Anticipating Opponent Responses via Conjectural Variations

Benjamin Chasnov, Tanner Fiez, Lillian Ratliff. University of Washington, Seattle WA.
{bchasnov,fiezt,ratliffl}@uw.edu
Electrical and Computer Engineering

NEURAL INFORMATION PROCESSING SYSTEMS
Smooth Games Optimization and Machine Learning Workshop
December 2019

Analysis of Gradient-based Learning in Continuous Games

Consider an $n-$player continuous game $\mathcal{G} = (f_1, \dots, f_n)$ with strategy space $X = X_1 \times \dots \times X_n$ and costs $f_i : X \to \mathbb{R}$. Players minimize costs using gradient-based updates with step sizes $\Gamma = \mathrm{diag}(\gamma_i)$,

$$x_{k+1} = x_k - \Gamma g_{(\cdot)}(x_k),$$

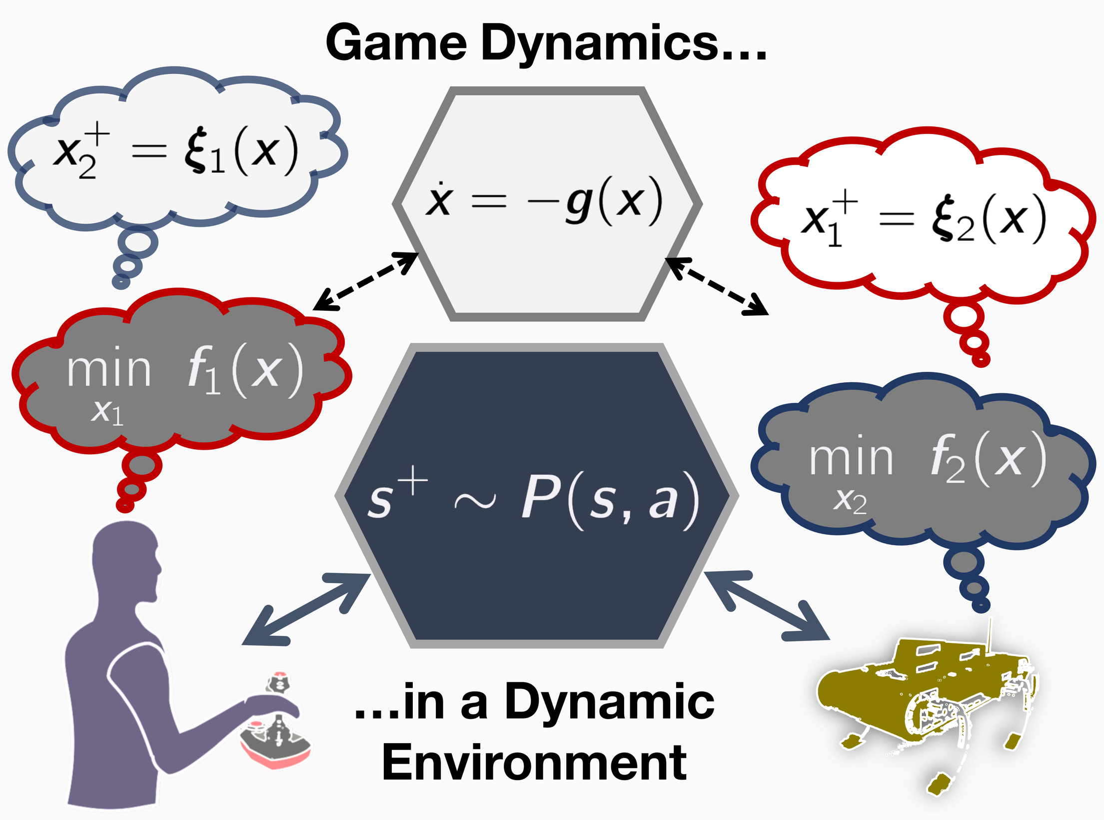
where $g_{(\cdot)}$ is the learning rule. For small step sizes, the learning rules are approximated by differential equation

$$\dot{x} = -\Gamma g_{(\cdot)}(x).$$

Convergence to critical point x^* is guaranteed when spectral radius $\rho(I - \Gamma D g_{(\cdot)}(x^*)) < 1$.

Introduction

Algorithms interact with each other and with humans. Understanding the fundamental characteristics of selfish interactions will enable robust, interpretable, and predictable outcomes.



Our proposed framework models agents that anticipate each other.

- Decentralized independent learning;
- *meaningful* equilibria;
- predictions of collective behaviors (qualitative and quantitative).

