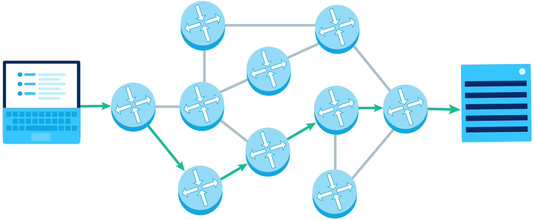
THE IMPACT OF UNCERTAINTY ON LEARNING IN GAMES

Pierre-Louis Cauvin, Davide Legacci, Panayotis Mertikopoulos
Université Grenoble Alpes (LIG)

⟨ SHARP+Foundry Workshop @COLT2025 | June 30, 2025 ⟩

Applications of learning in games

- **Multi-agent learning:** road traffic; network routing; recommender systems; ...



General setup — finite games in continuous time

- **continuous time** $t \geq 0$
- finite number of **players** $i \in \mathcal{N} = \{1, \dots, N\}$
- finite number of **actions** (or **pure strategies**) $\alpha_i \in \mathcal{A}_i = \{1, \dots, A_i\}$
- **action payoffs** $u_i(\alpha_1, \dots, \alpha_N)$

General setup – finite games in continuous time

- **continuous time** $t \geq 0$
- finite number of **players** $i \in \mathcal{N} = \{1, \ldots, N\}$
- finite number of **actions** (or **pure strategies**) $\alpha_i \in \mathcal{A}_i = \{1, \ldots, A_i\}$
- **action payoffs** $u_i(\alpha_1, \ldots, \alpha_N)$

Continuous-time learning procedure

for all $t \geq 0$ **do simultaneously** for all players $i \in \mathcal{N}$:

  Choose **mixed strategy** $x_i \in \mathcal{X}_i := \Delta(\mathcal{A}_i)$ # ensures exploration

  Sample **action** $\alpha_i \sim x_i$

  Observe **mixed payoff vector** $v_i(x) := (u_i(\alpha_i; x_{-i}))_{\alpha_i \in \mathcal{A}_i}$ # perfect feedback

end for

Learning dynamics

Continuous-time exponential weights dynamics

$$\dot{y}_i(t) = v_i(x(t)) \qquad \text{\# cumulative payoff}$$
$$x_i(t) = e^{y_i(t)} / \|e^{y_i(t)}\|_1 \qquad \text{\# soft-max function} \tag{EW}$$

- Continuous-time version of the **multiplicative weights algorithm** — Auer et al., 1995; Sorin, 2009
- (EW) is equivalent to the **replicator dynamics** of Taylor & Jonker (1978) defined by

$$\dot{x}_{i\alpha_i} = x_{i\alpha_i} \left[v_{i\alpha_i}(x) - \sum\nolimits_{\beta_i} x_{i\beta_i} v_{i\beta_i}(x) \right] \tag{RD}$$

Learning dynamics

Continuous-time exponential weights dynamics

$$\dot{y}_i(t) = v_i(x(t)) \qquad \text{\# cumulative payoff}$$
$$x_i(t) = e^{y_i(t)} / \|e^{y_i(t)}\|_1 \qquad \text{\# soft-max function} \qquad \text{(EW)}$$

- Continuous-time version of the **multiplicative weights algorithm** ➻ Auer et al., 1995; Sorin, 2009
- (EW) is equivalent to the **replicator dynamics** of Taylor & Jonker (1978) defined by

$$\dot{x}_{i\alpha_i} = x_{i\alpha_i}\left[v_{i\alpha_i}(x) - \sum\nolimits_{\beta_i} x_{i\beta_i} v_{i\beta_i}(x)\right] \qquad \text{(RD)}$$

Rational behaviors: ➻ Mertikopoulos & Sandholm (2016), Kwon & Mertikopoulos (2017)

- Achieves **no-regret**
- Underperforming actions become **extinct**
- "**Folk theorem**" (convergence vs equilibria)

Stochastic learning dynamics

Exponential weights dynamics

$$dY_{i\alpha_i}(t) = v_{i\alpha_i}(X(t))dt \qquad \text{\# cumulative payoff}$$

$$X_i(t) = e^{Y_i(t)} / \|e^{Y_i(t)}\|_1 \qquad \text{\# update strategy}$$

What is the impact of random perturbations on the exponential weights dynamics?

