Adaptively Perturbed Mirror Descent for Learning in Games

Kenshi Abe[1,2], Kaito Ariu[1], Mitsuki Sakamoto[1], Atsushi Iwasaki[2]

[1]CyberAgent, Japan [2]University of Electro-Communications, Japan

arXiv:

Introduction

Learning in Games

- This paper proposes a payoff perturbation technique for the Mirror Descent algorithms to find a Nash equilibrium (NE) in monotone games.
- ***N*-Player Monotone Games**
  - A family of games including: Cournot competition [Bravo et al. 2018]; λ-cocoercive games [Lin et al., 2020]; Concave-convex games and zero-sum polymatrix games [Cai & Daskalakis, 2011; Cai et al., 2016]
- Various learning algorithms have been developed and scrutinized to compute NE efficiently.

Mirror Descent and Average-Iterate Convergence

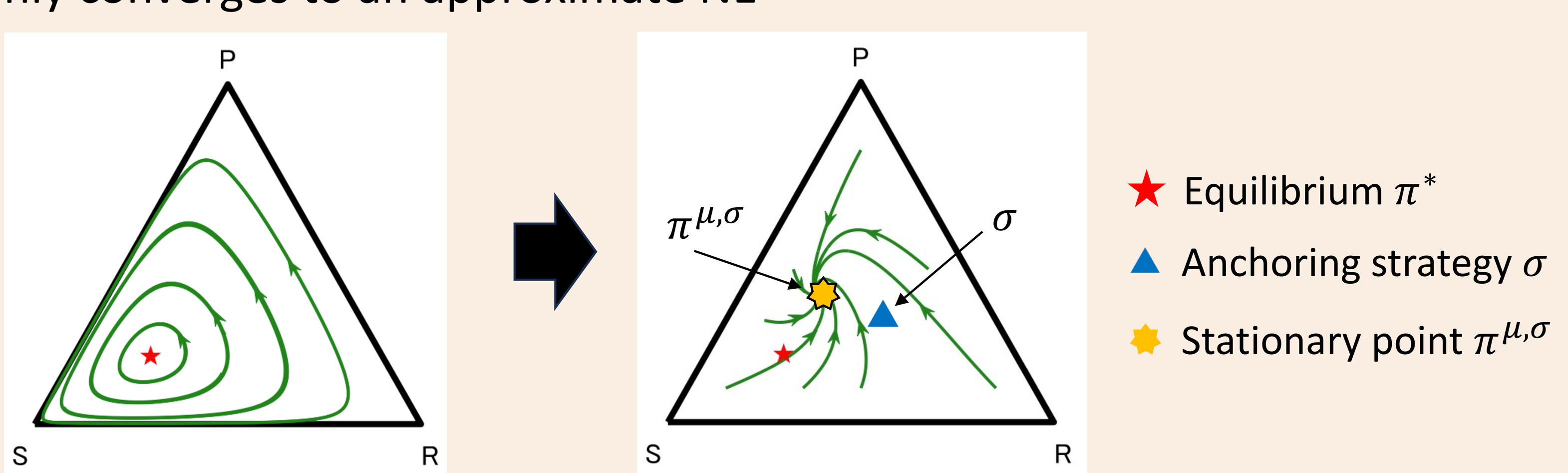
- Mirror Descent (MD) updates the strategy π_i^t based on the gradient feedback $\widehat{\nabla}_{\pi_i} v_i(\pi^t)$

$$\underbrace{\pi_i^{t+1}}_{\text{Next strategy}} = \arg\max_{x\in\mathcal{X}_i}\{\underbrace{\eta_t\langle\widehat{\nabla}_{\pi_i} v_i(\pi^t), x\rangle}_{\text{Choose strategies with higher expected payoffs}} - \underbrace{D_\psi(x,\pi_i^t)}_{\text{Does not move too far away from the current strategy}}\}$$

  - η_t: learning rate
  - $D_\psi(\pi_i,\pi_i')$: Bregman divergence with strongly convex function
- The average strategies $\frac{1}{T}\sum_{t=1}^T \pi^t$ converge to NE (**average-iterate convergence**). However, the actual trajectory of π^t may fail to converge [Mertikopoulos et al., 2018].

Last-Iterate Convergence and Perturbation Approach

- **Last-Iterate Convergence**
  - The updated strategy profile itself converges to NE
  - Optimistic learning algorithms are representative algorithms that achieve last-iterate convergence [Daskalakis et al., 2018; Daskalakis & Panageas, 2019; Mertikopoulos et al., 2019; Wei et al., 2021]. However, they perform suboptimally with feedback contaminated by some noise.
- **Payoff Perturbation Approach** (e.g., [Facchinei & Pang, 2003])
  - Introducing strongly convex penalties to the players' payoff functions
  - Only converges to an approximate NE



w/o perturbation ($\mu = 0$) wtih perturbation ($\mu > 0$)