Definition: Differential GCVE

A *differential general conjectural variations equilibrium* (GCVE) $x^* = (x_1^*, \dots, x_n^*)$ for conjectures $(\xi_1, \dots, \xi_n)$ satisfies

$$\text{First order: } (D_i f_i + D_i \xi_i^\top \circ D_j f_i)(x^*) = 0,$$

$$\text{Second order: } D_i(D_i f_i + D_i \xi_i^\top \circ D_j f_i)(x^*) > 0, \forall i.$$

Special case of $D_i \xi_i \equiv 0$, $\forall i$ recovers the optimality conditions for Nash. Conjecture reaction functions and conjectural variations can be defined explicitly or implicitly.

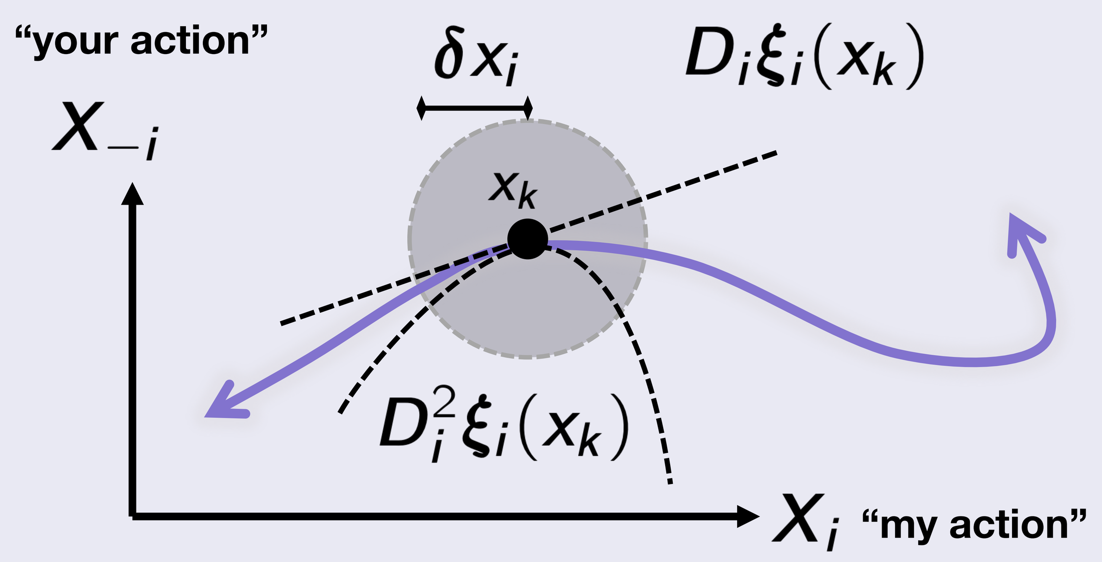
Conjectural Variations

Agents formulate conjecture reaction functions about others $\xi_i : X \to X_{-i}$ and optimize:

$$\min_{x_i} f_i(x_i, \xi_i(x_i, x_{-i})).$$

Of importance is conjectural variation $D_i \xi_i : X_i \to X_{-i}$.

"What I think you will do in reaction to what I do."



Player i conjectures the infinitesimal variation of its strategy δx_i by an infinitesimal variation of its opponent's strategy δx_j:

$$D_i \xi(x_k) \delta x_i = \delta x_j$$

where $D_i \xi(x_k)$ is the conjectural variation for a given benchmark strategy profile x_k at iteration k. "Optimal" critical points are GCVE.

Future Work

Within the conjectures framework, future work includes

- **analysis** of equilibrium concepts,
  "why should agents adopt conjectures?"
- **synthesis** of new algorithms,
  "how to compute equilibria using limited information?"
- **demonstration** of novel applications,
  "what scenarios benefit from game-theoretic outcomes?"

References and Related Work

Non-cooperative games and conjectural variations:
1 Basar and Olsder, Dynamic noncooperative game theory, 1999.
2 Figuieres et al. Theory of Conjectural Variations, 2004.
Related machine learning updates:
3 Balduzzi et al. The mechanics of n-player differentiable games. ICML 2018.
4 Zhang and Lesser. Multi-agent learning with policy prediction. AAAI 2010.
5 Foerster et al. Learning with opponent-learning awareness. AAMAS 2018.
6 Mescheder et al. On the convergence properties of gan training. NIPS 2017.
Our work
7 Fiez, Chasnov, and Ratliff. Convergence of Learning Dynamics in Stackelberg Games. arXiv 2019.
8 Chasnov et al. Convergence Analysis of Gradient-Based Learning in Continuous Games. UAI 2019.
9 Chasnov et al. Gradient Conjectures for Strategic Multi-Agent Learning. Pre-print 2019.

