Higher-order Learning in Multi-Agent Games

Lacra Pavel

University of Toronto
*based on joint work with Bolin Gao

Workshop 9: Large Population Teams:
Control, Equilibria, and Learning
63^{rd} IEEE Conference on Decision and Control
Milan, December 15, 2024

Game Setup

$$\mathcal{G} = (\mathcal{N}, (\Omega^p)_{p\in\mathcal{N}}, (\mathcal{U}^p)_{p\in\mathcal{N}})$$

$\mathcal{N} = \{1, \ldots, N\}$
N players

$x^p \in \Omega^p \subseteq \mathbb{R}^{n_p}$
set of strategies

$\mathcal{U}^p : \Omega = \prod_{p=1}^{N} \Omega^p \to \mathbb{R}$
scalar-valued payoff function

Assumption. $\mathcal{G}$ is continuous concave, i.e., $\forall p$,

- Ω^p non-empty, closed, convex subset of $\mathbb{R}^{n_p}$,
- $\mathcal{U}^p(x^p; x^{-p}) = \mathcal{U}^p(x)$ (jointly) continuous in $x = (x^p; x^{-p})$,
- $\mathcal{U}^p(x^p; x^{-p})$ concave and $\mathcal{C}^1$ in x^p, $\forall x^{-p} \in \Omega^{-p}$.

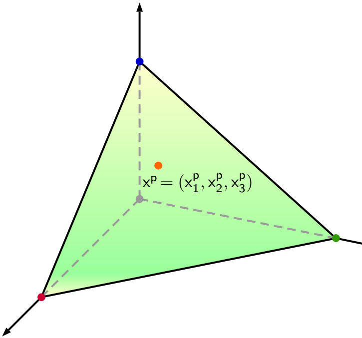
Example 1: Rock-Paper-Scissors

$$
\begin{array}{cc}
 & \begin{array}{ccc} \mathbf{r} & \mathbf{p} & \mathbf{s} \end{array} \\
\begin{array}{c} \mathbf{r} \\ \mathbf{p} \\ \mathbf{s} \end{array} &
\begin{bmatrix} (0,0) & (-l,w) & (w,-l) \\ (w,-l) & (0,0) & (-l,w) \\ (-l,w) & (w,-l) & (0,0) \end{bmatrix}
\end{array}
$$

$$l: \text{ loss } > 0, w: \text{ win } > 0$$

- $i^p \in \{\mathbf{r}, \mathbf{p}, \mathbf{s}\}$
- $\mathbf{i} = (i^1; i^2) \sim$ reward
- x^p prob. of selection
- $\mathcal{U}^p$ expected reward,

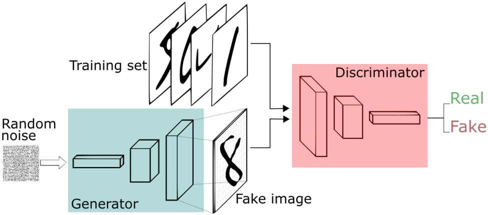
  $\mathcal{U}^p(x) = \pm x^{p\top} \begin{bmatrix} 0 & -l & w \\ w & 0 & -l \\ -l & w & 0 \end{bmatrix} x^{\neg p}$



Example 2: Saddle-point problems

$$\max_{x^1 \in \Omega^1} \min_{x^2 \in \Omega^2} f(x^1, x^2)$$

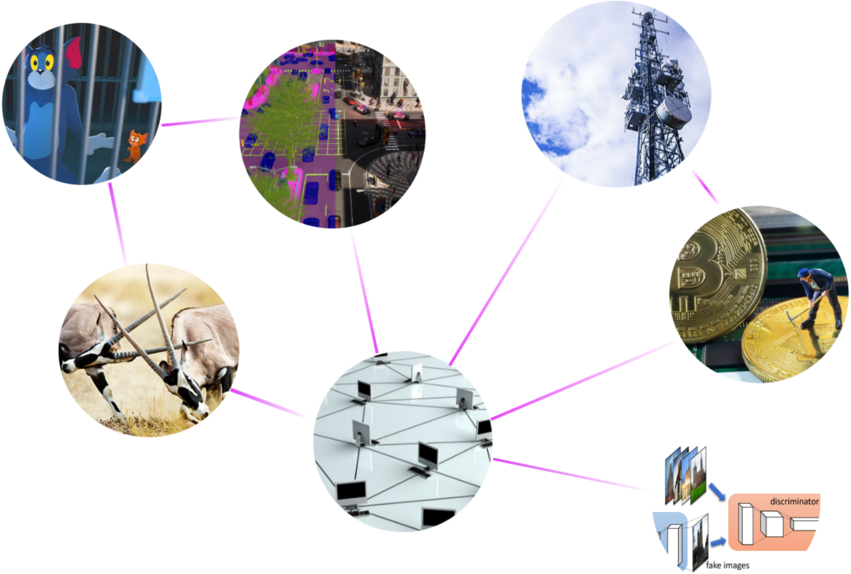
- designer and critic
- designer: submits design, x^1
- critic: submits appraisal, x^2
- $\mathcal{U}^1 = f(x^1, x^2) = -\mathcal{U}^2$
  "agreement/satisfaction"



Generative Adversarial Network (GAN)

After a "few" simplifying assumptions:

$$f(x^1, x^2) = x^1 \cdot x^2$$



Behavioral science, ecology, wireless networks, (virtual) economy, traffic modeling...

Game Solution

Player p's goal: given $x^{-p} \in \Omega^{-p}, \forall p \in \mathcal{N}, \max_{x^p \in \Omega^p} \mathcal{U}^p(x^p; x^{-p})$

$x^\star = (x^{p\star}; x^{-p\star}) \in \Omega$ is **Nash equilibrium** (NE) when no players can benefit from unilateral deviation:

$$\mathcal{U}^p(x^{p\star}; x^{-p\star}) \geq \mathcal{U}^p(x^p; x^{-p\star}), \forall x^p \in \Omega^p, \forall p \in \mathcal{N} \tag{1}$$

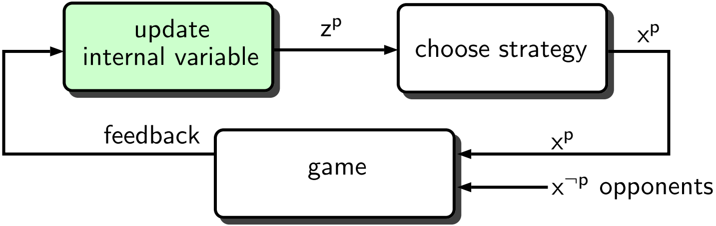
Equiv., under our concave game assumption,

$$(x - x^\star)^\top U(x^\star) \leq 0, \forall x \in \Omega \tag{2}$$

$U(x) = (U^p)_{p \in \mathcal{N}} = (\nabla_{x^p} \mathcal{U}^p(x^p; x^{-p}))_{p \in \mathcal{N}}$ (**pseudo-gradient**).

