Boosting Perturbed Gradient Ascent for Last-Iterate Convergence in Games

Kenshi Abe[1,2], Mitsuki Sakamoto[1], Kaito Ariu[1], Atsushi Iwasaki[2]
[1]CyberAgent, Japan [2]University of Electro-Communications, Japan

arXiv:

Introduction

Learning in Games

- This paper proposes a **novel payoff-perturbed learning algorithm, Gradient Ascent with Boosting Payoff Perturbation (GABP)**, enabling it to achieve fast convergence to Nash equilibrium (NE) in games.
- **N-Player Monotone Games**
  - ☐ A family of games including: Concave-convex games; Cournot competition [Monderer & Shapley, 1996] and zero-sum polymatrix games [Cai & Daskalakis, 2011; Cai et al., 2016]
- **Average-Iterate Convergence**
  - ☐ In online learning setting, learning algorithms update the strategy π_i^t based on the gradient feedback $\widehat{\nabla}_{\pi_i} v_i(\pi^t)$, as in algorithms like Gradient Ascent:

Learning rate

$$\boxed{\pi_i^{t+1}} = \prod_{\mathcal{X}_i}\left[\pi_i^t + \eta_t \boxed{\widehat{\nabla}_{\pi_i} v_i(\pi^t)}\right].$$

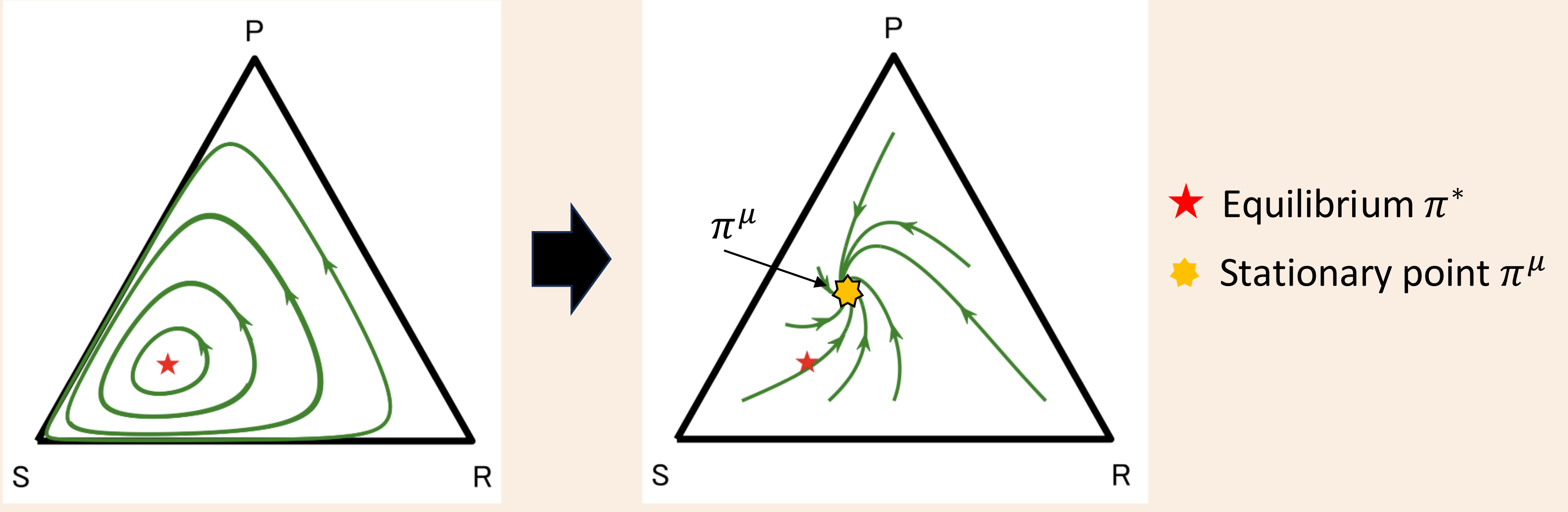
Next strategy

Gradient feedback

  - ☐ In many learning algorithms, the average strategies $\frac{1}{T}\sum_{t=1}^{T}\pi^t$ converge to NE.
  - ☐ However, the actual trajectory of π^t may fail to converge [Mertikopoulos et al., 2018].
- **Last-Iterate Convergence**
  - ☐ The updated strategy profile itself converges to NE
  - ☐ A stronger and more desirable property than average-iterate convergence