Synthesis of Algorithms for Agents with Bounded Rationality

The benchmark strategy x_k under perturbation of player i's action δx_i reveals a perturbed conjecture reaction function $\xi_i(x_k + [\delta x_i \; 0])$ given by

$$\xi_i(x_k) + \delta x_i^\top D_i \xi_i(x_k) + \delta x_i^\top D_i^2 \xi(x_k) \delta x_i + \mathcal{O}(\delta x_i^3).$$

The order of conjectural variations corresponds to layers of rationality [1],

"...what I think you think I think you think..."

Perfect knowledge of opponents' best-response is modeled using the *implicit function theorem*: there exists conjectural variation $D_i \xi_i(x)$ defined implicitly in a neighbourhood by $D_j f_j(x) = 0$, i.e.

$$D_i \xi_i \equiv D_j f_j^{-2} \circ D_{ji} f_j.$$

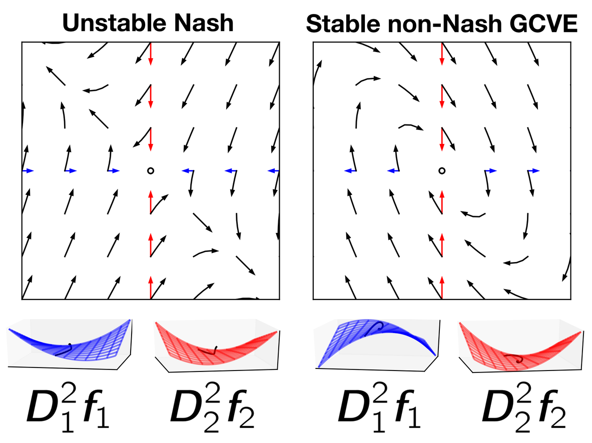
We form optimization problems that extremize an individual objective given (approximate) conjectures.

For general-sum game $\mathcal{G} = (f_1, f_2)$ and conjectural variations $\xi_i \equiv 0$, the fixed point x^* of its learning rule, simultaneous gradient descent $g_{\mathtt{simgrad}} \equiv (D_1 f_1, D_2 f_2)$, is stable iff

$$D g_{\mathtt{simgrad}}(x^*) = \begin{bmatrix} D_1^2 f_1 & D_{12} f_1 \\ D_{21} f_2 & D_2^2 f_2 \end{bmatrix}(x^*) > 0.$$

A non-symmetric game Jacobian gives rise to rich behaviors:

- stable non-Nash equilibria,
- unstable Nash equilibria,
- periodic orbits...



Algorithm: Implicit Conjectures

Players assume opponents play best-response,

$$\min_{x_i} f_i(x_i, \xi_i(x_i))$$
$$\text{s.t. } \xi_i(x_i) = \arg\min_\theta f_j(x_i, \theta),$$

giving rise to learning rule

$$g_{\mathtt{impconj},i} \equiv D_i f_i - D_{ij} f_j \circ D_j^{-2} f_j \circ D_j f_i.$$

The Jacobian at a fixed point $x^* \in X$ of a zero-sum game $\mathcal{G} = (f, -f)$ is

$$D g_{\mathtt{impconj}}(x^*) = \begin{bmatrix} S_1(x^*) & 0 \\ 0 & -S_2(x^*) \end{bmatrix}$$

where $S_i \equiv D_i^2 f - D_{ij} f \circ D_j^{-2} f \circ D_{ji} f$ are Schur complements of $D g_{\mathtt{simgrad}}$.

Related Equilibrium Concepts

$$\boxed{\begin{aligned} x_1^* &\in \arg\min f_1(x) \\ x_2^* &\in \arg\min f_2(x) \end{aligned}} \quad \textbf{Nash}$$

$$\boxed{\begin{aligned} x_1^* &\in \arg\min_{\xi_1(x) \in \arg\min f_2(x)} f_1(x_1, \xi_1(x)) \\ x_2^* &\in \arg\min f_2(x) \end{aligned}} \quad \textbf{Stackelberg}$$

$$\boxed{\begin{aligned} x_1^* &\in \arg\min_{\xi_1(x) \in \arg\min f_2(x)} f_1(x_1, \xi_1(x)) \\ x_2^* &\in \arg\min_{\xi_2(x) \in \arg\min f_1(x)} f_2(\xi_2(x), x_2) \end{aligned}} \quad \textbf{Conjectures}$$

See [7,8,9]. We approximate the best-response conjectures with gradient steps:

Algorithm: Gradient Conjectures

Players assume opponents take a gradient step,

$$\min_{x_i} f_i(x_i, \xi_i(x_i, x_{-i}))$$
$$\text{s.t. } \xi_i(x) = x_j - \eta_i D_j f_j(x),$$

giving rise to learning rule

$$g_{\mathtt{gradconj},i} \equiv D_i f_i - \eta_i D_{ij} f_j \circ D_j f_i.$$

Learning with Opponent-Learning Awareness [5] employs a similar update.