Stochastic learning dynamics

Stochastic exponential weights dynamics

$$dY_{i\alpha_i}(t) = v_{i\alpha_i}(X(t))dt + \sigma_{i\alpha_i} dW_{i\alpha_i} \quad \text{\# cumulative payoff}$$
$$X_i(t) = e^{Y_i(t)} / \|e^{Y_i(t)}\|_1 \quad \text{\# update strategy} \tag{S-EW}$$

- $W(t) := (W_{i\alpha_i}(t))_{i_\alpha \in \mathcal{A}_i, i \in \mathcal{N}}$ is a standard Brownian motion # continuous uncorrelated noise
- $\sigma_{i\alpha_i} > 0$ is the level of noise
- (S-EW) understood as a **stochastic differential equation**

Stochastic learning dynamics

Stochastic exponential weights dynamics

$$dY_{i\alpha_i}(t) = v_{i\alpha_i}(X(t))dt + \sigma_{i\alpha_i} dW_{i\alpha_i} \quad \text{\# cumulative payoff}$$
$$X_i(t) = e^{Y_i(t)} / \|e^{Y_i(t)}\|_1 \quad \text{\# update strategy} \tag{S-EW}$$

- $W(t) := (W_{i\alpha_i}(t))_{i_\alpha \in \mathcal{A}_i, i \in \mathcal{N}}$ is a standard Brownian motion # continuous uncorrelated noise
- $\sigma_{i\alpha_i} > 0$ is the level of noise
- (S-EW) understood as a **stochastic differential equation**

Related works:

- ☞ (S-EW) already studied by Mertikopoulos & Moustakas (2010) and Bravo & Mertikopoulos (2017)
- ☞ Other stochastic variants of (RD): Foster & Young (1990) (pairwise imitation), Fudenberg & Harris (1992) (biological reproduction)
- Further works by Cabrales (2000), Imhof (2005), Hofbauer & Imhof (2009), Mertikopoulos & Viossat (2016), Engel & Piliouras (2023), ...

Uncertainty favors extremes

Theorem — Evolution close to pures

In **any game** and for **any level of noise**, every player reaches an arbitrarily small neighborhood of one of their pures strategies in finite time

Uncertainty favors extremes

Theorem — Evolution close to pures

In **any game** and for **any level of noise**, every player reaches an arbitrarily small neighborhood of one of their pures strategies in finite time

Corollary — Limits of (S-EW)

The only possible limits of (S-EW) are **pure strategies**

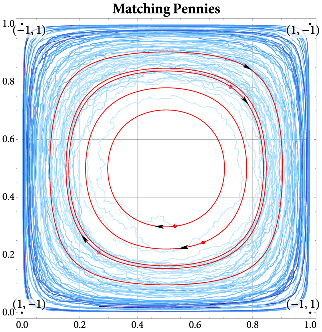
Uncertainty favors extremes

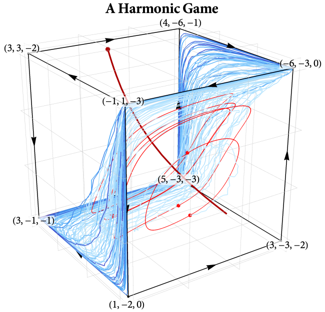
Theorem — Evolution close to pures

In **any game** and for **any level of noise**, every player reaches an arbitrarily small neighborhood of one of their pures strategies in finite time

Corollary — Limits of (S-EW)

The only possible limits of (S-EW) are **pure strategies**





"Uncertainty favors extreme decisions"

Stability & attractiveness of pure strategies

Which pure strategies are stable and attracting for (S-EW)?

