Aligning Individual and Collective Objectives in Multi-Agent Cooperation

Yang Li[1], Wenhao Zhang[2], Jianhong Wang[3,1], Shao Zhang[2], Yali Du[4], Ying Wen[2], Wei Pan[1]

1: The University of Manchester; 2: Shanghai Jiao Tong University; 3. University of Bristol; 4: King's College London

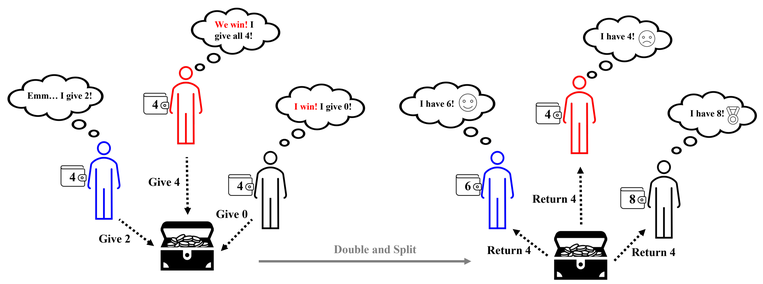
TL;DR

We propose the Altruistic Gradient Adjustment (AgA) for mixed-motive games—a method that aligns individual and collective objectives through gradient adjustments, supported by theoretical proofs and empirical validation.

Background & Overview

A **mixed-motive game** is a game where players have both cooperative and competitive motivations, and their incentives are partly aligned and partly opposed.

Here is *an example of mixed-motive game*, each player has 4 dollars and need to contribute any dollars to the public pool. The public pool will double the money in pool and split them among players.



In the example, the altruistic player ends up the poorest; the selfish player becomes the wealthiest; the collective outcome is suboptimal.

In this paper, we aim to align individual and collective objectives to foster cooperation in mixed-motive settings.

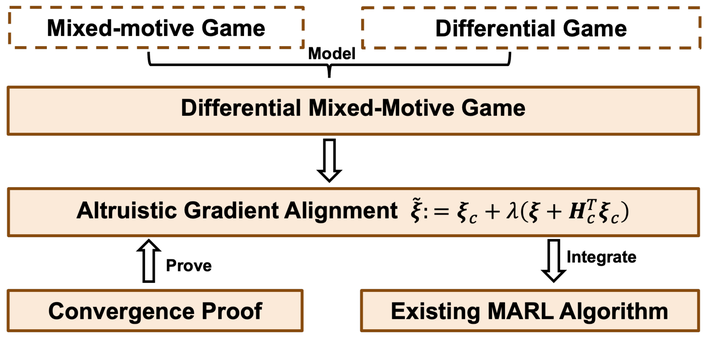


Figure 1. Overview of our proposed altruistic gradient alignment method.

Differential Mixed-motive Game

Differential Mixed-motive Game (DMG) is defined as a tuple $\{\mathcal{N}, \boldsymbol{w}, \boldsymbol{\ell}\}$, where

- $\mathcal{N} = \{1, \ldots, n\}$ is the set of players.
- The parameter set $\boldsymbol{w} = [\boldsymbol{w}_i]^n \in \mathbb{R}^d$ is defined, each with $\boldsymbol{w}_i \in \mathbb{R}^{d_i}$ and $d = \sum_{i=1}^n d_i$.
- $\boldsymbol{\ell} = \{\ell_i : \mathbb{R}^d \to \mathbb{R}\}_{i=1}^n$ represents the corresponding losses. These losses are assumed to be at least twice differential.

Here, Each player $i \in \mathcal{N}$ is equipped with a policy, parameterized by $\boldsymbol{w}_i$, aiming to minimize its loss ℓ_i.

Differentiable losses exhibit the ***mixed motivation property***: minimization of individual losses can result in a conflict between individuals or between individual and collective objectives (e.g., maximizing individual stats and winning the game are often conflict in basketball matches).

Altruistic Gradient Adjustment (AgA)

Building on differential mixed-motive game and gradient-adjustment optimization, we propose the Altruistic Gradient Adjustment (AgA) algorithm to align individual and collective objectives from a gradient-based perspective.

