Consensus Multiplicative Weights Update: Learning to Learn using Projector-based Game Signatures

Nelson Vadori[1], Rahul Savani[2], Thomas Spooner[1], Sumitra Ganesh[1]
[1] J.P. Morgan AI Research
[2] University of Liverpool

Overview

1. Consensus Multiplicative Weights Update: new algorithm with local last-iterate convergence guarantees to Nash Equilibria for zero-sum bimatrix games.
   - *generalizes consensus optimization from the unconstrained to the constrained case.*

2. Learning the coefficients of the above update rule for general bimatrix games (non zero-sum).
   - *can we learn last-iterate convergence in 2 Player games using a featurization of such games, the game signature?*

Setting

- constrained case: x, y players' mixed strategies in probability simplex.
- 2P bimatrix game, payoffs $f_1(x, y) = x^T A y$, $f_2(x, y) = x^T B y$.
  - *First focus on the zero-sum case $B = -A$*
- Existing algorithms displaying last-iterate convergence in the zero-sum case, with constant step-size:
  - *Optimistic Multiplicative Weights Update,*
  - *Optimistic mirror descent with various projection methods.*

Consensus Multiplicative Weights Update

- Naively adding to the gradient the 2^{nd} order term AA^T as in the unconstrained case does **not** work!
  - *Nash Equilibria are not fixed points*
  - *A is typically not invertible (Rock-Paper-Scissors)*
    - problematic regarding the Jacobian spectral radius condition

Consensus Multiplicative Weights Update

- Define simplex-Hessians $H_x := A^T \mathrm{diag}(x) A$, $H_y := A \mathrm{diag}(y) A^T$.
    - *spectrum "rescaling"* $sp(H_x) = sp(\mathrm{diag}(x) A A^T)$.
- simplex-Hessians make Nash Equilibria be fixed points
- Consensus Multiplicative Weights Update (CMWU) update:
    - h learning rate, ϵ Hessian coefficient.
    - Multiplicative Weights projection P

$$x_{t+1} = P_{x_t}\left(hAy_t - h\epsilon H_{y_t} x_t\right)$$
$$y_{t+1} = P_{y_t}\left(-hA^T x_t - h\epsilon H_{x_t} y_t\right)$$

- What about invertibility of H_x and H_y?
    - *rock-paper-scissors still not invertible*

Consensus Multiplicative Weights Update

- **Definition:** a matrix A is weakly V-invertible if:

$$X \in V \Rightarrow AX \neq 0, \quad \text{i.e.} \quad Ker(A) \cap V = \{0\}$$

- coincides with classical invertibility when A square and $V = \mathbb{R}^d$
- We require A to be weakly Z-invertible
  - *Z space of vectors that sum to 0*
  - *Rock-Paper-Scissors is weakly Z-invertible.*
- CMWU has local convergence guarantees in the zero-sum case and performs competitively compared to existing methods.
  - *new proof technique based on eigenvalue perturbation*

Learning Last-iterate convergence in Games

- Can we learn the update rule coefficients h and ϵ?
- Recent research shows that the update rule needs to depend on the *nature* of the game (Cheung & Piliouras, 2020).
  - *we need to encode/featurize the game*
- We allow different coefficients for the 2 players
  - 4 coefficients in total

Learning Last-iterate convergence in Games

- RL problem over trajectory τ, minimize the distance to Nash δ:

$$\min_{\pi} \mathbb{E}_{\pi} \sum_{t=1}^{\tau} \frac{\delta(x^t, y^t)}{\delta(x^0, y^0)}$$

- MDP state: game signature + trajectory information of both players: distances to Nash, gradient, Hessian, payoffs.
- Policy actions. $a_t = (h_1, h_2, h_1\epsilon_1, h_2\epsilon_2)$.
- Train on mixtures of 3 components

Game decomposition

- How to featurize/encode a 2P game?
- Use linear operators ρ that satisfy $\rho^2 = \rho$, a.k.a. projectors.
- The whole space is the direct sum of the Kernel and Range of any projector:

$$\text{2P Games} = Ker(\rho) \oplus Range(\rho)$$

- For example, ρ_Z computes the zero-sum component of any 2P game, and $\text{Id} - \rho_Z$ its cooperative component

$$\boldsymbol{\rho_Z} : (f_1, f_2) \to \frac{1}{2}(f_1 - f_2, f_2 - f_1)$$

Game decomposition

- Given n commutative projectors we have:

$$\text{2P Games} = \bigoplus_{\mathcal{C}_i \in \{K_{\rho_i}, R_{\rho_i}\}} \bigcap_{i=1}^{n} \mathcal{C}_i \qquad (2^n \text{ components})$$

- This view generalizes cyclic/transitive games in (Balduzzi, 2019) from the zero-sum symmetric case to any 2P game, and unifies it with trivial games of (Cheung & Piliouras, 2020).
- Application to the case $n = 3$ gives **8 components** associated to zero-sum/cooperative, symmetric/antisymmetric, transitive/cyclic games.