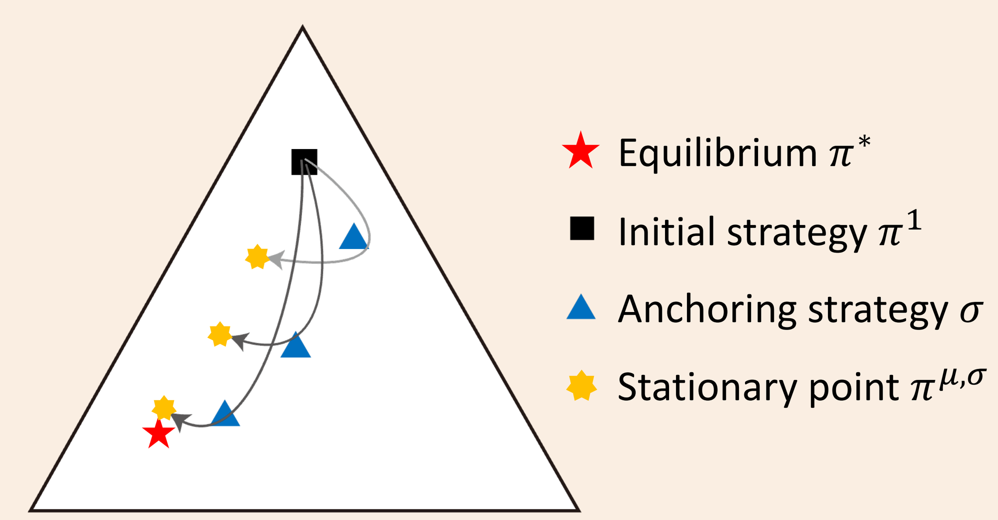
Proposed Algorithm

Adaptively Perturbed MD (APMD)

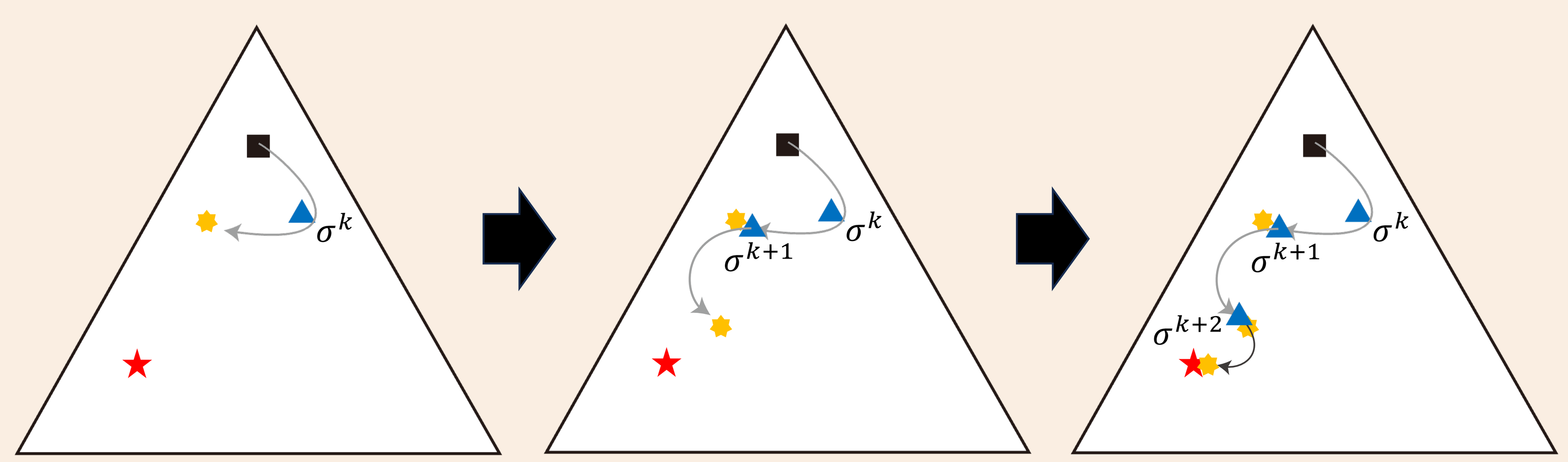
- APMD perturbs the payoff function v_i by the divergence function $G(\cdot,\cdot)$ between the current and anchoring strategies (i.e., π_i^t and σ_i)

$$\pi_i^{t+1} = \arg\max_{x\in\mathcal{X}_i}\{\eta_t\langle\widehat{\nabla}_{\pi_i} v_i(\pi^t) - \underbrace{\mu}_{\text{Perturbation strength}}\overbrace{\nabla_{\pi_i} G(\pi_i^t, \underbrace{\sigma_i}_{\text{Anchoring strategy}})}^{\text{Strongly convex divergence function}}, x\rangle - D_\psi(x,\pi_i^t)\}$$

  - The strong convexity of G enables the current strategy π^t to converge to a stationary point $\pi^{\mu,\sigma}$, an approximate equilibrium
  - To ensure convergence to NE, the magnitude of perturbation requires careful adjustment [Koshal et al., 2010; Tatarenko & Kamgarpour, 2019; Liu et al., 2023]
- APMD adaptively determines the magnitude of perturbation, while maintaining the perturbation strength parameter μ constant
  - This is inspired by the fact that different anchoring strategies σ lead to different stationary points $\pi^{\mu,\sigma}$



