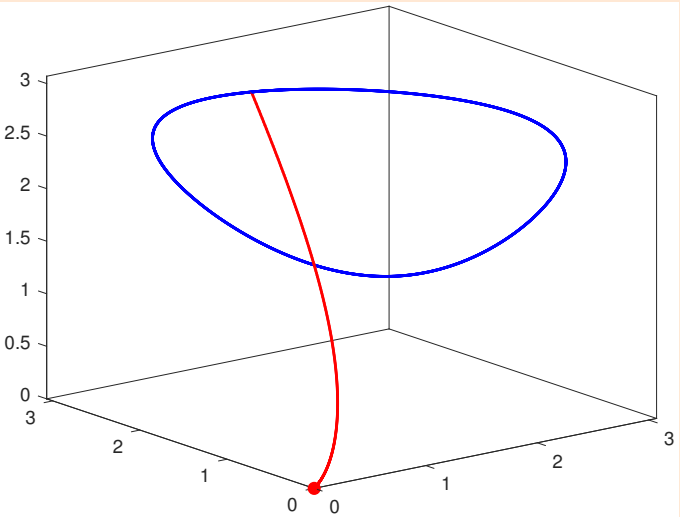
DXL $P: \begin{cases} \dot{z} = U(x) - z \\ x = \boldsymbol{\sigma}_\epsilon(z) \end{cases}$

Observation

- Σ in XL is EIP (integrator) versus Σ in DXL output strict EIP (low-pass filter)
- For $-U$ monotone (EIP), XL is merely stable, while DXL is asy. stable $\forall \epsilon > 0$.
- DXL convergens to $x^\epsilon \to x^*$, $\epsilon \to 0$.

[Gao & Pavel, TAC'21a]

Standard Zero-sum RPS game



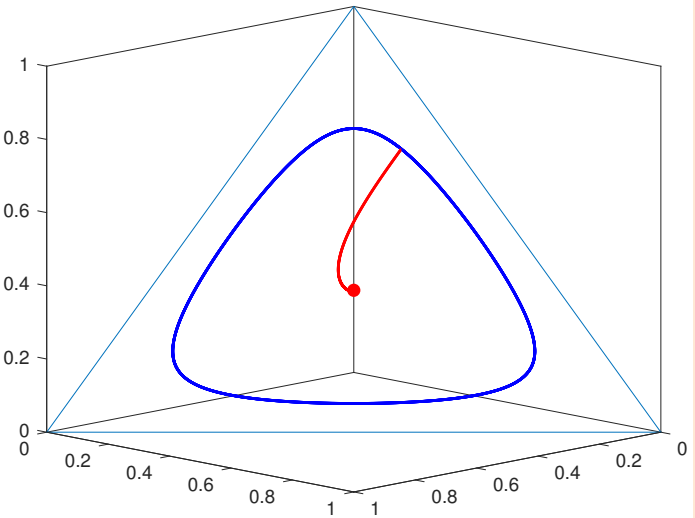
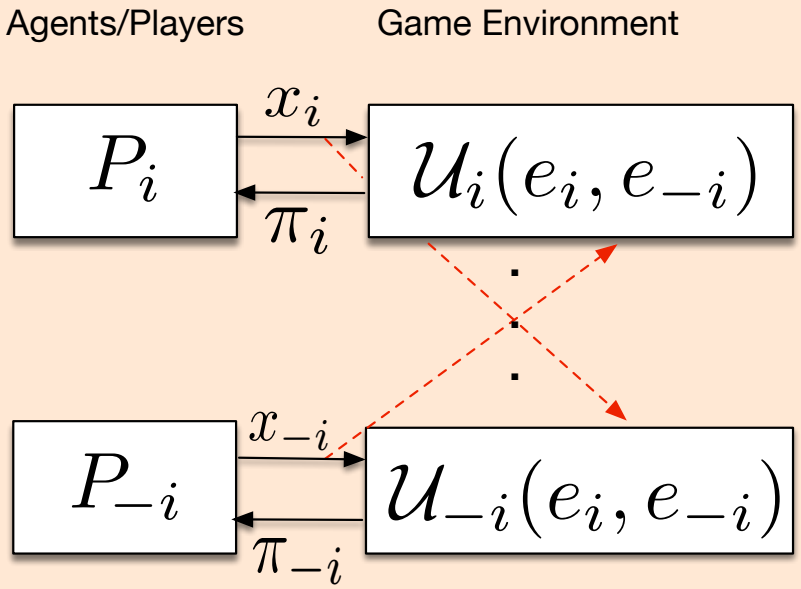


Figure 1: Score trajectories (left), and (strategy trajectories (right) for player 1
XL dynamics (blue) and DXL dynamics (red).
XL dynamics cycles, DXL converges.

$$A = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix}, B = A^T.$$

- Unique Nash distribution $x^\epsilon = x^*$ the Nash equilibrium.
- U is (null) monotone.

Reinforcement Learning



Feedback is π_i own realized payoff.