Players use simple rules/models to convert game information to their strategies, hopefully leading to a NE.

NE seeking via online learning

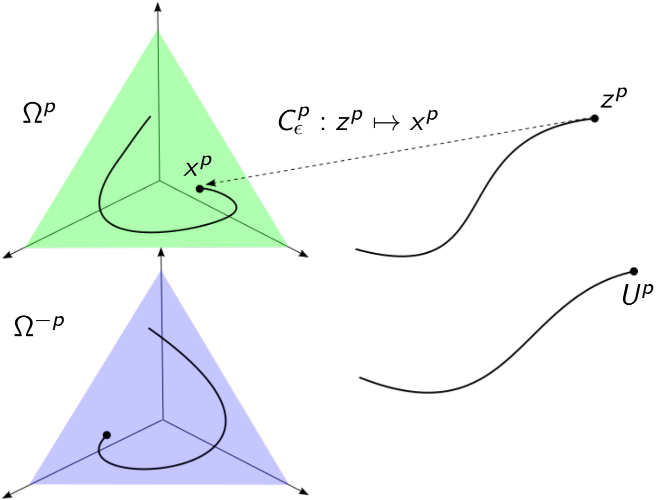


A group of players **learns** a NE by individually:

- choose strategies via an internal variable†
- receive information (feedback) from the game
- update own internal variable (← "learning")

†Q-value, dual aggregate, score, perception, model weights, etc.

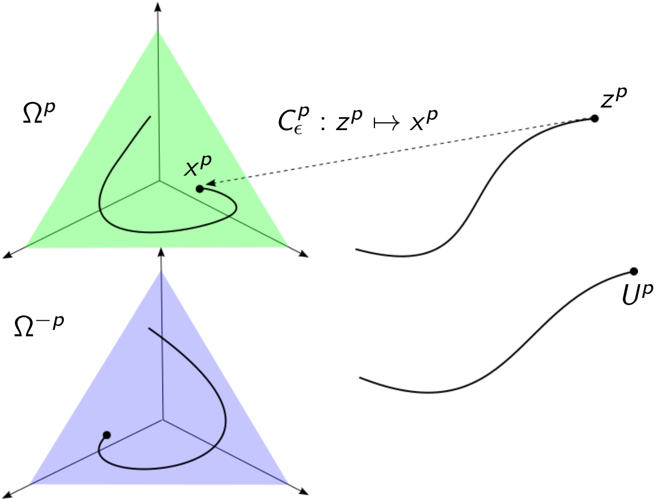
Each player maps its own variable z^p into a strategy $x^p \in \Omega^p$ through a *mirror map* $C_\epsilon^p : \mathbb{R}^{n_p} \to \Omega^p$



$$C_\epsilon^p(z^p) = \text{argmax}_{y^p \in \Omega^p} \left[{y^p}^\top z^p - \epsilon v^p(y^p) \right], \epsilon > 0 \qquad (\star)$$

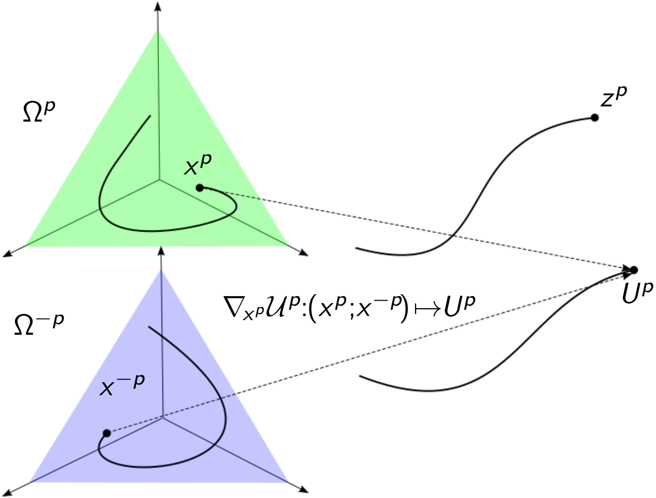
v^p: regularizer, ϵ: regularization constant

Each player maps its own variable z^p into a strategy $x^p \in \Omega^p$ through a *mirror map* $C_\epsilon^p : \mathbb{R}^{n_p} \to \Omega^p$

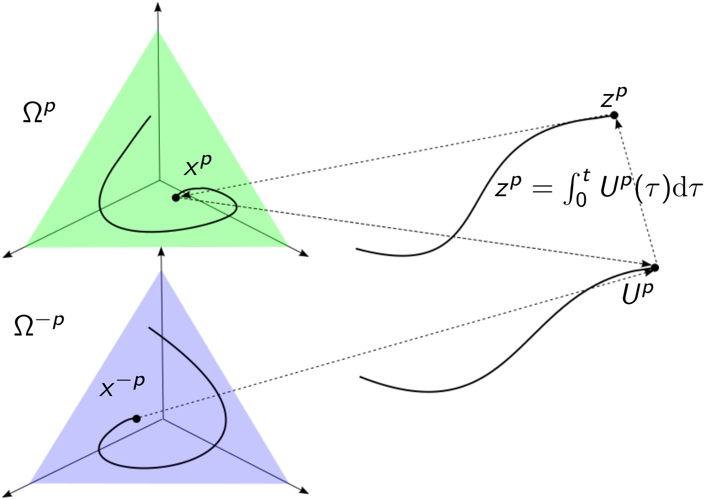


- $v^p = \|x^p\|_2^2/2 \Rightarrow C_\epsilon^p(z^p) = \pi_{\Omega^p}(\epsilon^{-1} z^p)$ (Projection)
- $v^p = {x^p}^\top \log(x^p) \Rightarrow C_\epsilon^p(z^p) = \frac{\exp(\epsilon^{-1} z^p)}{\sum_{q \in \mathcal{N}} \exp(\epsilon^{-1} z^q)}$ (Softmax)

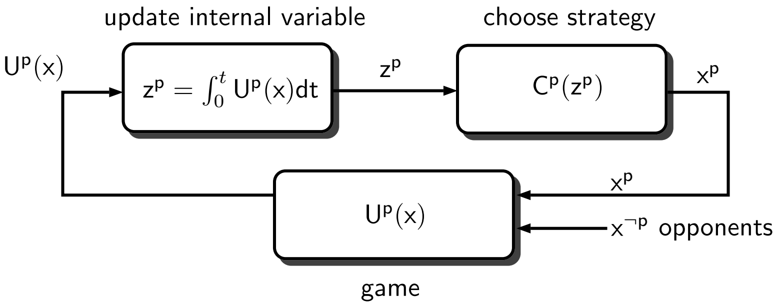
Each player plays the game $\mathcal{G}$ using strategy x^p (taking into account opponents' $x^{-p} \in \Omega^{-p}$) and obtains partial-gradient U^p