Existing Algorithms

- **Optimistic Learning Algorithms**
  - ☐ Representative algorithms that achieve last-iterate convergence [Daskalakis et al., 2018; Daskalakis & Panageas, 2019; Mertikopoulos et al., 2019]
  - ☐ However, they perform suboptimally with feedback contaminated by some noise.
- **Payoff-Perturbed Learning Algorithms** (e.g., [Facchinei & Pang, 2003])
  - ☐ Introduces strongly convex penalties to the players' payoff functions
  - ☐ Only converges to an approximate NE



without perturbation ($\mu = 0$)

with perturbation ($\mu > 0$)

Proposed Algorithm

Gradient Ascent with Boosting Payoff Perturbation

- GABP perturbs the payoff function v_i by the distance $||\pi_i^t - \sigma_i^k||^2$ between the current and anchoring strategies (i.e., π_i^t and σ_i^k) :

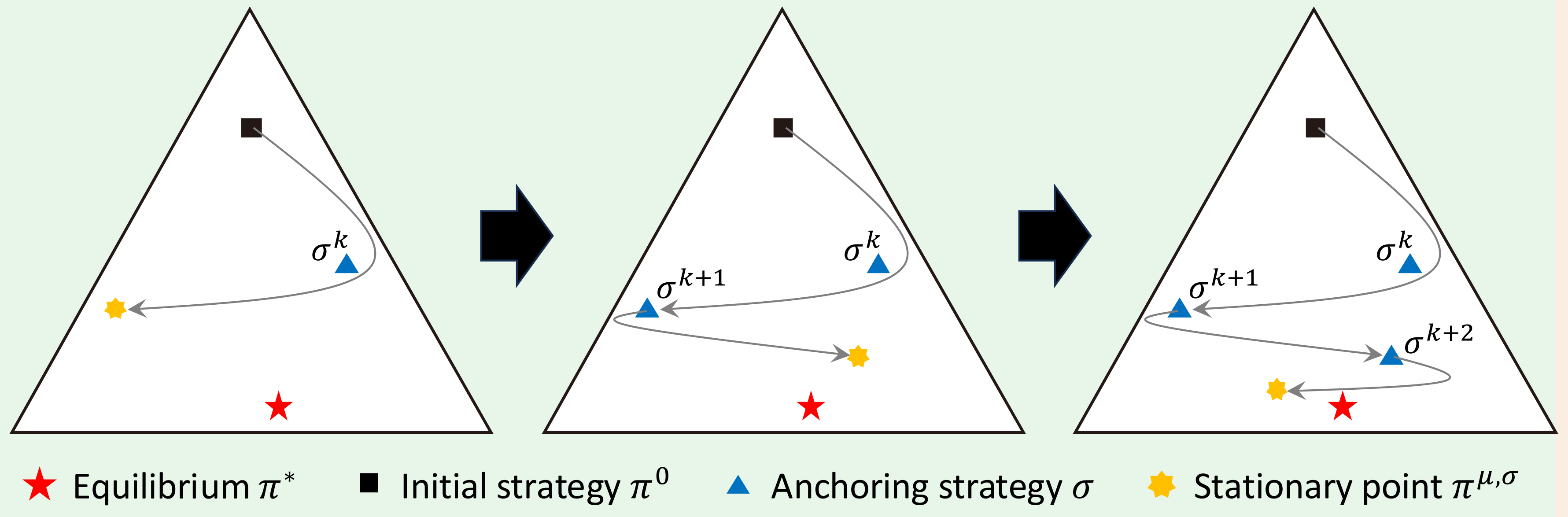
Perturbation payoff

$$\pi_i^{t+1} = \prod_{\mathcal{X}_i}\left[\pi_i^t + \eta_t\left(\widehat{\nabla}_{\pi_i} v_i(\pi^t) - \mu\boxed{(\pi_i^t - \sigma_i^k)}\right)\right].$$