Stability & attractiveness of pure strategies

Which pure strategies are stable and attracting for (S-EW)?

- $\mathcal{B} = \prod_{i \in \mathcal{N}} \mathcal{B}_i$ **product of pure strategies** with $\mathcal{B}_i \subseteq \mathcal{A}_i$
- $S = \operatorname{span}(\mathcal{B})$ face of strategies with **support** in $\mathcal{B}$ $\quad$ # $\operatorname{supp}(x_i) := \{\alpha_i \in \mathcal{A}_i : x_{i\alpha_i} > 0\}$

☞ S is **stochastically asymptotically stable** if trajectories starting nearby remain nearby and eventually converge to it with arbitrarily high probability

Stability & attractiveness of pure strategies

Which pure strategies are stable and attracting for (S-EW)?

- $\mathcal{B} = \prod_{i \in \mathcal{N}} \mathcal{B}_i$ **product of pure strategies** with $\mathcal{B}_i \subseteq \mathcal{A}_i$
- $S = \operatorname{span}(\mathcal{B})$ face of strategies with **support** in $\mathcal{B}$ $\quad \# \operatorname{supp}(x_i) := \{\alpha_i \in \mathcal{A}_i : x_{i\alpha_i} > 0\}$
- ☞ S is **stochastically asymptotically stable** if trajectories starting nearby remain nearby and eventually converge to it with arbitrarily high probability

Theorem — Stable $\Longleftrightarrow$ club

For **any level of noise**, S is stochastically asymptotically stable if and only if it is **closed under better replies**

- ☞ S is **closed under better replies** if

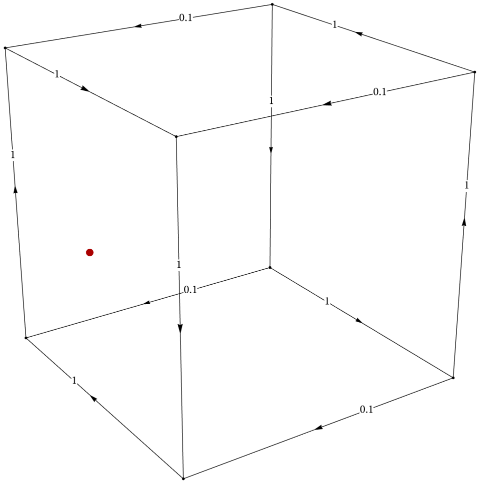
$$u_i(\beta_i; \alpha_{-i}) < u_i(\alpha) \quad \text{for all } \alpha \in \mathcal{B}, \beta_i \notin \mathcal{B}_i, i \in \mathcal{N}$$

- Extends results of Ritzberger & Weibull (1995) and Boone & Mertikopoulos (2023)

Closedness under better replies

- S is **closed under better replies** if

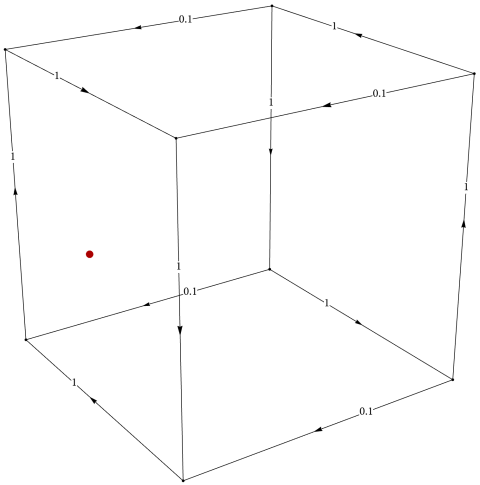
$$u_i(\beta_i; \alpha_{-i}) < u_i(\alpha) \quad \text{for all } \alpha \in \mathcal{B}, \beta_i \notin \mathcal{B}_i, i \in \mathcal{N}$$