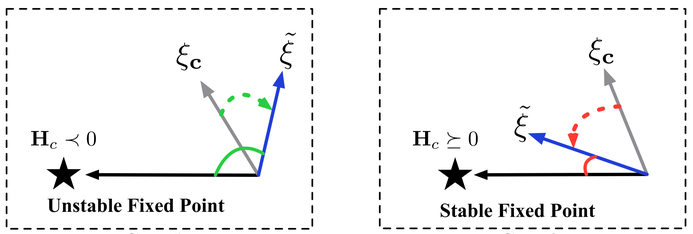
Definition (Altruistic Gradient Adjustment.) *Altruistic gradient adjustment (AgA) extends the gradient term in the learning dynamic as*

$$\tilde{\boldsymbol{\xi}} := \boldsymbol{\xi}_c + \lambda \boldsymbol{\xi}_{adj} = \boldsymbol{\xi}_c + \lambda \left(\boldsymbol{\xi} + \boldsymbol{H}_c^T \boldsymbol{\xi}_c \right), \tag{1}$$

where $\lambda \in \mathbb{R}$ is alignment parameter, $\boldsymbol{\xi}$ is the gradient of the losses with respect to the parameters of the respective players, $\lambda \boldsymbol{\xi}_{adj}$ is called adjustment term. In $\boldsymbol{\xi}_{adj}$, $\boldsymbol{\xi}_c$ and $\boldsymbol{H}_c$ is the gradient vector and Hessian matrix of the game about collective loss.

In our AgA method, we only need to compute the Hessian-vector product $\boldsymbol{H}_c^T \xi_c$, instead of the full Hessian matrix $\boldsymbol{H}_c^T$, reducing the time complexity to $\boldsymbol{H}_c^T \xi_c$ is $\mathcal{O}(n)$ for n weights.

We theoretically show that selecting the appropriate sign of λ ensures that: (1) AgA pushes the gradient away from unstable points, and (2) AgA pulls the gradient toward stable fixed points via an adjustment term.



Corollary. *In the neighborhood of fixed points of the collective objective, AgA will **pull** the gradient **toward** stable fixed points, which means $\theta(\tilde{\boldsymbol{\xi}}, \nabla \mathcal{H}_c) \leq \theta(\boldsymbol{\xi}_c, \nabla \mathcal{H}_c)$, and **push away** from unstable ones, indicated by $\theta(\tilde{\boldsymbol{\xi}}, \nabla \mathcal{H}_c) \geq \theta(\boldsymbol{\xi}_c, \nabla \mathcal{H}_c)$, if λ satisfies $\lambda \cdot \langle \boldsymbol{\xi}_c, \nabla \mathcal{H}_c \rangle (\langle \boldsymbol{\xi}, \nabla \mathcal{H}_c \rangle + \|\nabla \mathcal{H}_c\|^2) \geq 0$.*

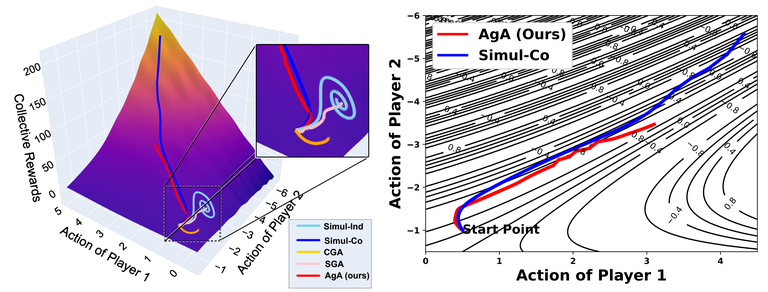
Case Study: A Toy Experiment

We conducted a series of toy experiments in a two-player differentiable mixed-motive game. The game is defined as follows.

Example. *Consider a two-player DMG with $\ell_1(a_1, a_2) = -sin(a_1 a_2 + a_2^2)$ and $\ell_2(a_1, a_2) = -[cos(1 + a_1 - (1 + a_2)^2) + a_1 a_2^2]$, where a_i represents the action of the player i $(i = 1, 2)$, and $a_i \in \mathbb{R}$. The rewards for the two players are the negation of their respective losses.*

AgA Successfully Aligning Individual and Collective Objectives

Fig. 2a shows the trajectories over the collective reward landscape, with deeper orange indicating higher rewards. Only Simul-Co and AgA move toward the social optimum. However, Simul-Co ignores Player 1's interests, focusing on its own reward peaks and valleys.