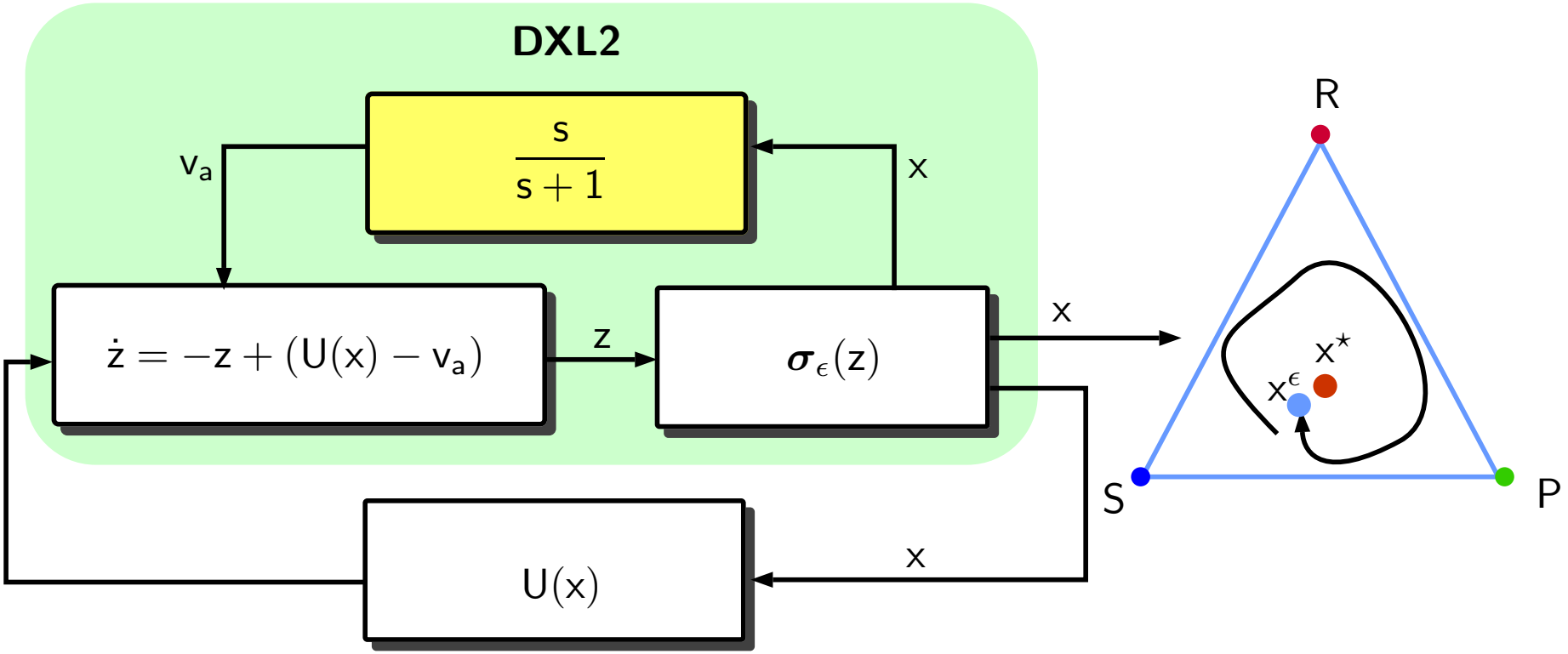
Two popular RL schemes

- Payoff-learning (P-RL) [Coucheney & al'15] $\Rightarrow$ XL.
- Individual Q-learning (Q-L) [Leslie & Collins'05] $\Rightarrow$ DXL.
  by using stochastic approximation method [Benaim'99].
- Passivity framework explains (non)convergence in two-player zero-sum games with mixed NE.

Higher-order dynamics

- What happens if players used some additional processing ?

Higher-order dynamics: DXL2

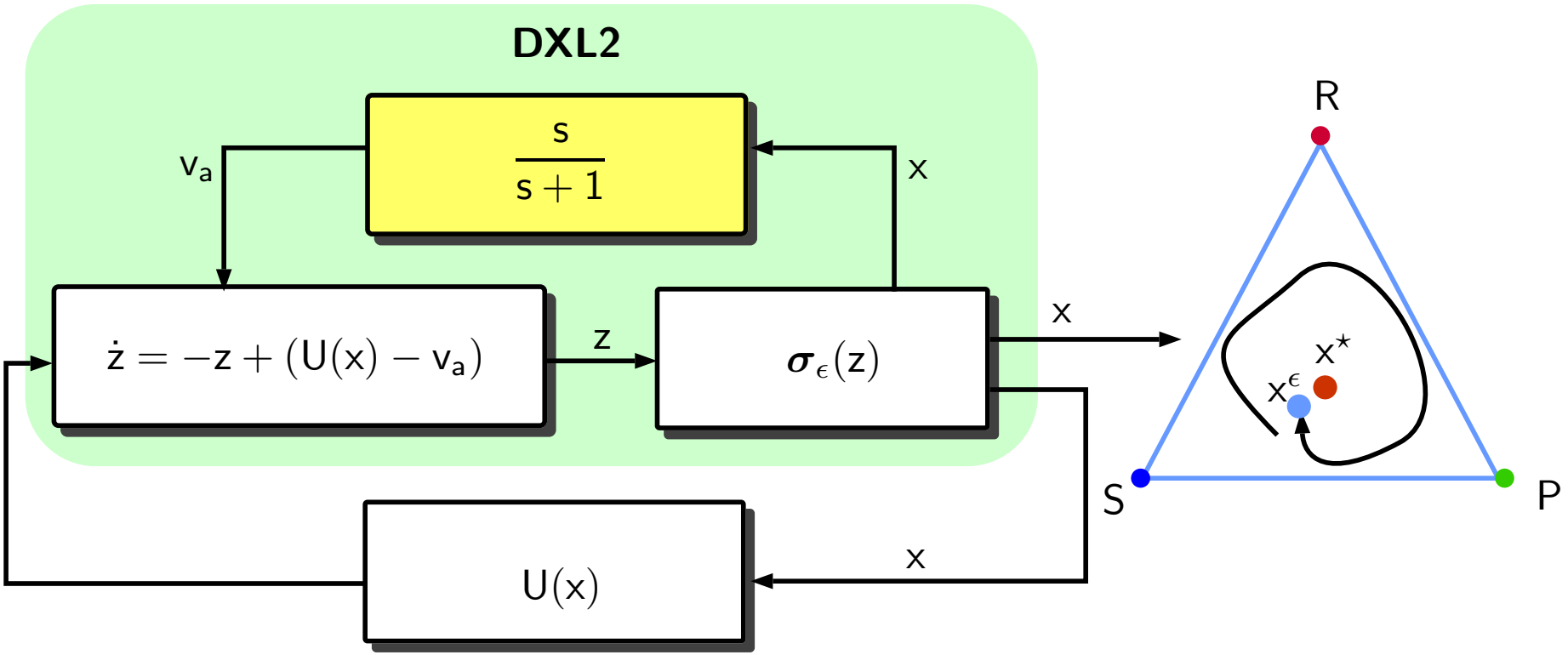


Observation

Negative feedback interconnection preserves passivity.

- Add a strictly passive HPF in feedback. Call this **second-order DXL** (DXL2).
- ODE: $\begin{cases} \dot{z} = U(x) - z - \dot{y} \\ \dot{y} = x - y \end{cases}$, $x = \boldsymbol{\sigma}_\epsilon(z) \quad \Longrightarrow$ Equilibrium at x^ϵ (logit eq.)
- Intuition interpretation: add exponential weighting of the strategies.

Higher-order dynamics: DXL2



Observation

DXL2 converges to x^ϵ under same conditions, but empirically more robust against game parameter perturbations.