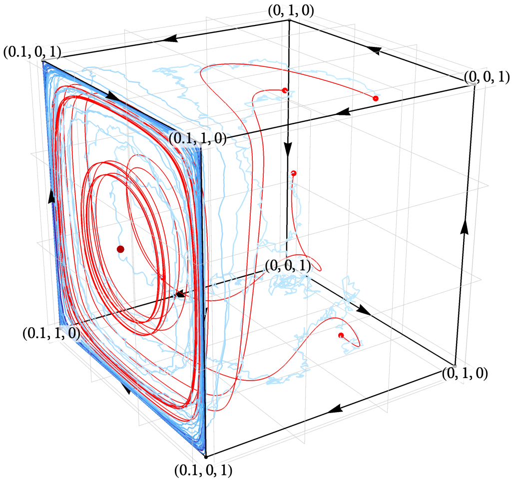


Closedness under better replies

☞ S is **closed under better replies** if

$$u_i(\beta_i; \alpha_{-i}) < u_i(\alpha) \quad \text{for all } \alpha \in \mathcal{B}, \beta_i \notin \mathcal{B}_i, i \in \mathcal{N}$$





Consequences for recurrent dynamics

2-players zero-sum game: $u_1(\alpha, \beta) = -u_2(\alpha, \beta)$ for all $\alpha \in \mathcal{A}_1, \beta \in \mathcal{A}_2$

☞ Trajectories of (EW) are **recurrent** if game admits fully mixed Nash equilibrium ➻ Mertikopoulos et al. (2018)

Consequences for recurrent dynamics

2-players zero-sum game: $u_1(\alpha, \beta) = -u_2(\alpha, \beta)$ for all $\alpha \in \mathcal{A}_1, \beta \in \mathcal{A}_2$

- ☞ Trajectories of (EW) are **recurrent** if game admits fully mixed Nash equilibrium — Mertikopoulos et al. (2018)

Theorem 3 — Failure of recurrence under uncertainty

In **any** 2-players zero-sum game (with a fully mixed Nash equilibrium) and for **any level of noise**, players' choices converge "on average" toward the strategy space boundary.

- ☞ Mathematically, $\sum_{i \in \mathcal{N}} \mathbb{E}[\mathrm{KL}(q_i, X_i(t))] \to \infty$ for some fully mixed $q \in \operatorname{ri} \mathcal{X}$ — # $\mathcal{X} := \prod_{i \in \mathcal{N}} \mathcal{X}_i$
- ⚠ Trajectories **do not** necessarily converge **almost surely** to the boundary

Consequences for recurrent dynamics

2-players zero-sum game: $u_1(\alpha,\beta) = -u_2(\alpha,\beta)$ for all $\alpha \in \mathcal{A}_1, \beta \in \mathcal{A}_2$

☞ Trajectories of (EW) are **recurrent** if game admits fully mixed Nash equilibrium ➡ Mertikopoulos et al. (2018)

Theorem 3 — Failure of recurrence under uncertainty

In **any** 2-players zero-sum game (with a fully mixed Nash equilibrium) and for **any level of noise**, players' choices converge "on average" toward the strategy space boundary.

☞ Mathematically, $\sum_{i\in\mathcal{N}} \mathbb{E}[\mathrm{KL}(q_i, X_i(t)] \to \infty$ for some fully mixed $q \in \mathrm{ri}\,\mathcal{X}$ # $\mathcal{X} := \prod_{i\in\mathcal{N}} \mathcal{X}_i$

⚠ Trajectories **do not** necessarily converge **almost surely** to the boundary

Theorem 4 — Irreducibility of boundary

In **any** 2-players zero-sum game (with a fully mixed Nash equilibrium) and for **any level of noise**, no proper product of pure strategies span a stochastically asymptotically stable face.