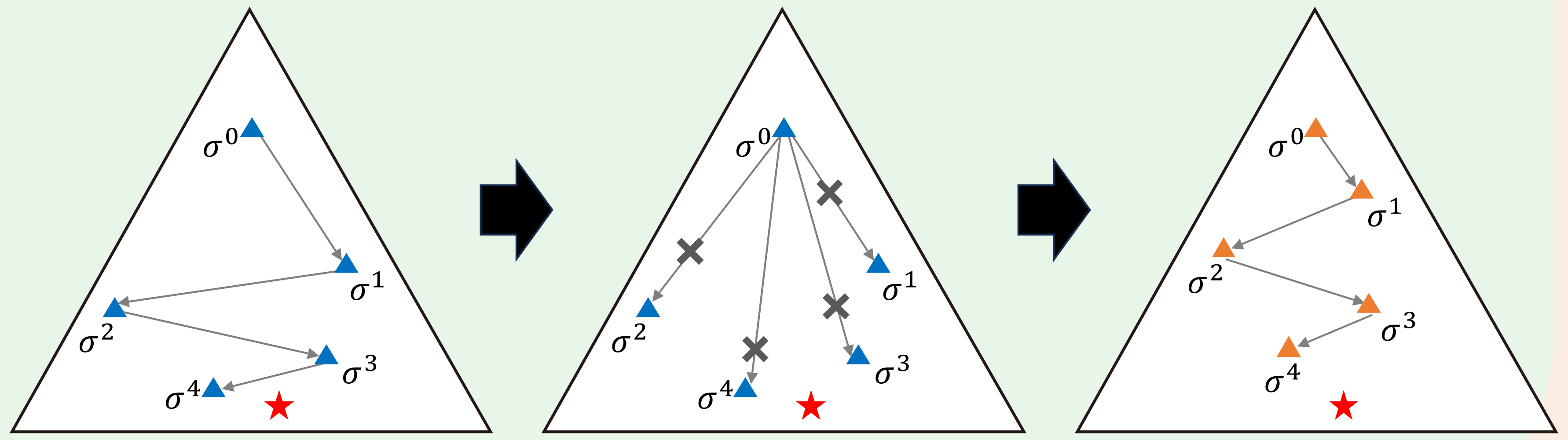
Perturbation strength

Anchoring strategy

- **Anchoring Strategy Update**
  - ☐ To ensure convergence to NE, the anchoring strategy σ^k is updated based on the current strategy π^t at a predefined interval T_σ.
  - ☐ Although existing approaches directly overwrite σ^k with π^t [Perolat et al., 2021; Abe et al., 2023; 2024], the anchoring strategy can change dramatically, which often results in instability in the learning dynamics.



- **Blended Anchoring Strategy**
  - ☐ GABP employs a convex combination $\sigma^k = \frac{k\pi^t+\pi^0}{k+1}$ of the current strategy π^t and the initial strategy π^0 as an anchoring strategy.
  - ☐ This enables the anchoring strategy σ^k to gradually evolve from π^0 to π^t, **which contributes to both stabilization and faster convergence!**



Trajectory of anchoring strategies in existing approaches

Convex combination

GABP (Ours)

Theoretical/Experimental Results

Last-Iterate Convergence Results

- A metric of proximity to NE:

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

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 $\tilde{O}(1/T)$:

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

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 $T_\sigma = \Theta(T^{6/7})$, then π^T converges to NE:

$$\mathbb{E}[\mathbf{GAP}(\boldsymbol{\pi}^T)] = \boldsymbol{O}\left(\frac{\ln T}{T^{1/7}}\right).$$

- **GABP achieves improved rates over APGA under both full and noisy feedback!**

Algorithm	Full Feedback	Noisy Feedback
Optimistic Gradient (OG) [Golowich et al., 2020b; Gorbunov et al., 2022; Cai et al., 2022a]	$O(1/\sqrt{T})$	✗
Accelerated Optimistic Gradient (AOG) [Cai & Zheng, 2023]	$O(1/T)$	✗
Adaptively Perturbed Gradient Ascent (APGA) [Abe et al., 2024]	$\tilde{O}(1/\sqrt{T})$	$\tilde{O}(1/T^{1/10})$
GABP (Ours)	$\tilde{\boldsymbol{O}}(\mathbf{1}/\boldsymbol{T})$	$\tilde{\boldsymbol{O}}(\mathbf{1}/\boldsymbol{T}^{\mathbf{1/7}})$

Experimental Results

- GABP performs better under both full and noisy feedback!