Modified RPS Games

[Benaim, Hofbauer, & Hopkins'09]

Game 1: $A = \begin{bmatrix} 0 & -1 & 3 \\ 2 & 0 & -1 \\ -1 & 3 & 0 \end{bmatrix}$,

strictly monotone (stable)

Game 2: $A = \begin{bmatrix} 0 & -3 & 1 \\ 1 & 0 & -2 \\ -3 & 1 & 0 \end{bmatrix}$

0.5-hypo (weakly) monotone (unstable)

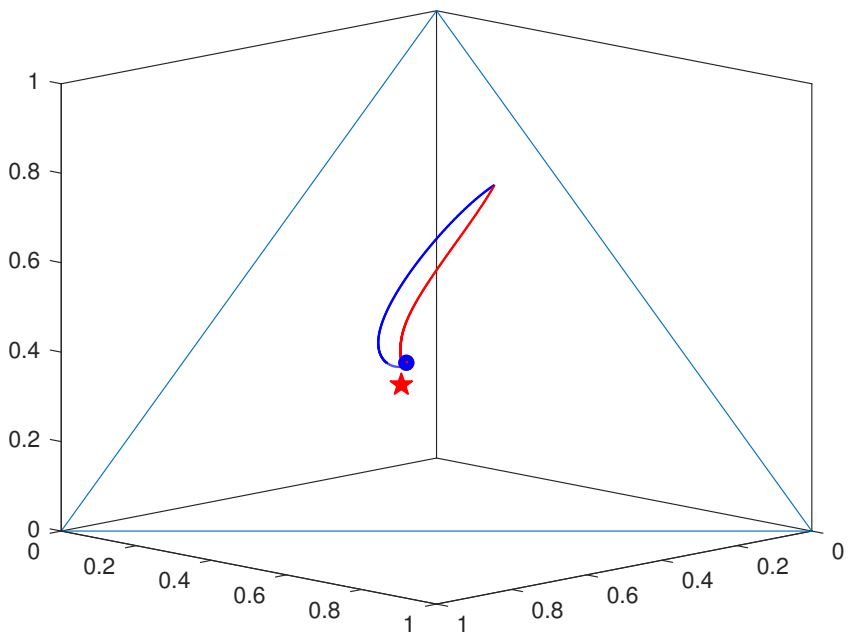
For both games $x^\epsilon \neq x^*$.

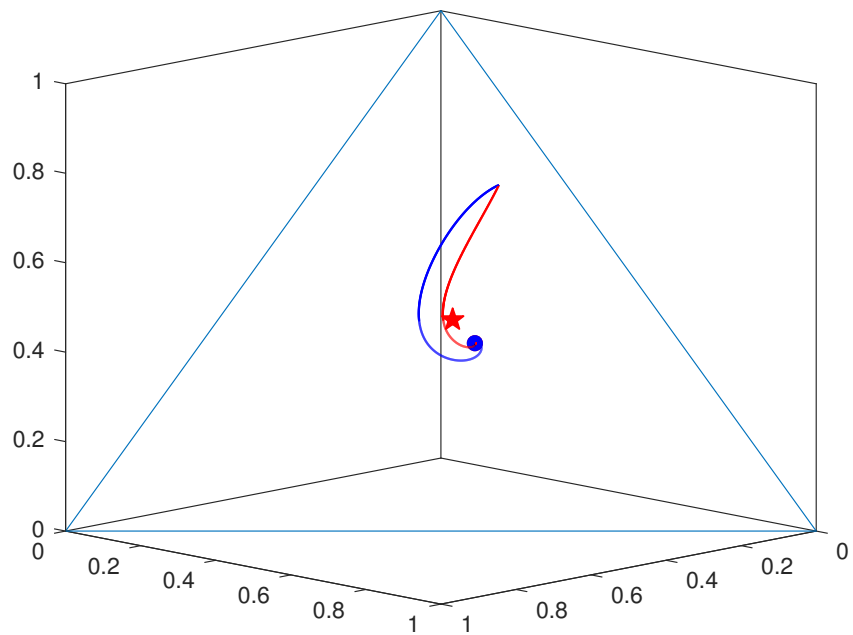
Sufficient condition: $\forall \epsilon > 0$

Sufficient condition: $\forall \epsilon > 0.5$

DXL and DXL2 converge to $x^\epsilon \neq x^*$.

Example 6: $\epsilon = 1$ Both DXL (—) and DXL2 (—) converge in both games





Modified RPS Games

[Benaim, Hofbauer, & Hopkins'09]

Game 1: $A = \begin{bmatrix} 0 & -1 & 3 \\ 2 & 0 & -1 \\ -1 & 3 & 0 \end{bmatrix}$,

strictly monotone (stable)

Game 2: $A = \begin{bmatrix} 0 & -3 & 1 \\ 1 & 0 & -2 \\ -3 & 1 & 0 \end{bmatrix}$

0.5-hypo (weakly) monotone (unstable)

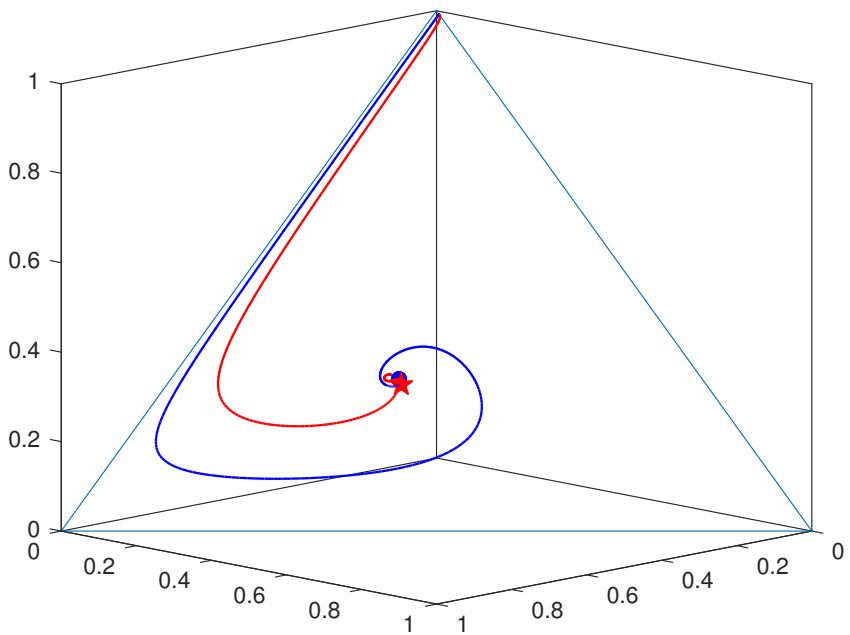
For both games $x^\epsilon \neq x^*$.