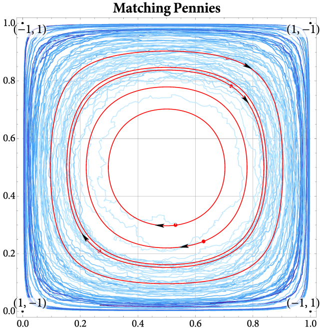
Consequences for recurrent dynamics

Theorem 3 — Failure of recurrence under uncertainty

In **any** 2-players zero-sum game (with a fully mixed Nash equilibrium) and for **any level of noise**, players' choices converge "on average" toward the strategy space boundary.

Theorem 4 — Irreducibility of boundary

In **any** 2-players zero-sum game (with a fully mixed Nash equilibrium) and for **any level of noise**, no proper product of pure strategies span a stochastically asymptotically stable face.



Concluding remarks

Framework extension

General regularizer & correlated noise:

$$\begin{aligned} dY_i(t) &= v_i(X(t))dt + dM_i(t) \\ X_i(t) &= Q_i(Y_i(t)) \end{aligned} \tag{S-FTRL}$$

Framework extension

General regularizer & correlated noise:

$$
\begin{aligned}
dY_i(t) &= v_i(X(t))dt + dM_i(t) \\
X_i(t) &= Q_i(Y_i(t))
\end{aligned}
\qquad \text{(S-FTRL)}
$$

Harmonic games: there exist weights $m_i > 0$ and a fully mixed strategy $q \in \operatorname{ri} \mathcal{X}$ such that ➜ Legacci et al. (2024)

$$\sum_{i \in \mathcal{N}} m_i \langle v_i(x), x_i - q_i \rangle \quad \text{for all } x \in \mathcal{X}$$

Concluding remarks

Framework extension

General regularizer & correlated noise:

$$
\begin{aligned}
dY_i(t) &= v_i(X(t))dt + dM_i(t) \\
X_i(t) &= Q_i(Y_i(t))
\end{aligned}
\tag{S-FTRL}
$$

Harmonic games: there exist weights $m_i > 0$ and a fully mixed strategy $q \in \operatorname{ri} \mathcal{X}$ such that ➾ Legacci et al. (2024)

$$
\sum_{i \in \mathcal{N}} m_i \langle v_i(x), x_i - q_i \rangle \quad \text{for all } x \in \mathcal{X}
$$

Open directions:

- Vanishing learning rate?
- Discontinuous noise?
- Discrete time?
- Continuous action space?

References I

[1] Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pp. 322–331, October 1995.

[2] Boone, V. and Mertikopoulos, P. The Equivalence of Dynamic and Strategic Stability under Regularized Learning in Games. In *Thirty-Seventh Conference on Neural Information Processing Systems*, November 2023.

[3] Bravo, M. and Mertikopoulos, P. On the robustness of learning in games with stochastically perturbed payoff observations. *Games and Economic Behavior*, 103:41–66, May 2017.

[4] Cabrales, A. Stochastic Replicator Dynamics. *International Economic Review*, 41(2):451–481, 2000.

[5] Engel, M. and Piliouras, G. A stochastic variant of replicator dynamics in zero-sum games and its invariant measures. *Physica D: Nonlinear Phenomena*, 456:133940, December 2023.

[6] Foster, D. and Young, P. Stochastic evolutionary game dynamics. *Theoretical Population Biology*, 38(2):219–232, October 1990.

[7] Fudenberg, D. and Harris, C. Evolutionary dynamics with aggregate shocks. *Journal of Economic Theory*, 57(2):420–441, August 1992.

[8] Hofbauer, J. and Imhof, L. A. Time averages, recurrence and transience in the stochastic replicator dynamics. *The Annals of Applied Probability*, 19(4):1347–1368, August 2009.

[9] Imhof, L. A. The long-run behavior of the stochastic replicator dynamics. *The Annals of Applied Probability*, 15(1B):1019–1045, February 2005.

[10] Kwon, J. and Mertikopoulos, P. A continuous-time approach to online optimization. *Journal of Dynamics and Games*, 4(2):125–148, 2017.