Each player maps U^p back into z^p via aggregation $z^p = \int_0^t U^p(\tau)\mathrm{d}\tau$ and the process continues indefinitely



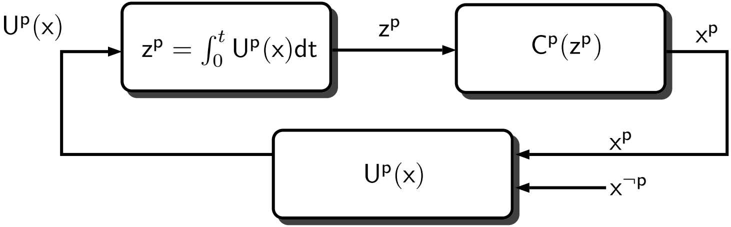
(Continuous-Time) Mirror Descent (Nemirovski '83)



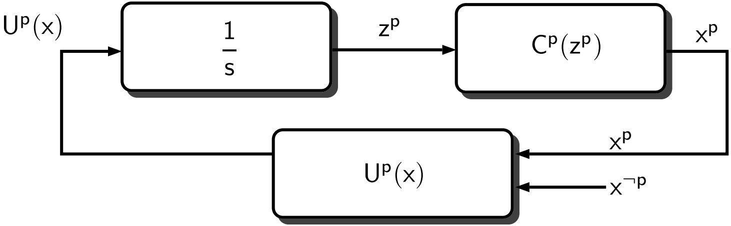
- z^p: a vector of player p's internal state
- C^p: "mirror map", $z^p \mapsto x^p$ e.g., Euclidean projection
- U^p: a partial-gradient

$$U^p(x) = \nabla_{x^p} \mathcal{U}^p(x^p; x^{\neg p}) = \frac{\partial \mathcal{U}^p(x^p; x^{\neg p})}{\partial x^p}$$

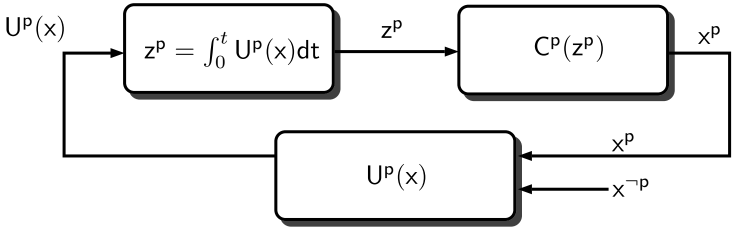
- $x = (x^p; x^{\neg p})$: joint strategy



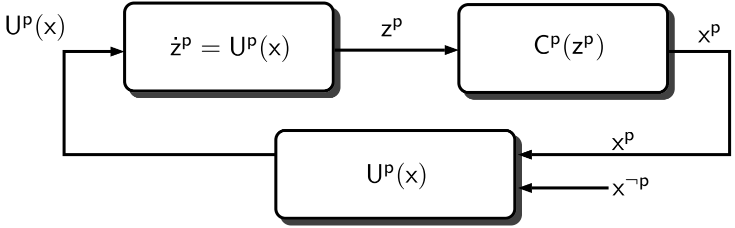
Up(x)
zp = ∫_0^t Up(x)dt
zp
Cp(zp)
xp
xp
Up(x)
x¬p



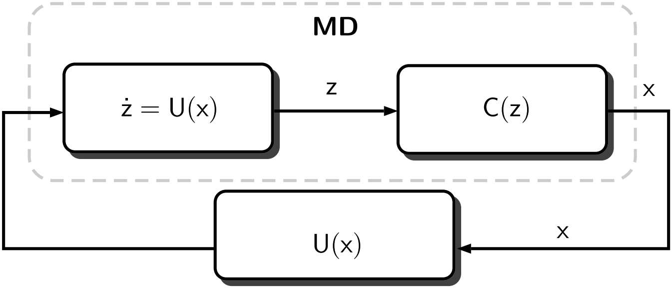
Up(x)
1/s
zp
Cp(zp)
xp
xp
Up(x)
x¬p



$U^p(x)$
$z^p = \int_0^t U^p(x)dt$
z^p
$C^p(z^p)$
x^p
x^p
$U^p(x)$
$x^{\neg p}$



$U^p(x)$
$\dot{z}^p = U^p(x)$
z^p
$C^p(z^p)$
x^p
x^p
$U^p(x)$
$x^{\neg p}$



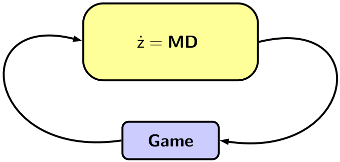
To analyze collective behavior,
stack all $p = 1, \ldots, N$:

$$z = \begin{bmatrix} z^1 \\ \vdots \\ z^N \end{bmatrix} \quad U = \begin{bmatrix} U^1 \\ \vdots \\ U^N \end{bmatrix}$$

U is called a **pseudo-gradient**.

$$x = \begin{bmatrix} x^1 \\ \vdots \\ x^N \end{bmatrix} \quad C = \begin{bmatrix} C^1 \\ \vdots \\ C^N \end{bmatrix}$$

Convergence of MD



Using stacked notation,

$$z = \begin{bmatrix} z^1 \\ \vdots \\ z^N \end{bmatrix} \quad U = \begin{bmatrix} U^1 \\ \vdots \\ U^N \end{bmatrix} \quad x = \begin{bmatrix} x^1 \\ \vdots \\ x^N \end{bmatrix} \quad C_\epsilon = \begin{bmatrix} C_\epsilon^1 \\ \vdots \\ C_\epsilon^N \end{bmatrix}$$