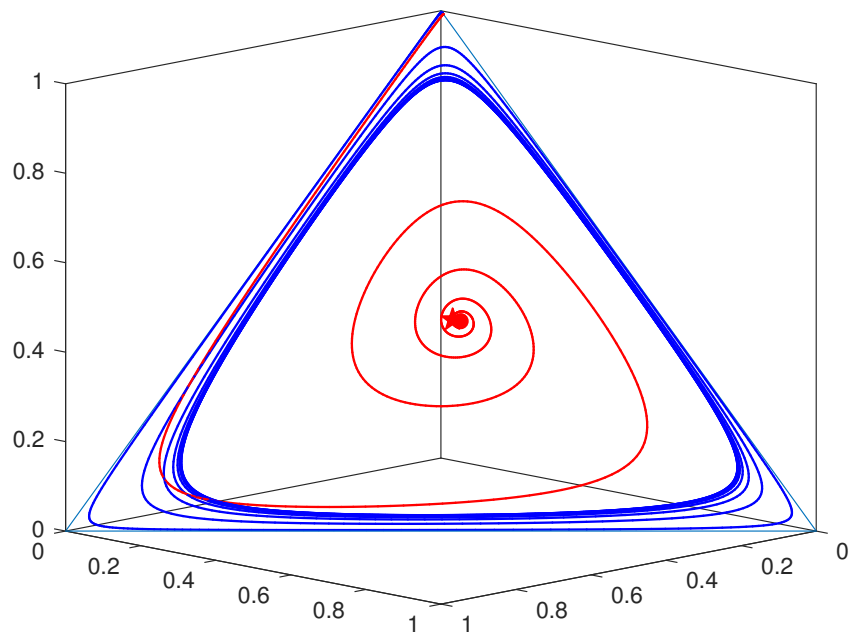
Sufficient condition: $\forall \epsilon > 0$

Sufficient condition: $\forall \epsilon > 0.5$

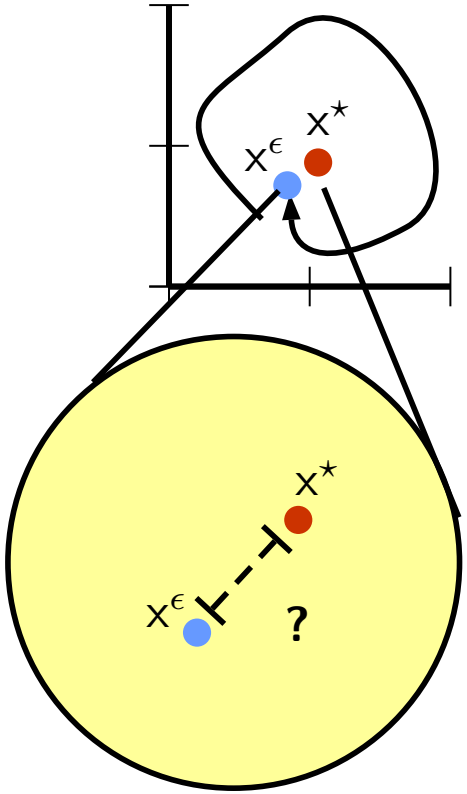
DXL and DXL2 converge to $x^\epsilon \neq x^*$.

Example 7: $\epsilon = 0.2$ DXL (—) cycles while DXL2 (—) converges in Game 2





Second-Order Dynamics from XL



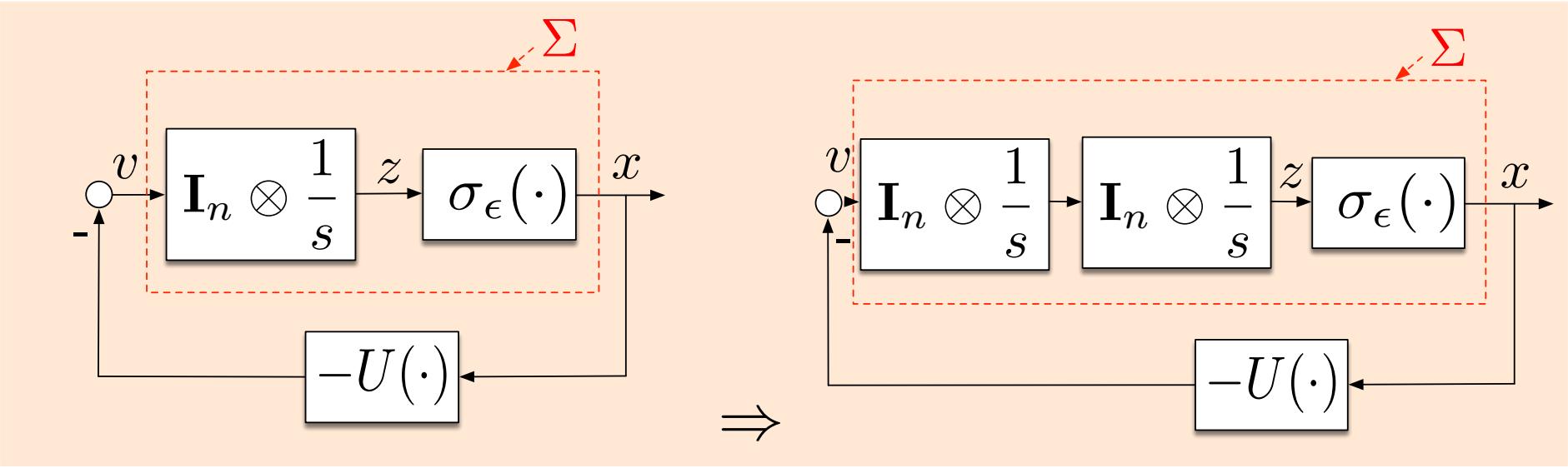
DXL and DXL2 converge to $x^\epsilon \neq x^*$

Can $x(t) \to x^*$ exactly (irrespective of ϵ)?