  - When the anchoring strategy σ is close to NE, the corresponding stationary point $\pi^{\mu,\sigma}$ is also close to NE
- In order to bring the anchoring strategy σ closer to NE, APMD re-initializes σ at a predefined interval T_σ by the current strategy π^t (i.e., $\sigma^k \leftarrow \pi^t$)
  - This means that σ^k is overrode by the approximation of π^{μ,σ^k}
  - Although the same idea is utilized by [Perolat et al., 2021; Abe et al., 2023], they provide the convergence in an asymptotic manner



Theoretical/Experimental Results

Last-Iterate Convergence Results

- A metric of proximity to NE:

$$\mathrm{GAP}(\pi) := \max_{\tilde{\pi}\in\mathcal{X}} \sum_{i=1}^N \langle \nabla_{\pi_i} v_i(\pi), \tilde{\pi}_i - \pi_i\rangle.$$

- Consider a setting where both D_ψ and G is set to the squared ℓ^2-distance, i.e., $D_\psi(\pi_i,\pi_i') = G(\pi_i,\pi_i') = \frac{1}{2}\|\pi_i - \pi_i'\|^2$

Theorem 1 (Full Feedback: $\widehat{\nabla}_{\pi_i} v_i(\pi^t) = \nabla_{\pi_i} v_i(\pi^t)$)

If we set $T_\sigma = \Theta(\ln T)$, then π^T converges to NE at the rate of $\tilde{O}(1/\sqrt{T})$:

$$\mathrm{GAP}(\pi^T) = O\left(\frac{\ln T}{\sqrt{T}}\right).$$

Theorem 2 (Noisy Feedback: $\widehat{\nabla}_{\pi_i} v_i(\pi^t) = \nabla_{\pi_i} v_i(\pi^t) + \xi_i^t$)

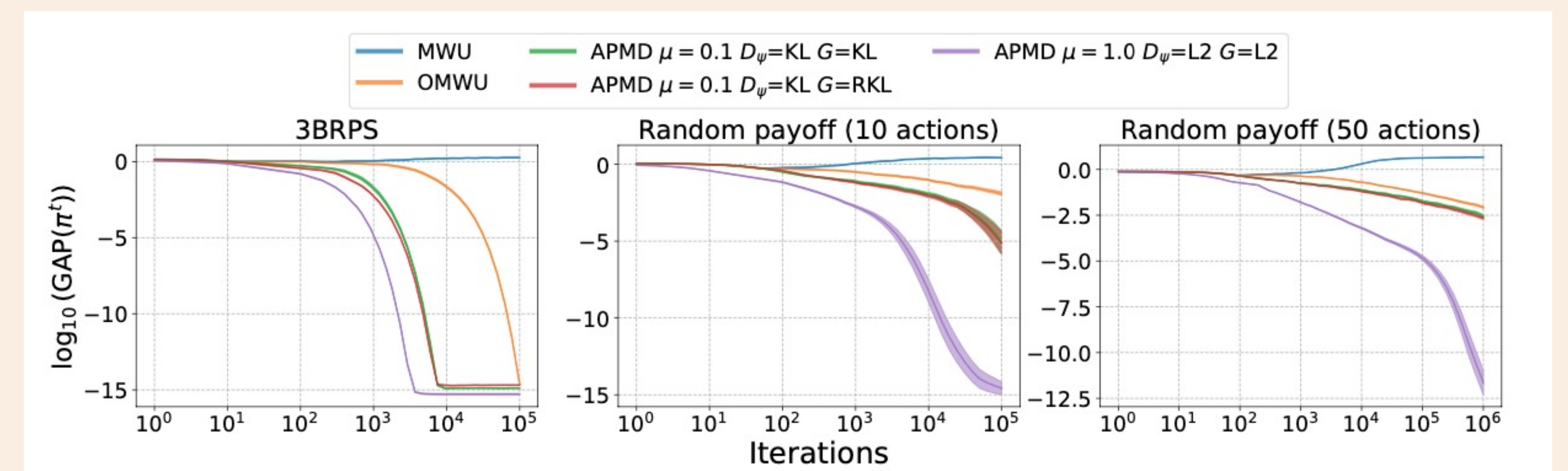
If we set $T_\sigma = \Theta(T^{4/5})$, then π^T converges to NE:

$$\mathbb{E}[\mathrm{GAP}(\pi^T)] = O\left(\frac{\ln T}{T^{1/10}}\right).$$

- **Asymptotic convergence beyond squared ℓ^2-distance can be achieved**
  - Bregman divergence, Reverse KL, α-divergence, Rényi divergence

Experimental Results

- **Full Feedback (Three-Player Biased Rock-Paper-Scissors)**



- **Noisy Feedback (Three-Player Biased Rock-Paper-Scissors)**