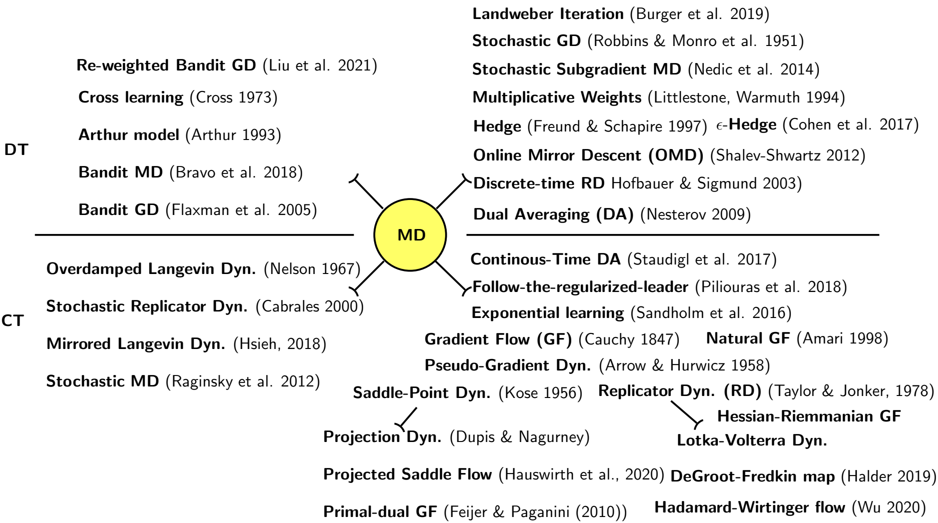
Can represent this entire process as,

$$z = \int_0^t U(x)\mathrm{d}\tau, \quad x = C_\epsilon(z), \tag{3}$$

$$\Updownarrow$$

$$\dot{z} = U(x), x = C_\epsilon(z), \qquad \text{MD}$$

Question: when does $x(t) = C_\epsilon(z(t))$ converge to $x^\star$?


MD
DT
CT
Re-weighted Bandit GD (Liu et al. 2021)
Cross learning (Cross 1973)
Arthur model (Arthur 1993)
Bandit MD (Bravo et al. 2018)
Bandit GD (Flaxman et al. 2005)
Landweber Iteration (Burger et al. 2019)
Stochastic GD (Robbins & Monro et al. 1951)
Stochastic Subgradient MD (Nedic et al. 2014)
Multiplicative Weights (Littlestone, Warmuth 1994)
Hedge (Freund & Schapire 1997)
ϵ-Hedge (Cohen et al. 2017)
Online Mirror Descent (OMD) (Shalev-Shwartz 2012)
Discrete-time RD Hofbauer & Sigmund 2003)
Dual Averaging (DA) (Nesterov 2009)
Overdamped Langevin Dyn. (Nelson 1967)
Stochastic Replicator Dyn. (Cabrales 2000)
Mirrored Langevin Dyn. (Hsieh, 2018)
Stochastic MD (Raginsky et al. 2012)
Continous-Time DA (Staudigl et al. 2017)
Follow-the-regularized-leader (Piliouras et al. 2018)
Exponential learning (Sandholm et al. 2016)
Gradient Flow (GF) (Cauchy 1847)
Natural GF (Amari 1998)
Pseudo-Gradient Dyn. (Arrow & Hurwicz 1958)
Saddle-Point Dyn. (Kose 1956)
Replicator Dyn. (RD) (Taylor & Jonker, 1978)
Hessian-Riemmanian GF
Projection Dyn. (Dupis & Nagurney)
Lotka-Volterra Dyn.
Projected Saddle Flow (Hauswirth et al., 2020)
DeGroot-Fredkin map (Halder 2019)
Primal-dual GF (Feijer & Paganini (2010))
Hadamard-Wirtinger flow (Wu 2020)

MD converges under one of a trio of strict assumptions [1]

1. $\mathcal{G}$ is **strictly monotone** (Rosen '65),

$$(U(x) - U(x'))^\top (x - x') < 0, \forall x \in \Omega \setminus \{x'\} \tag{4}$$

2. $x^\star$ is a **strict variationally stable state** (VSS) (Smith '73),

$$U(x)^\top (x - x^\star) < 0, \forall x \in \Omega \setminus \{x^\star\} \tag{5}$$

NE of strictly monotone game $\implies$ strict VSS

3. $x^\star$ is a **strict Nash equilibrium** (Harsanyi '73),

$$\mathcal{U}^p(x^p; x^{-p\star}) < \mathcal{U}^p(x^{p\star}; x^{-p\star}), \forall x^p \in \Omega^p \setminus \{x^{p\star}\}, \forall p \in \mathcal{N} \tag{6}$$

or,

$$U(x^\star)^\top (x - x^\star) < 0, \forall x \in \Omega \setminus \{x^\star\} \tag{7}$$

strict NE = strict VSS in finite games

[1] See e.g., Mertikopoulos & Sandholm '16, Mertikopoulos & Staudigl '17, Mertikopoulos & Zhou '19, Migot & Cojocaru '21, Laraki & Mertikopoulos '13, Higher order game dynamics

MD converges under one of a trio of strict assumptions [1]

1. $\mathcal{G}$ is **strictly monotone** (Rosen '65),

$$(U(x) - U(x'))^\top (x - x') < 0, \forall x \in \Omega \backslash \{x'\} \tag{4}$$

2. $x^\star$ is a **strict variationally stable state** (VSS) (Smith '73),

$$U(x)^\top (x - x^\star) < 0, \forall x \in \Omega \backslash \{x^\star\} \tag{5}$$

NE of strictly monotone game $\implies$ strict VSS

3. $x^\star$ is a **strict Nash equilibrium** (Harsanyi '73),

$$\mathcal{U}^p(x^p; x^{-p\star}) < \mathcal{U}^p(x^{p\star}; x^{-p\star}), \forall x^p \in \Omega^p \backslash \{x^{p\star}\}, \forall p \in \mathcal{N} \tag{6}$$

or,

$$U(x^\star)^\top (x - x^\star) < 0, \forall x \in \Omega \backslash \{x^\star\} \tag{7}$$

strict NE = strict VSS in finite games

[1] See e.g., Mertikopoulos & Sandholm '16, Mertikopoulos & Staudigl '17, Mertikopoulos & Zhou '19, Migot & Cojocaru '21, Laraki & Mertikopoulos '13, Higher order game dynamics

MD converges under one of a trio of strict assumptions [1]

1. $\mathcal{G}$ is **strictly monotone** (Rosen '65),

$$(U(x)-U(x'))^\top(x-x')<0, \forall x \in \Omega\backslash\{x'\} \quad (4)$$

2. $x^\star$ is a **strict variationally stable state** (VSS) (Smith '73),

$$U(x)^\top(x-x^\star)<0, \forall x \in \Omega\backslash\{x^\star\} \quad (5)$$

NE of strictly monotone game $\implies$ strict VSS

3. $x^\star$ is a **strict Nash equilibrium** (Harsanyi '73),

$$\mathcal{U}^p(x^p;x^{-p\star})<\mathcal{U}^p(x^{p\star};x^{-p\star}), \forall x^p \in \Omega^p\backslash\{x^{p\star}\}, \forall p \in \mathcal{N} \quad (6)$$

or,

$$U(x^\star)^\top(x-x^\star)<0, \forall x \in \Omega\backslash\{x^\star\} \quad (7)$$

strict NE = strict VSS in finite games

[1] See e.g., Mertikopoulos & Sandholm '16, Mertikopoulos & Staudigl '17, Mertikopoulos & Zhou '19, Migot & Cojocaru '21, Laraki & Mertikopoulos '13, Higher order game dynamics