Go back and start from XL

Second-order XL: Two-fold aggregation of payoff

- From XL

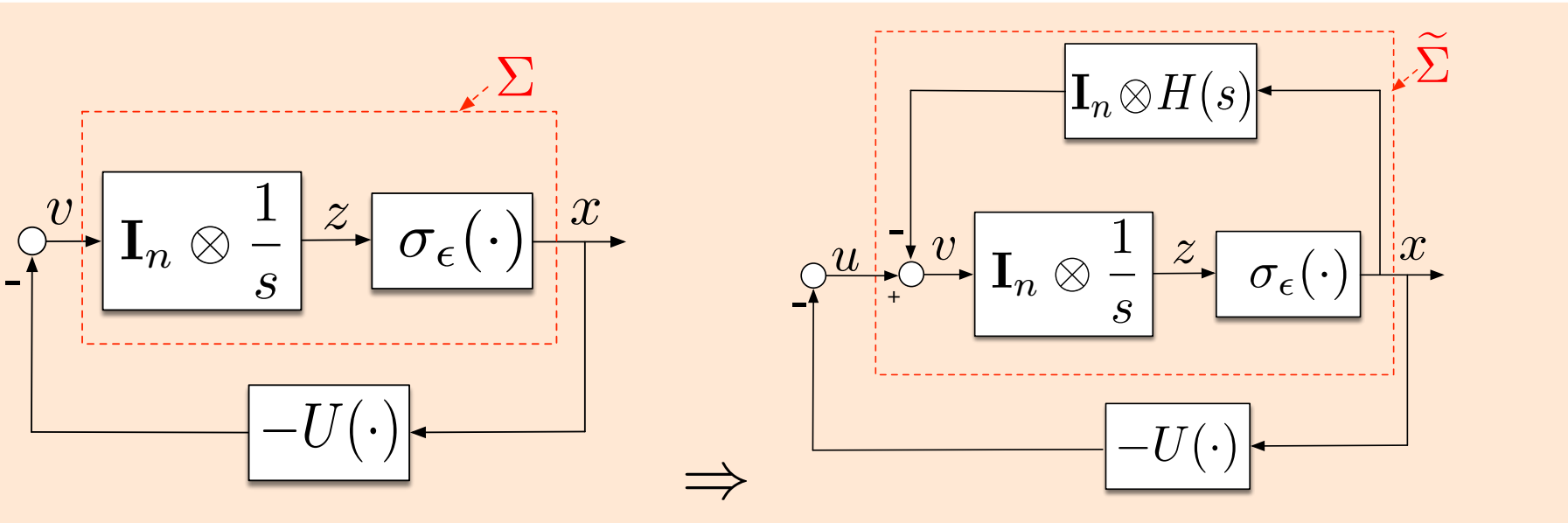


Second-order XL via double aggregation

- ODE: $\ddot{z} = U(x), \quad x = \sigma_\epsilon(z)$. Leads to ODE for x as RD_2 Replicator Dynamics.
- Some nice properties, see [Laraki & Mertikopoulos'13].
- But has the same (non)-convergence properties in RPS games as XL.
- Cascade interconnection does not preserve passivity properties.

Higher-order XL Dynamics: Passivity-based

- From XL

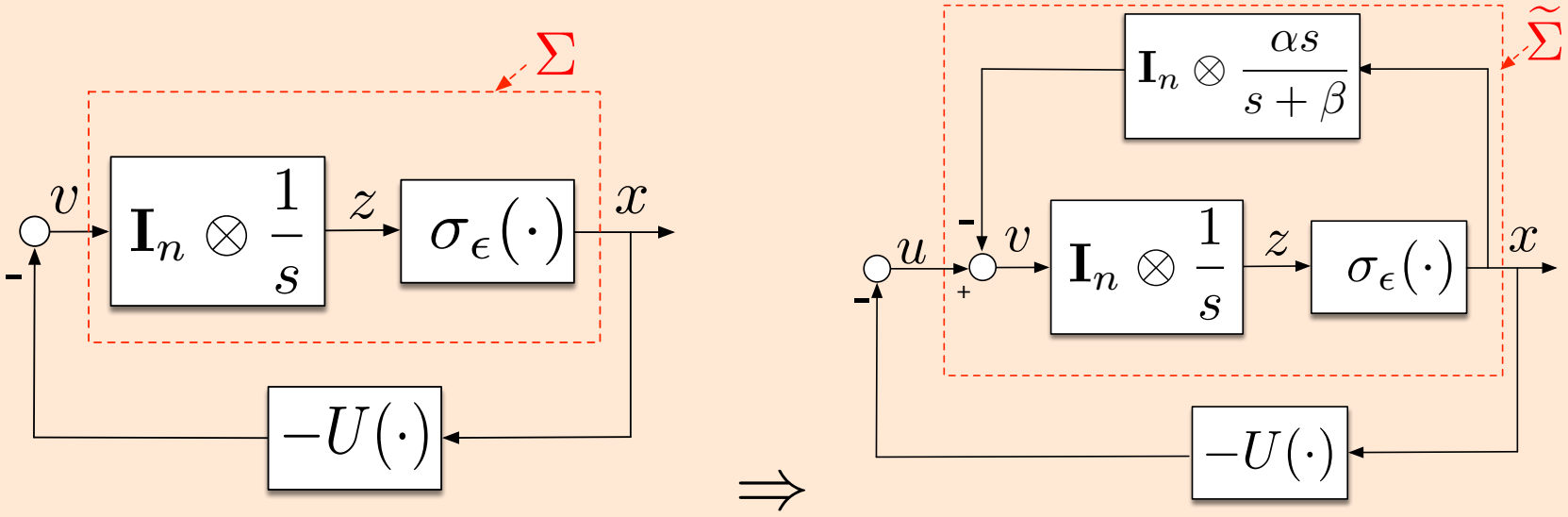


Feedback modified XL

- Feedback interconnection preserves passivity properties.
- Use $H(s)$ strictly passive system, $H(0) = 0$.
- This can improve convergence properties in RPS games.

Second-order XL: via Passivity (XL2)