Algorithm: Fast Conjectures

Players approximate objective by Taylor expansion around benchmark x_k,

$$\min_{x_i} f_i(x_i, x_j) + \delta x_j^\top D_j f_i(x_i, x_{k,j})$$
$$\text{s.t. } \delta x_j = -\eta_i D_j f_j(x_k),$$

giving rise to learning rule

$$g_{\mathtt{fastconj},i} \equiv D_i f_i - \eta_i D_{ij} f_i \circ D_j f_j.$$

Lookahead [4] employs a similar update. For $\mathcal{G} = (f, -f)$, the game Jacobians $D g_{\mathtt{gradconj}}(x^*) = D g_{\mathtt{fastconj}}(x^*)$ are both

$$\begin{bmatrix} D_1^2 f + \eta_1 P & D_{12} f \circ (I - \eta_1 D_2^2 f) \\ -D_{21} f \circ (I - \eta_2 D_1^2 f) & -D_2^2 f + \eta_2 Q \end{bmatrix}(x^*)$$

where $P \equiv D_{12} f \circ D_{21} f \geq 0$ and $Q \equiv D_{21} f \circ D_{12} f \geq 0$

Deriving other Multi-agent Learning Updates

We consider regularized conjectures and derive various other updates:

- Symplectic Gradient Adjustment [3]

$$\min_{x_i} f_i(x) + \eta \delta(x_k)^\top D_i f_i(x) + \eta \|\delta(x)\|_2^2$$
$$\text{s.t. } \delta(x) = -D_j f_j(x), \; \eta = \mathrm{sign}(\mathrm{align}(g, g_{\mathrm{sga}}))$$

- Consensus Optimization [6]

$$\min_{x_i} f_i(x) + \eta \|g(x)\|_2^2 \quad \text{s.t. } g \equiv g_{\mathtt{simgrad}}$$

Applications of Decentralized Game Dynamics

When agents have limited communication or lack trust of each other, they are boundedly rational.

The conjecture framework can be used to predict selfish or cooperative behaviors in a population of economic agents.

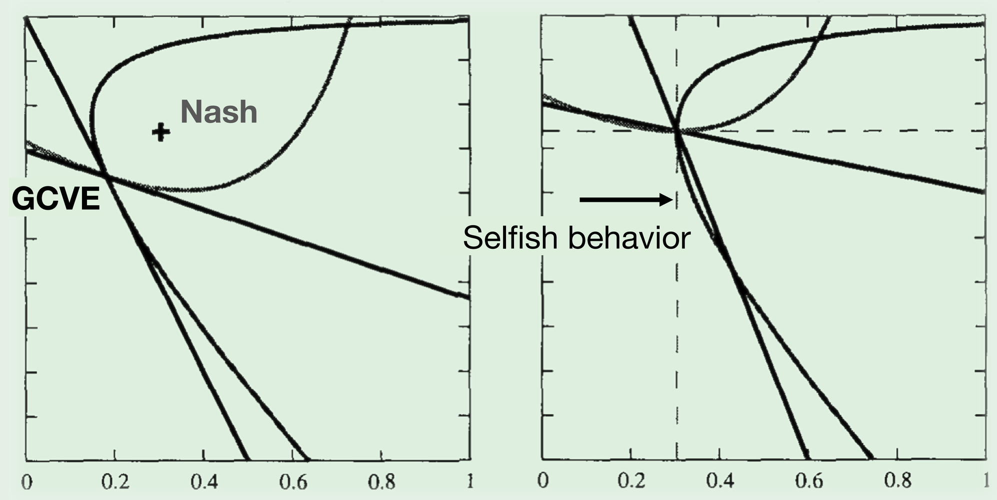
Example: Voluntary Contribution to a Public Good [2]

Consider $n-$player game $\mathcal{G} = (V_i) \; i \in [1, n]$ where player utility is modeled by

$$V_i(x_i, x_{-i}) = (I_i - x_i)_i^\alpha (x_i + X_{-i})^{1-\alpha_i}$$

where player i has

- scalar action $x_i \in \mathbb{R}$,
- utility $V_i : \mathbb{R}^n \to \mathbb{R}$,
- model parameters (I_i, α_i),
- and observes opponent behaviors $X_{-i} = \sum_{j \neq i} x_j$.