Learning to converge to Nash Equilibria

The RL policy is able to exploit the game signature across a wide range of game types.

Game Type	Avg. Distance to Nash
ZT	6
ZCy	4
CT	4
CCy	4
Avg. Mixtures of 1	5
ZT + ZCy	**7**
ZT + CT	3
ZT + CCy	6
ZCy + CT	15
ZCy + CCy	10
CT + CCy	5
Avg. Mixtures of 2	**8**
ZT + ZCy + CT	12
ZT + ZCy + CCy	19
ZT + CT + CCy	5
ZCy + CT + CCy	20
ZT + ZCy + CT + CCy	15
Avg. Mixtures of 3-4	14
Avg. k-Mixtures	**9**

Learnt coefficients - Pure components

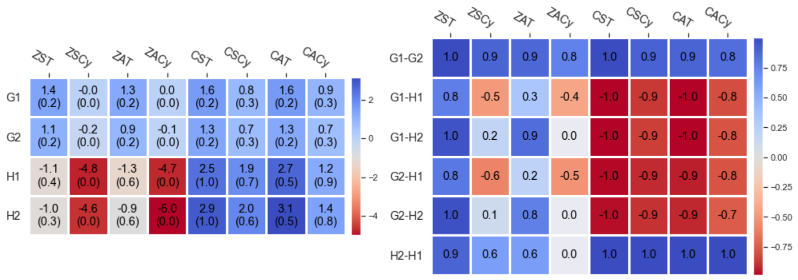

Figure 1: Learnt coefficients of gradient $G_k = h_k$ and Hessian $H_k = -h_k \epsilon_k$ of players k across 8 pure game components. Mean value **(left)**. Pairwise correlations **(right)**.

- Gradient stronger in transitive games, Hessian stronger in cyclic games.
- mirror behavior between zero-sum and cooperative games.

Coefficient trajectory Shape - Pure Game components

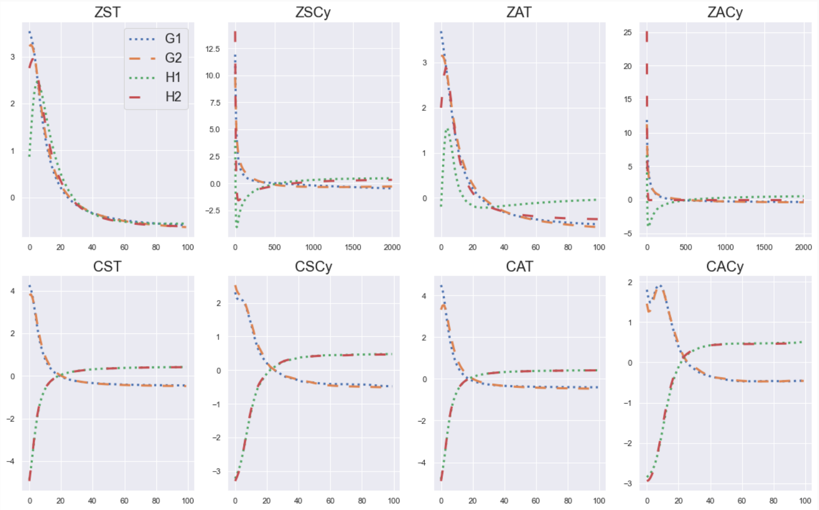

Figure 2: Average standardized per-episode-trajectory ("shape") of coefficients of gradient $G_k = h_k$ and Hessian $H_k = -h_k\epsilon_k$ of players $k = 1, 2$ across game types, as a function of time t.

SHAP analysis

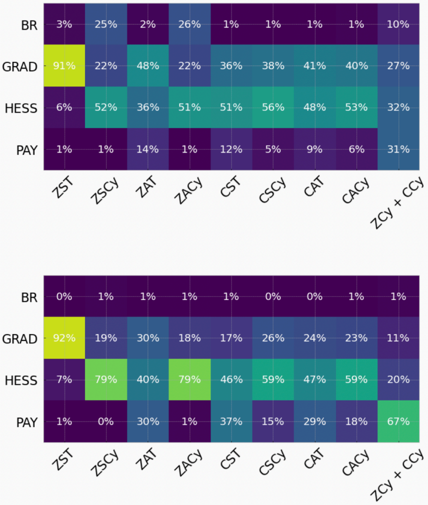

Figure 3: SHAP importance with respect to **(top)** gradient learning rates $G_k = h_k$ and **(bottom)** Hessian learning rates $H_k = -h_k\epsilon_k$ for players k for the four input groups and eight unique game types.