- From XL



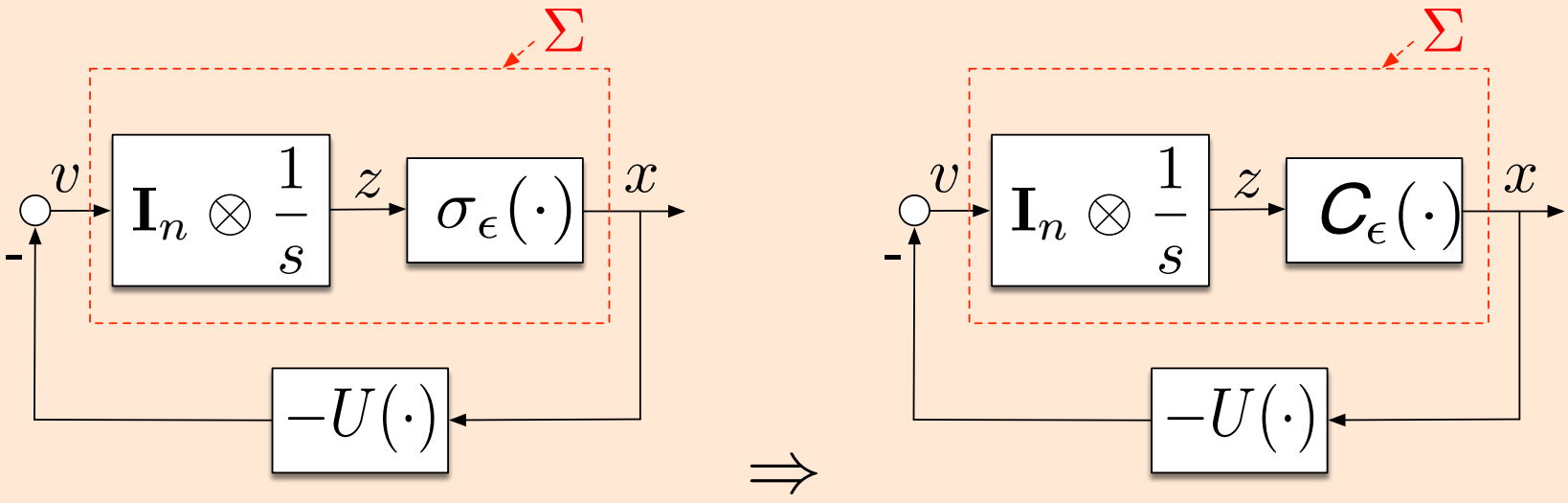
XL2: Second-order XL via Passivity

- $H(s) = \frac{\alpha s}{s+\beta}$, $\alpha, \beta > 0$, lead, output strict EIP.
- ODE: $\dot{z} = U(x) - \frac{\alpha}{\beta}\dot{y}$, $\dot{y} = \beta(x - y)$, $x = \sigma_\epsilon(z)$.
- At rest, $0 = U(x) \implies$ equilibrium is x^* an interior NE.
- Anticipatory term, similar to [Arslan & Shamma'06], can improve convergence.
- ODE for x a form of anticipatory modified RD_2 Replicator Dynamics.

[Gao & Pavel, TAC'24]

Generalizations: XL $\Rightarrow$ MD

- From XL in finite games to Mirror-descent (MD) in continuous games



- MD: dual-space dynamics: $\dot{z} = U(x)$, $x = C_\epsilon(z)$.

  $\Delta \Rightarrow \Omega$, $x^T \log x \Rightarrow h(x)$ regularizer (steep/non-steep)

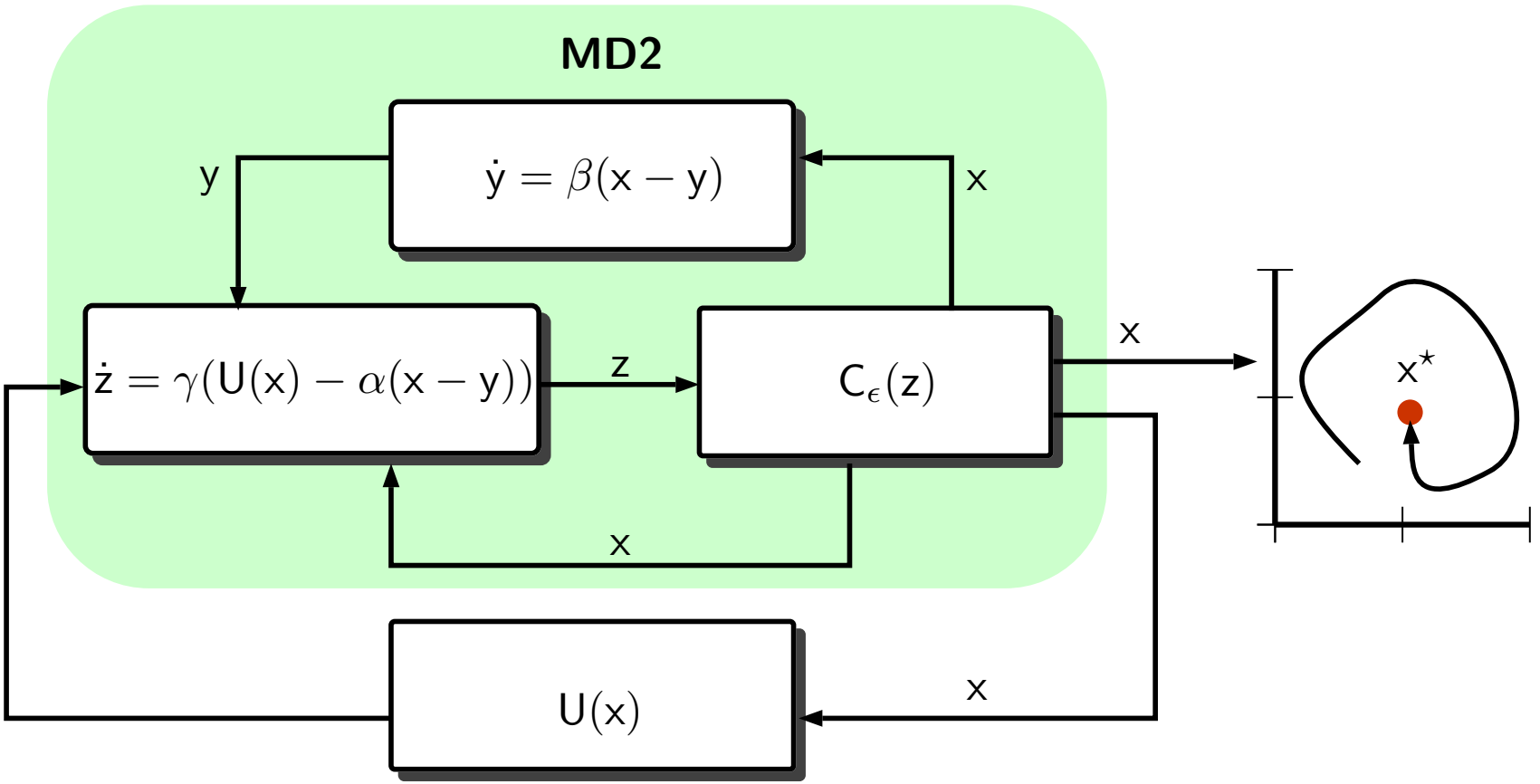
  Softmax $\sigma_\epsilon \Rightarrow C_\epsilon$ mirror map, $C_\epsilon(z_i) = \text{argmax}_{x_i \in \Omega_i} \left[{x_i}^\top z_i - \epsilon h_i(x_i) \right]$.