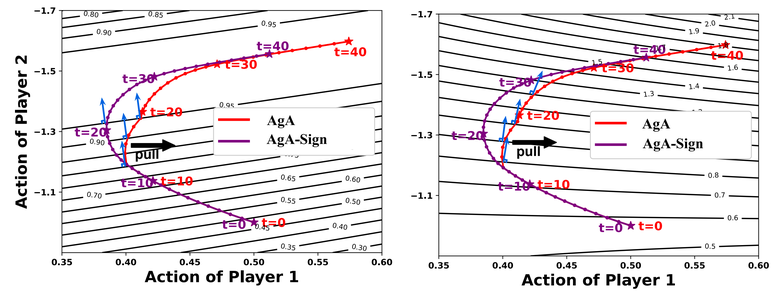


(a) Collective Reward Landscape (b) Reward Contour of Player 1

Figure 2. Trajectories of optimization in a two-player DMG.

Alignment Effectiveness of Corollary 4.3

The comparison between AgA (shown in red) and AgA without sign alignment (AgA-Sign, in purple) trajectories spans 40 steps, marked at every tenth step. Starting from the 14th step, sign alignment pulls the gradient toward the steepest direction, resulting in AgA reducing the number of steps by approximately 15% compared to AgA-Sign by the end of the trajectory.



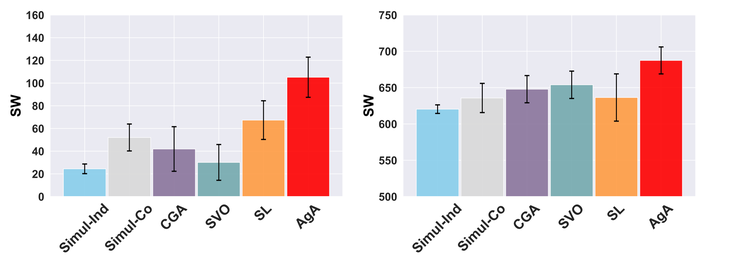
Main Experiment Results

Two-player public good games

A two-player public goods matrix game: Players $1, 2$ contribute amounts a_i from a budget $[0, b]$, and their payoffs are $p_i = b - a_i + \frac{c}{2}(a_1 + a_2)$. Social welfare is $SW = p_1 + p_2$. In our experiments, we set $b = 1$ and $c = 1.5$. We show the mean of value and 95% confidence interval utilizing 50 random runs.

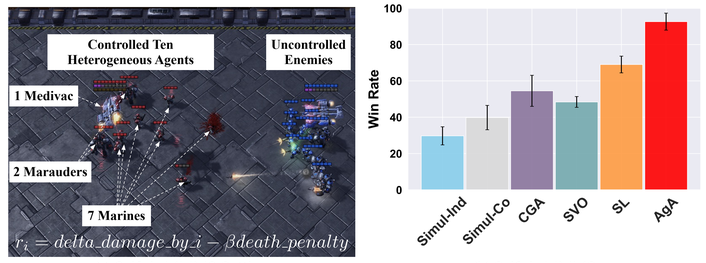
Metrics	Simul-Ind	CGA	SGA	SVO	Simul-Co	SL	AgA
$\mathbf{r_1}$	1.133 ± 0.063	1.156 ± 0.060	1.175 ± 0.062	1.104 ± 0.054	1.433 ± 0.056	1.314 ± 0.062	1.443± 0.042
$\mathbf{r_2}$	1.184 ± 0.065	1.150 ± 0.057	1.137 ± 0.063	1.060 ± 0.051	1.381 ± 0.065	1.371 ± 0.057	1.459 ± 0.041
SW	2.316± 0.039	2.306 ± 0.039	2.312± 0.044	2.164 ± 0.026	2.814 ± 0.033	2.684 ± 0.049	2.903± 0.023
Equality	0.923± 0.014	0.929± 0.012	0.924± 0.013	0.930 ± 0.011	0.941 ± 0.014	0.940 ± 0.011	0.960 ± 0.008

Cleanup and Harvest: common-used mixed-motive testbeds



(a) Cleanup (b) Harvest

Selfish-MMM2: a large-scale mixed-motive testbed



(a) Selfish-MMM2 Setting (b) Selfish-MMM2