Two wider settings MD can fail to converge

1. $\mathcal{G}$ is **merely monotone** ($=$ stable game) (Hofbauer et al. '09)

$$(U(x) - U(x'))^\top (x - x') \leq 0, \forall x \in \Omega \tag{8}$$

2. $x^\star$ is a **mere variationally stable state** (VSS) (Smith, '82)

$$U(x)^\top (x - x^\star) \leq 0, \forall x \in \Omega \tag{9}$$

NE of merely monotone game $\implies$ mere VSS

Example. (Two-Player Rock-Paper-Scissors (RPS))

$$\mathcal{A} = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix} \quad \mathcal{U}^p(x^p; x^q) = {x^p}^\top \mathcal{A} x^q, x^p \in \textbf{Simplex}$$

RPS with $\mathcal{A}$ is merely monotone $\implies$ unique interior NE
$x^\star = (x^{p\star}), x^{p\star} = (1/3, 1/3, 1/3)$ is a mere VSS.

Two wider settings MD can fail to converge

1. $\mathcal{G}$ is **merely monotone** (= stable game) (Hofbauer et al. '09)

$$(U(x)-U(x'))^\top(x-x') \leq 0, \forall x \in \Omega \tag{8}$$

2. $x^\star$ is a **mere variationally stable state** (VSS) (Smith, '82)

$$U(x)^\top(x-x^\star) \leq 0, \forall x \in \Omega \tag{9}$$

NE of merely monotone game $\Longrightarrow$ mere VSS

Example. (Non-Monotone Game with Mere VSS)

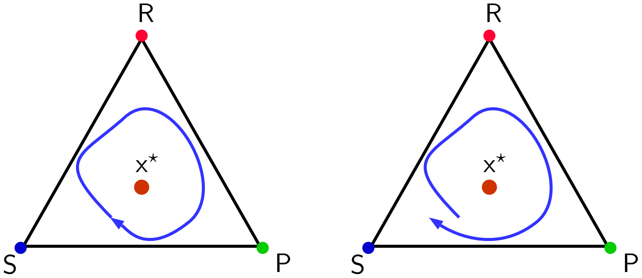
$$\mathcal{U}^p(x^p, x^{-p}) = -\prod_{p=1}^{N} x^p, x^p \in [0,1] \tag{10}$$

Game not merely monotone for $N > 2$, NE $x^\star = \mathbf{0}$ is a mere VSS.

Long-run strategy generated by MD corresponds to NE in some games

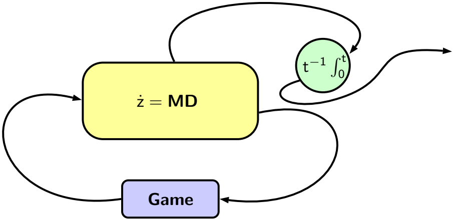
But...

MD fails to converge to NE for a wide range of games, even simple ones



Fundamental limitation in applications

Overcoming non-convergence of MD: time-averaging



Mirror descent with time-averaging (MDA) (e.g., Mertikopoulos & Sandholm, '16, Hofbauer, Sorin & Viossat '09),

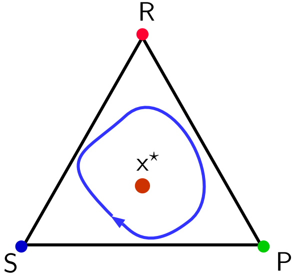
$$\dot{z} = U(x),\, x = C_\epsilon(z),\, x_{\text{avg}}(t) = t^{-1} \int_0^t x(\tau)\mathrm{d}\tau \qquad \text{(MDA)}$$

- Converges exactly in ZS game with interior NE, which could be a mere VSS.
- Requires all to use time-averaging $\implies$ not "game-realistic".
- Not robust to slight game parameter perturbation.

Alternative: Design via passivity principles

Observation: MD fails in games with specific types of properties

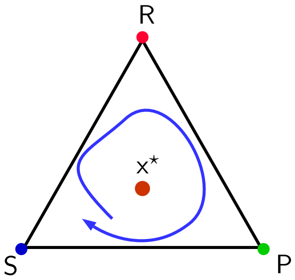
$$\begin{bmatrix} 0 & -l & w \\ w & 0 & -l \\ -l & w & 0 \end{bmatrix} \quad \color{red}{l = w}$$



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

$$\underbrace{(U(x) - U(x'))^\top (x - x')}_{\text{monotonicity product}} \leq 0, \forall x, x' \in \Omega. \quad (2)$$

$$\begin{bmatrix} 0 & -l & w \\ w & 0 & -l \\ -l & w & 0 \end{bmatrix} \quad l > w$$

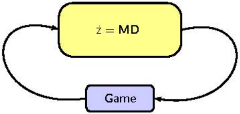


This is an example of a μ-**weakly monotone** game:

$$(U(x) - U(x'))^\top (x - x') \leq \mu \|x - x'\|_2^2, \forall x, x' \in \Omega. \qquad (3)$$

Higher $\mu \geq 0$, "harder" the game, "worse" the behavior.

Observation: (weak) monotonicity of game = (lack of) energy dissipation or **passivity/dissipativity** of feedback subsystem U.



Observation: (weak) monotonicity of game = (lack of) energy dissipation or **passivity/dissipativity** of feedback subsystem U.

3.1 Equilibrium Independent Dissipativity (EID)

Consider the system

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = f(x(t), u(t)) \quad (3.3)$$

$$y(t) = h(x(t), u(t)) \quad (3.4)$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $y(t) \in \mathbb{R}^p$, and suppose there exists a set $\mathscr{X} \subset \mathbb{R}^n$ where, for every $\bar{x} \in \mathscr{X}$, there is a unique $\bar{u} \in \mathbb{R}^m$ satisfying $f(\bar{x}, \bar{u}) = 0$. Thus $\bar{u}$ and $\bar{y} \triangleq h(\bar{x}, \bar{u})$ are implicit functions of $\bar{x}$.

Definition 3.1 We say that the system above is **equilibrium independent dissipative (EID)** with supply rate $s(\cdot, \cdot)$ if there exists a continuously differentiable storage function $V : \mathbb{R}^n \times \mathscr{X} \mapsto \mathbb{R}$ satisfying, $\forall (x, \bar{x}, u) \in \mathbb{R}^n \times \mathscr{X} \times \mathbb{R}^m$

$$V(x, \bar{x}) \geq 0, \qquad V(\bar{x}, \bar{x}) = 0, \qquad \nabla_x V(x, \bar{x})^T f(x, u) \leq s(u-\bar{u}, y-\bar{y}). \quad (3.5)$$

From Arcak, Meissen, Packard (2016)

Observation: (weak) monotonicity of game = (lack of) energy dissipation or **passivity/dissipativity** of feedback subsystem U.