  $-U$ monotone $\Rightarrow$ VSS equilibrium x^* s.t. $U(x)^\top (x - x^*) \leq 0$, $\forall x \in \Omega$.

- MD converges if x^* is a strict VSS, s.t. $U(x)^\top (x - x^*) < 0$, $\forall x \neq x^* \in \Omega$.

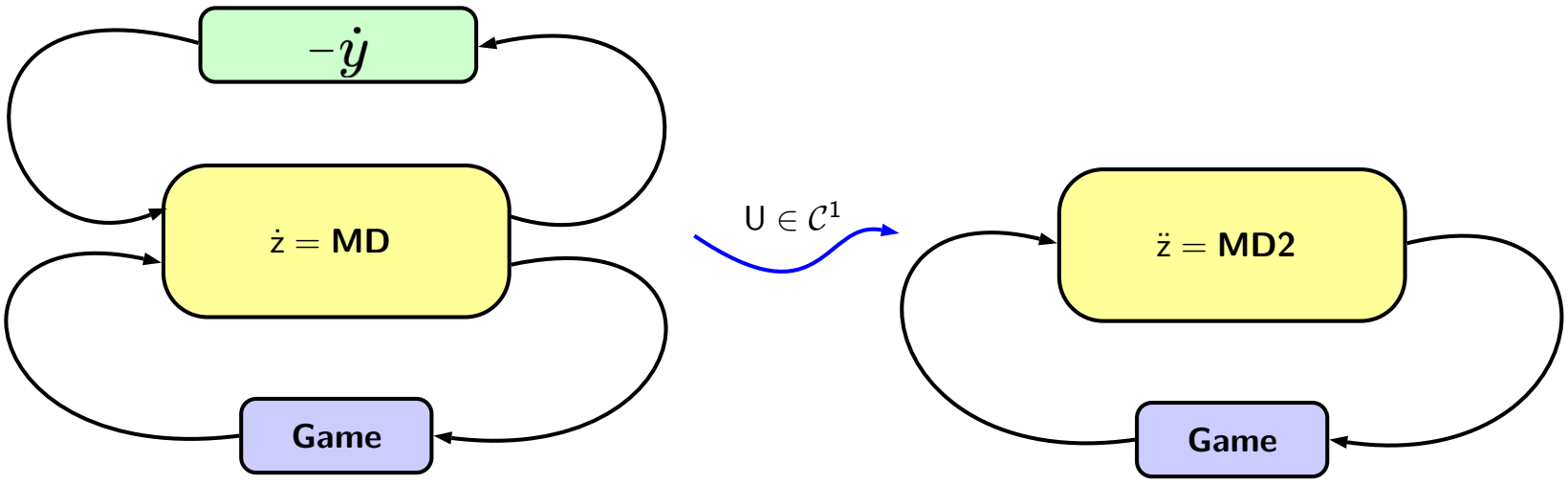
[Zhou, Mertikopoulos& al, CDC'17], [Mertikopoulos, Papdimitriou & Piliouras, ACM'18]

Generalization: XL2 $\Rightarrow$ MD2



- MD2: introduce $\dot{y} = \beta(x - y)$, $\beta > 0$.
- Feedback modified MD, built on passivity-inspired principles.
- MD2 $\cong$ dual-space version of heavy-anchor dynamics [Gadjov & Pavel, TAC'23].
- MD2 converges to interior mere VSS [Gao & Pavel, TAC'24].

Why Second-Order?



Take time derivative of $\dot{z}$ and re-arrange,

$$\begin{aligned} \ddot{z} &= \gamma \left[\mathbf{J}_{U \circ C_\epsilon}(z) - \alpha \mathbf{J}_{C_\epsilon}(z) - \beta \gamma^{-1} \mathbf{I} \right] \dot{z} + \gamma \beta U(x), \\ x &= C_\epsilon(z), \end{aligned} \tag{6}$$

- Unlike straight second-order integration of payoffs [Laraki & Mertikopoulos'13], which has the same (non-)convergence properties as MD.
- Feedback modified, built on passivity-inspired principles.

Simulation: RPS Game

$$A = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix}$$

(null) monotone$(A + A^\top = 0)$

$x^* = (x_1^*, x_2^*), \quad x_i^* = (1/3, 1/3, 1/3)$ is a mere VSS

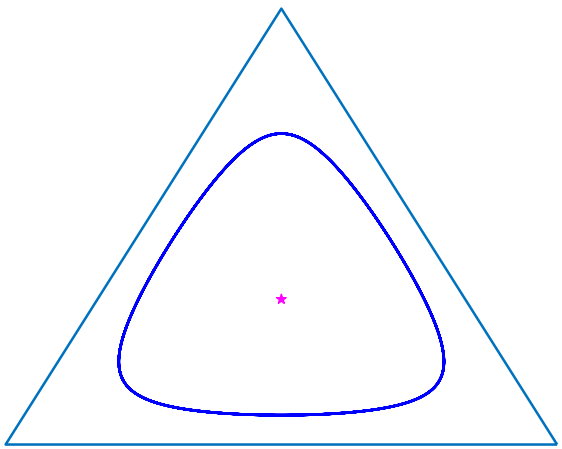


Figure 7: XL

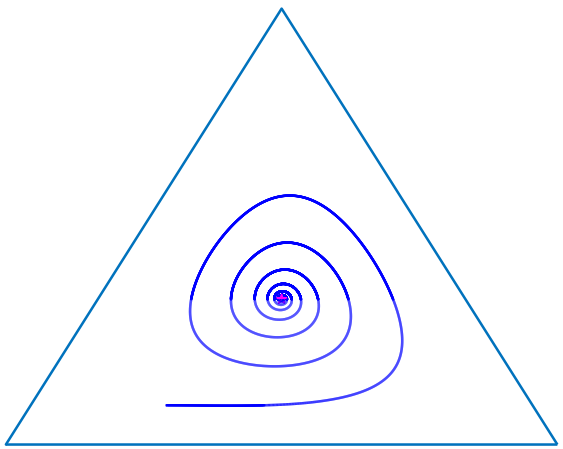


Figure 8: XL2

Simulation: RPS Game

$$A = \begin{bmatrix} 0 & -1.2 & 1 \\ 1 & 0 & -1.2 \\ -1.2 & 1 & 0 \end{bmatrix} \qquad \mu = 0.1\text{-weakly monotone}$$