[11] Legacci, D., Mertikopoulos, P., Papadimitriou, C., Piliouras, G., and Pradelski, B. No-regret Learning in Harmonic Games: Extrapolation in the Face of Conflicting Interests. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, November 2024.

[12] Mertikopoulos, P. and Moustakas, A. L. The emergence of rational behavior in the presence of stochastic perturbations. *The Annals of Applied Probability*, 20(4):1359–1388, August 2010.

References II

[13] Mertikopoulos, P. and Sandholm, W. H. Learning in Games via Reinforcement and Regularization. *Mathematics of Operations Research*, 41(4): 1297–1324, November 2016.

[14] Mertikopoulos, P. and Viossat, Y. Imitation dynamics with payoff shocks. *International Journal of Game Theory*, 45(1):291–320, March 2016.

[15] Mertikopoulos, P., Papadimitriou, C., and Piliouras, G. Cycles in Adversarial Regularized Learning. *Proceedings of the 2018 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 2703–2717, January 2018.

[16] Ritzberger, K. and Weibull, J. W. Evolutionary Selection in Normal-Form Games. *Econometrica*, 63(6):1371–1399, 1995.

[17] Sorin, S. Exponential weight algorithm in continuous time. *Mathematical Programming*, 116(1):513–528, January 2009.

[18] Taylor, P. D. and Jonker, L. B. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40(1):145–156, July 1978.

Stochastic FTRL dynamics

$$dY_i(t) = v_i(X(t))dt + dM_i(t) \quad \text{\# cumulative payoff}$$
$$X_i(t) = Q_i(Y_i(t)) \quad \text{\# update strategy} \qquad \text{(S-FTRL)}$$

- $Q_i(y) := \arg\max_{x_i \in \mathcal{X}_i}\{\langle y, x\rangle - h_i(x)\}$ **regularized best response map** $\rightsquigarrow$ ensures exploration
- $M_i(t)$ **continuous square-integrable martingale** $\rightsquigarrow$ captures all sources of randomness

Assumptions:

- $h_i(x_i) = \sum_{\alpha_i} \theta_i(x_{i\alpha_i})$ for $\theta_i : (0,1) \to \mathbb{R}$ **smooth**, **strongly convex** and **steep** at 0
- $dM_i(t) = \sigma_i(X(t))dW(t)$ for some Lipschitz function $\sigma_i : \mathcal{X} \to \mathbb{R}^{A_i \times d}$ # $W(t)$ d-dimensional BM
- The smallest eigenvalue of $\Sigma := \sigma\sigma^T$ is strictly positive uniformly on $\mathcal{X}$ # persistent noise

Examples of regularizers:

- Negative Gibbs entropy $h(z) = z \log z \rightsquigarrow$ **soft-max function**
- Fractional (Tsallis) entropy $h(z) = -4\sqrt{z}$

Failure of recurrence

Theorem 3 — Failure of recurrence under uncertainty, explicit version

In **any** 2-players zero-sum game (with a fully mixed Nash equilibrium) and for **any level of noise**, trajectories of (S-EW) verify:

1. $\sum_{i \in \mathcal{N}} \mathbb{E}[\mathrm{KL}(q_i, X_i(t))] \to \infty$ for some $q \in \mathrm{ri}\,\mathcal{X}$;
2. The first exit time $\tau_{\mathcal{K}}$ from a compact subset $\mathcal{K}$ is **finite in expectation** if $\mathcal{K}$ is not connected to $\mathrm{bd}\,\mathcal{X}$;
3. The first exit time $\tau_{\mathcal{K}}$ from a compact subset $\mathcal{K}$ is **infinite in expectation** if $\mathcal{K}$ contains $\mathrm{bd}\,\mathcal{X}$.

Convergence on average vs almost-surely:

- 2×2 game with $u_1 = u_2 = 0 \Longrightarrow$ **no almost-sure convergence to boundary**
- 3×3 game with $u_1 = u_2 = 0 \Longrightarrow$ **almost-sure convergence to boundary**