3.1 Equilibrium Independent Dissipativity (EID)

Consider the system

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = f(x(t), u(t)) \quad (3.3)$$

$$y(t) = h(x(t), u(t)) \quad (3.4)$$

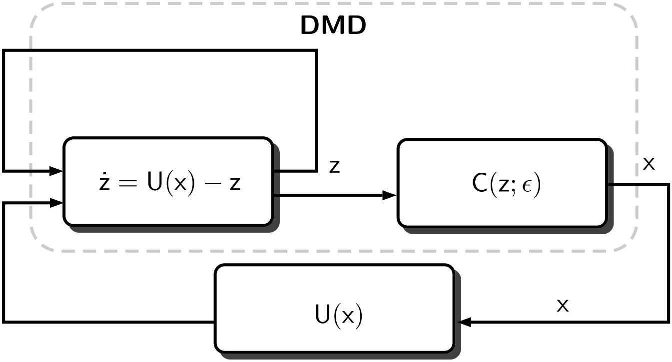
where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $y(t) \in \mathbb{R}^p$, and suppose there exists a set $\mathscr{X} \subset \mathbb{R}^n$ where, for every $\bar{x} \in \mathscr{X}$, there is a unique $\bar{u} \in \mathbb{R}^m$ satisfying $f(\bar{x}, \bar{u}) = 0$. Thus $\bar{u}$ and $\bar{y} \triangleq h(\bar{x}, \bar{u})$ are implicit functions of $\bar{x}$.

Definition 3.1 We say that the system above is **equilibrium independent dissipative (EID)** with supply rate $s(\cdot, \cdot)$ if there exists a continuously differentiable storage function $V : \mathbb{R}^n \times \mathscr{X} \mapsto \mathbb{R}$ satisfying, $\forall (x, \bar{x}, u) \in \mathbb{R}^n \times \mathscr{X} \times \mathbb{R}^m$

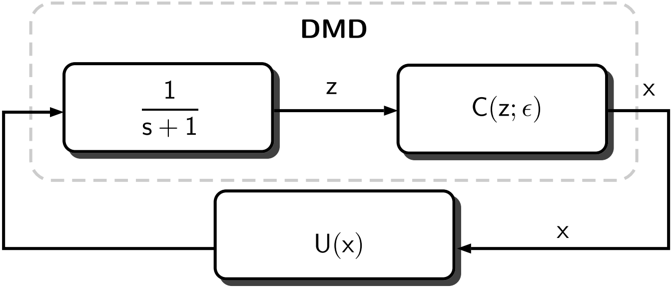
$$V(x, \bar{x}) \geq 0, \qquad V(\bar{x}, \bar{x}) = 0, \qquad \nabla_x V(x, \bar{x})^T f(x, u) \leq s(u - \bar{u}, y - \bar{y}). \quad (3.5)$$

From Arcak, Meissen, Packard (2016)

Idea: add **passivity/dissipativity** (more energy dissipation) to the entire system.



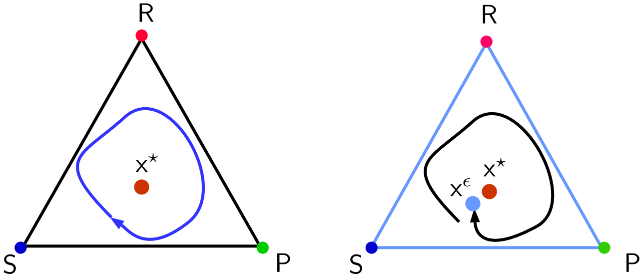
- Let players re-evaluate internal variable z during play.
- Call this **discounted mirror descent** (DMD).



DMD is “output-strictly” passive.

There exists a **storage function** V associated with DMD s.t.

$$\dot{V} \leq (x - \overline{x})^\top (U(x) - U(\overline{x})) - \epsilon \|x - \overline{x}\|_2^2$$



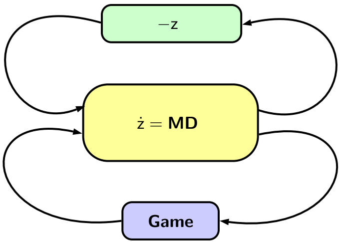
Theorem[1,2]: $\mathcal{G}$ merely monotone matrix game, $x(t) \to x^{\epsilon}$ (x^{ϵ} **logit equilibrium**).

$$x^{\epsilon} \to x^{\star} \text{ as } \epsilon \to 0$$

[1] B. Gao, L. Pavel, "On Passivity and Reinforcement Learning in Finite Games", in 57th IEEE CDC, 2018

[2] B. Gao, L. Pavel, "On Passivity, Reinforcement Learning and Higher-Order Learning in Multi-Agent Finite Games", in IEEE TAC, 2021

Overcoming non-convergence: discounting



Discounted mirror descent (DMD) (Coucheney et al. '15, Gao & Pavel '21) ,

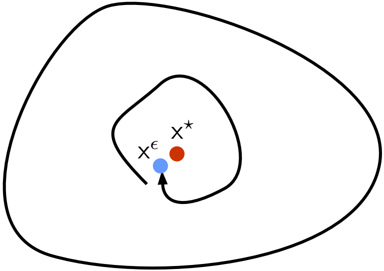
$$\dot{z} = U(x) - z\,,\; x = C_\epsilon(z) \qquad \text{DMD}$$

- Converges in merely monotone game, hence to a mere VSS.
- Does not converge exactly (in general).

Reason: when dynamics settle, feedback term $-z$ does not vanish, perturbing the solution away from $x^\star$:

$$\dot{z} = \textbf{DMD} = \textbf{MD} + \textbf{non-vanishing feedback}$$

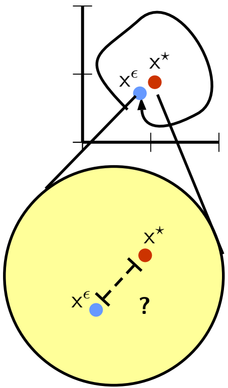
Higher-order learning MD



$x(t) \to x^\star$ inexact in general

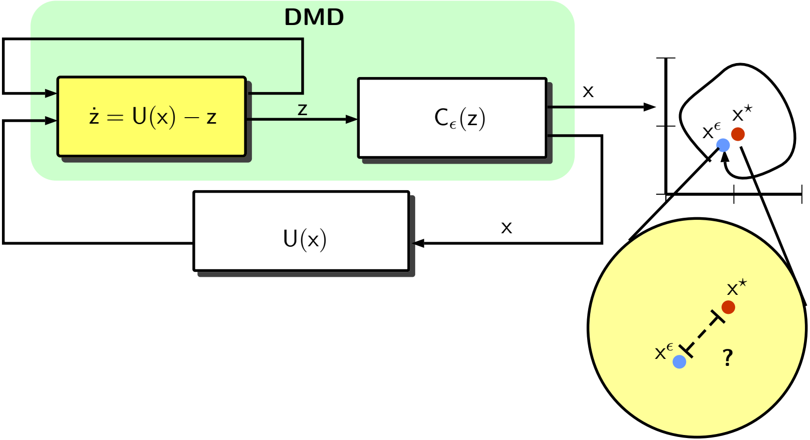
Many games lack monotonicity

Higher-Order learning MD

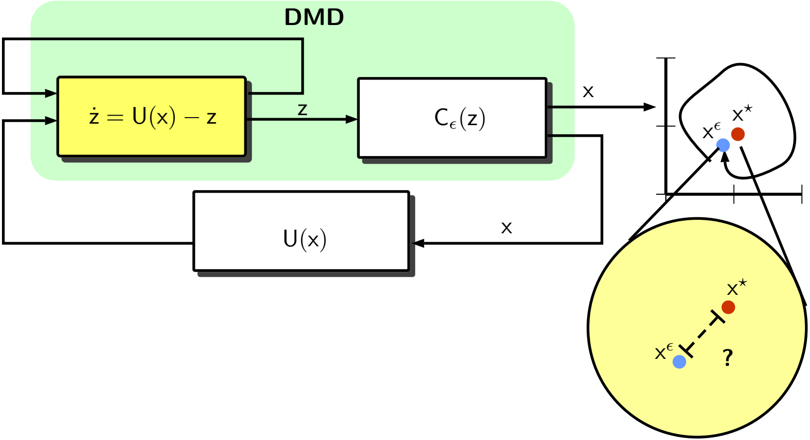


Can $x(t) \to x^\star$ exactly (irrespective of ϵ)?

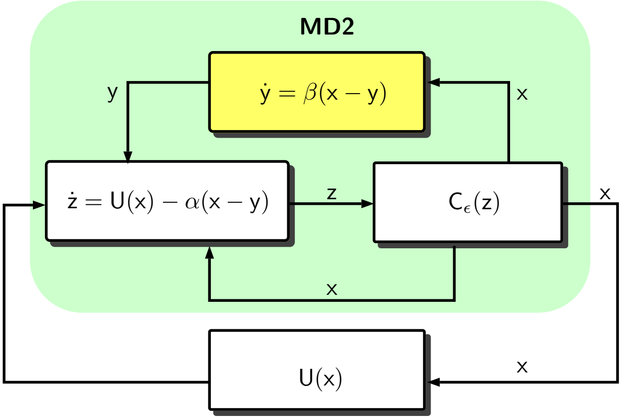
Can we do it without monotonicity?



Source of inexactness: feedback term $-z$ is **non-vanishing** at rest, perturbing the solution away from $x^{\star}$.

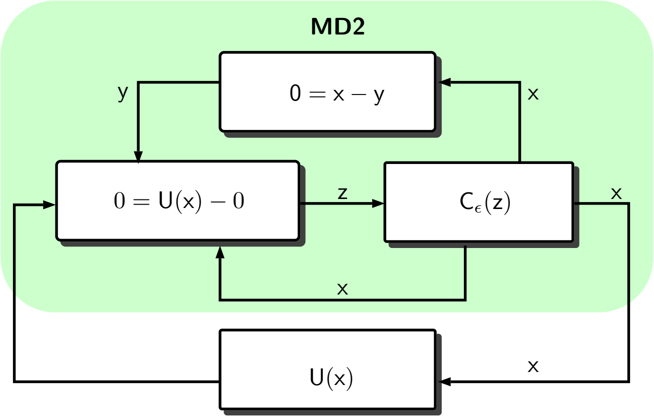


Idea: use a **new feedback** that goes away as x reaches $x^\star$.



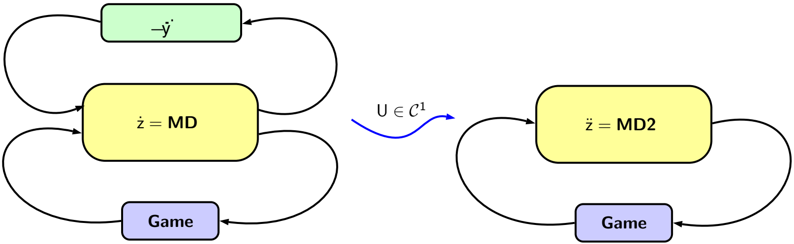
Introduce $\dot{y} = \beta(x - y)$, $\beta > 0$. Call it **second-order MD** (MD2).

Interpretation: Do regularization but without perturbing the optimal solution.



At rest, $0 = U(x) \implies x$ an interior NE.

Why Second-Order MD? What kind of second-order?



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

$$\begin{aligned} \ddot{z} &= \left[\mathbf{J}_{U\circ C_\epsilon}(z) - \alpha\, \mathbf{J}_{C_\epsilon}(z) - \beta \mathbf{I} \right] \dot{z} + \beta U(x), \\ x &= C_\epsilon(z), \end{aligned} \tag{12}$$

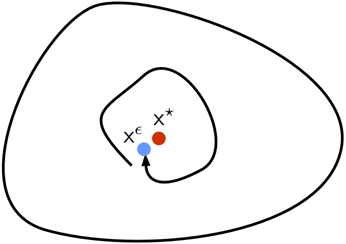
MD2 is Unlike straight second-order integration of payoffs, Laraki & Mertikopoulos'13, which has the same (non-)convergence properties as MD.

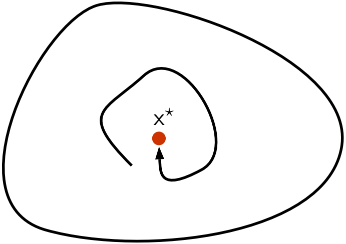
MD2 ≅ dual-space "heavy-ball method", feedback modified, built on passivity-inspired principles.

Main Result: MD2 converges to interior mere VSS

Assumption. Mirror map C_ϵ^p is generated from a regularizer v^p that is $\mathcal{C}^2$ and either Legendre (strictly convex + boundary conditions) or strongly convex.

This assumption covers virtually all mirror maps in the literature.





Theorem[5,6]: $x(t) \to x^\star$ **an interior mere variationally stable state (VSS)**, defined as,

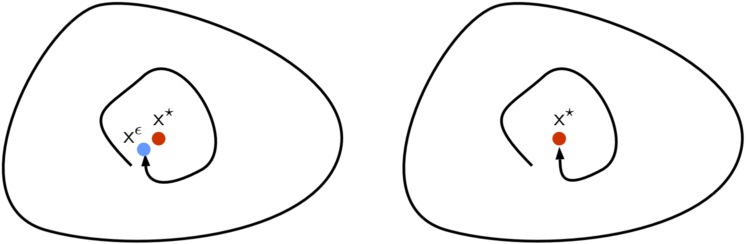
$$\underbrace{U(x)^\top (x - x^\star) \le 0}_{\text{variational inequality}}, \forall x \in \Omega. \tag{4}$$

A type of NE.