$x^* = (x_1^*, x_2^*), \quad x_i^* = (1/3, 1/3, 1/3)$ is not a mere VSS (nearly mere VSS)

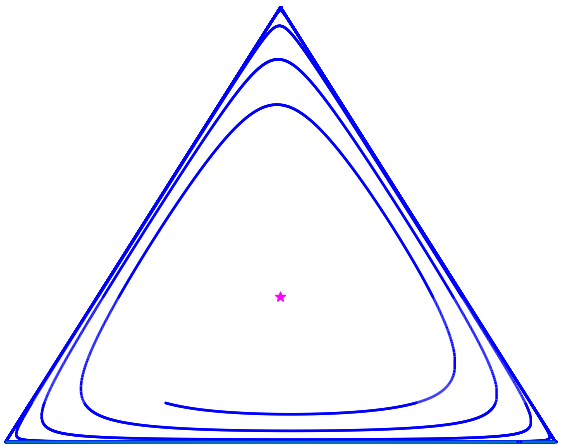


Figure 9: XL

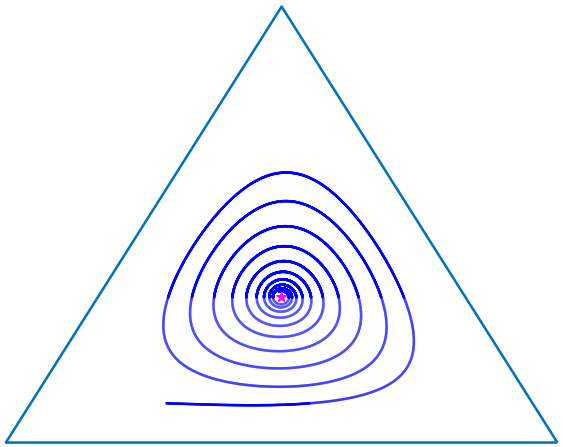
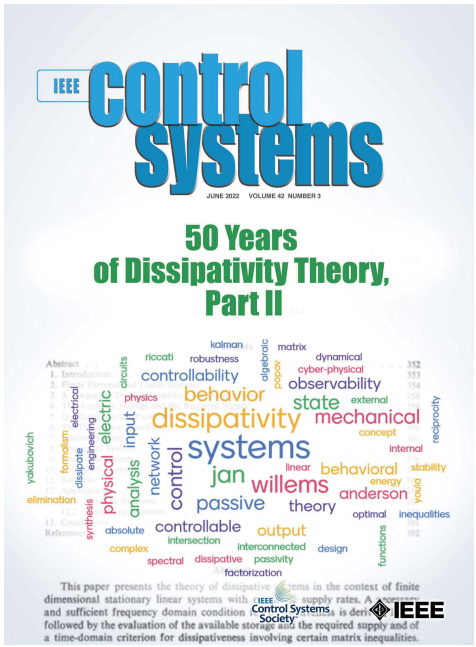


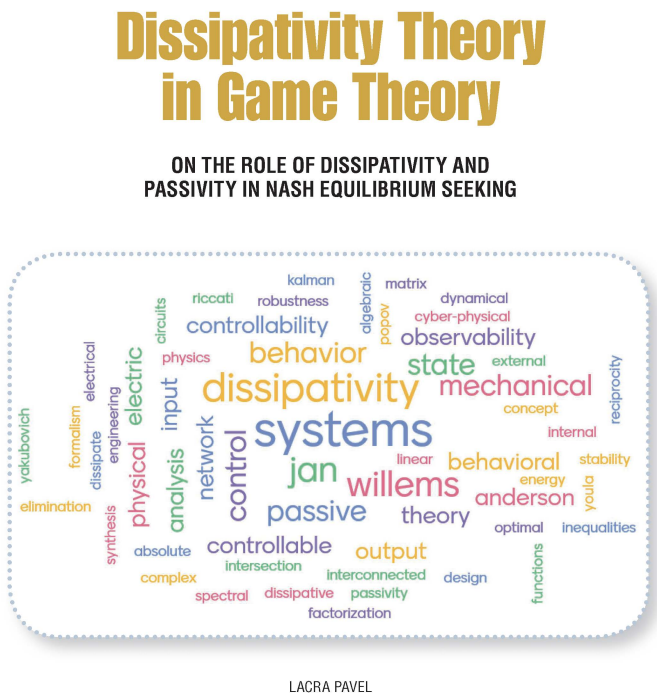
Figure 10: XL2

XL2 still converges (without any tuning) $\Rightarrow$ XL2 is “robust" near the mere VSS!

Conclusions

- Passivity: intuition for convergence of learning dynamics in games
- Interpretation and insights into generalizations and higher-order dynamics (balancing game properties with regularization).
- Other cases: see IEEE CSM Special Issue 2022





- Conjecture: other NE learning algorithms have dissipativity/passivity at play.

- Recent work: dynamic agents [Romano & Pavel, TAC'25] , GNE [Li & Pavel, TAC'26], CBF [Meng & Pavel, in progress]

References

[TAC'21a], Gao, B., L. Pavel, "On passivity, reinforcement learning & higher-order learning in multi-agent finite games," *IEEE Trans. on Automatic Control*, 66(1), 121-136, 2021.

[TAC'21b] Gao, B., L. Pavel, "Continuous-time discounted mirror-descent dynamics in monotone concave games," *IEEE Trans. on Automatic Control*, 66(11), 5451-5458, 2021.

[TAC'24] Gao, B., L. Pavel, "Second-Order Mirror Descent: Convergence in Games Beyond Averaging & Discounting," *IEEE Trans. on Automatic Control*, 69(4), 2143-2157, 2024.

[TAC'23] Gadjov, D., L. Pavel, "On the exact convergence to Nash equilibrium in hy- pomonotone regimes under full and partial-information," *IEEE Trans. on Automatic Control*, 68(8), 4539-4553, 2023.

[CSM'22] Pavel, L., "Dissipativity Theory in Game Theory: On the role of dissipativity and passivity in NE-seeking," *IEEE Control Systems Magazine*, Special Issue: 50 years of dissipativity theory - Part II, 42(3), 150-164, 2022.

[TAC'25] Romano, A., L. Pavel, "Game-theoretic steady-state control: A general framework for NE seeking with LTI agents," *IEEE Trans. on Automatic Control*, Early Access, 2025.

[TAC'26] Li, W. L. Pavel, "Passivity-based gradient-play dynamics for distributed Generalized Nash equilibrium seeking." *IEEE Trans. on Automatic Control, Early Access, 2026.*