---

[5] B. Gao and L. Pavel, "Second-order mirror descent: exact convergence beyond strictly stable equilibria in concave games," in 60th IEEE CDC, 2021.

[6] B. Gao and L. Pavel, "Second-Order Mirror Descent: Convergence in Games Beyond Averaging and Discounting," in IEEE TAC, 2024.



Theorem[5,6]: $x(t) \to x^\star$ **an interior mere variationally stable state (VSS)**, defined as,

$$\underbrace{U(x)^\top(x - x^\star) \le 0}_{\text{variational inequality}}, \forall x \in \Omega. \tag{4}$$

Key advantage: no monotonicity of $\mathcal{G}$ needs to be assumed.

[5] B. Gao and L. Pavel, "Second-order mirror descent: exact convergence beyond strictly stable equilibria in concave games," in 60th IEEE CDC, 2021.

[6] B. Gao and L. Pavel, "Second-Order Mirror Descent: Convergence in Games Beyond Averaging and Discounting," in IEEE TAC, 2024.

MD2 Enjoys No Regret

$$\text{Regret}^p(t) = \max_{y^p \in \Omega^p} \frac{1}{t} \int_0^t \mathcal{U}^p(y^p; x^{-p}(\tau)) - \mathcal{U}^p(x(\tau)) \mathrm{d}\tau \quad (15)$$

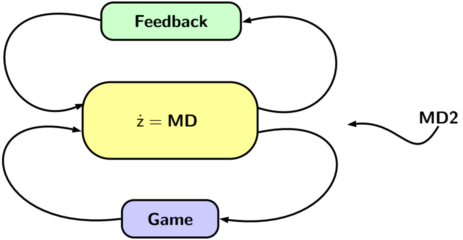
Previous result (Mertikopoulos, Papadimitriou, Piliouras, '18): MD achieves no-regret in finite games,

$$\limsup_{t \to \infty} \text{Regret}^p(t) \leq 0, \forall p, \quad (16)$$

but cannot converge in zero-sum finite games with interior NE.

Our result: MD2 achieves no-regret in concave games with compact action sets (e.g., finite games), and converges in zero-sum games with interior mere VSS.

Second-order variant of the continuous-time mirror descent dynamics (**MD**) $\Rightarrow$ **MD2**.



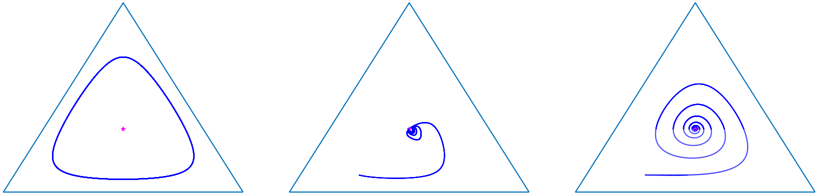
MD2 has the following benefits,

- Converges to Nash equilibrium without global game properties (e.g., (pseudo/quasi-)monotonicity)
- Converges beyond what **MD** is capable of without using additional techniques such as averaging or discounting
- Achieves exponential rate of convergence ($\sim$ slight mod.)
- Achieves no-regret

Simulation: RPS Game

$$\mathcal{A} = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix} \quad \mathcal{U}^p(x^p; x^q) = {x^p}^\top \mathcal{A} x^q$$

$$x^\star = (x^{p\star}), x^{p\star} = (1/3, 1/3, 1/3) \text{ is a mere VSS}$$

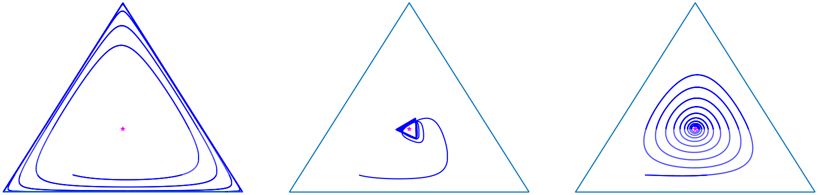


MD MDA MD2

Simulation: RPS Game

$$\mathcal{A} = \begin{bmatrix} 0 & -1.2 & 1 \\ 1 & 0 & -1.2 \\ -1.2 & 1 & 0 \end{bmatrix} \quad \mathcal{U}^p(x^p; x^q) = x^{p\top} \mathcal{A} x^q$$

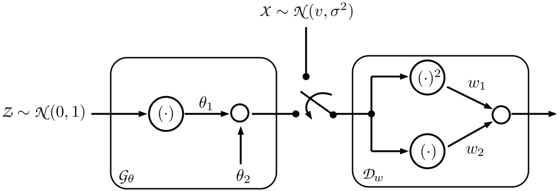
$x^\star = (x^{p\star}), x^{p\star} = (1/3, 1/3, 1/3)$ is not a mere VSS



MD MDA MD2

MD2 still converges (without any tuning) $\Rightarrow$ MD2 is "robust" near the mere VSS!

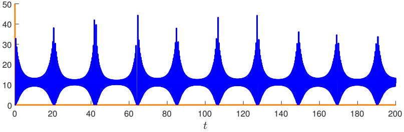
Simulation: Generative Adversarial Network



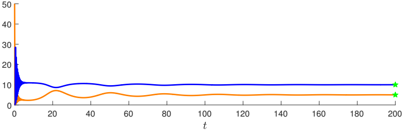
Construct $\theta = x^1, w = x^2$ s.t. $\mathcal{G}_\theta(\mathcal{Z})$ recovers mean and var of $\mathcal{X}$.

$$\mathcal{U}^1(x^1; x^2) = x_1^2(\sigma^2 + v^2 - \sum_{i=1}^2 (x_i^1)^2) + x_2^2(v - x_2^1) = -\mathcal{U}^2(x^2; x^1),$$

Game not merely monotone, but $x^\star = ((\sigma, v), (0,0))$ is mere VSS.
θ_1 , θ_2 are estimates of mean and var.



MD



MD2

Parting Message

- MD2 is the systematic generalization of MD.
  Allows to characterize discrete-time MD2 in semi-bandit and full-bandit setups.

- Future Work:
  - Systematic methods: higher-order for learning in games with coupled constraints (GNE)? (see CDC'24), continuous to discrete-time algorithms while preserving properties?
  - Other questions: distributed settings, additional constraints, discontinuity, non-convexity, even higher-orders, asynchrony, delays, time-varying, more sophisticated players...