Conjectural best-response and iso-payoff curves

Nash best-response and iso-payoff curves

Solving for fixed points of $g_{\mathtt{simgrad}}$, $g_{\mathtt{impconj}}$, and

$$g_{\mathtt{stack}} \equiv \begin{bmatrix} g_{\mathtt{impconj},i} \\ g_{\mathtt{simgrad},-i} \end{bmatrix},$$

we obtain (stable) fixed points corresponding to

- Nash with highest cost;
- Stackelberg leader with a lower cost;
- conjectural equilibrium that is Pareto-efficient.

Agents forming different conjectures about each other can reveal emergent behaviors in dynamic and stochastic environments.

Example: Multi-agent Reinforcement Learning with Conjectures

Consider $n-$player game $\mathcal{G} = (R_i) \; i \in [1, n]$ where $R_i : \Theta_1 \times \dots \times \Theta_n \to \mathbb{R}$ and $R_i(\theta_i, \theta_{-i})$ is

$$\mathbb{E}_{(s_t, a_t) \sim \mathrm{traj}} \sum_{t=0}^{T} \widehat{R}_{t,i} \pi_{\theta_i}(s_t, a_{t,i}) \Pi_{-i}(\theta_{-i}, a_{t,-i}).$$

Rewards are sampled from rollouts, represented by

$$\widehat{R}_{t,i} = P(s_t, a_{t,i}, a_{t,-i}) R_i(s_t, a_{t,i}, a_{t,-i}).$$

where player i has at time t

- action $a_{t,i} \sim \pi_{\theta_i}(s_t)$ with stochastic policy...
- ...parameterized by θ_i.
- sampled rewards $\widehat{R}_{t,i}$,
- stochastic model dynamics $s_{t+1} \sim P(s_t, a_t)$

We obtain the stochastic policy gradient $\nabla_{\theta_i} \log \pi_{\theta_i}(s, a_i) \in \Theta_i$, its conjectural variations

$$(\nabla_{\theta_j} \log \pi_{\theta_i} \circ \nabla_{\theta_i} \log \pi_{\theta_j}^\top)(s, a_i, a_j) : \Theta_i \to \Theta_j$$

and individual Hessians

$$(\nabla_{\theta_i}^2 \log \pi_{\theta_i} + \nabla_{\theta_i} \log \pi_{\theta_i} \circ \nabla_{\theta_i} \log \pi_{\theta_i}^\top)(s, a_i) : \Theta_i \to \Theta_i.$$

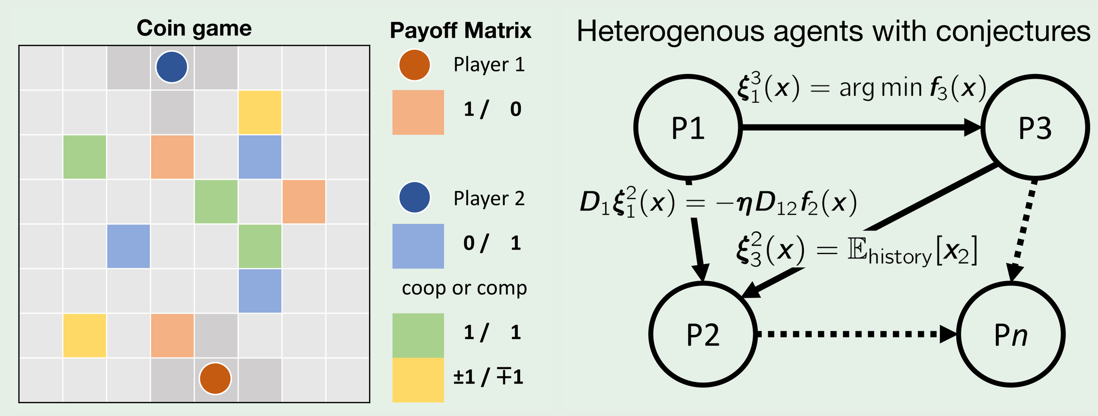


Figure: Multi-player coin game with general-sum rewards. Agents rewards are coupled through stochastic state dynamics.

The algorithms are implemented efficiently on large-scale ML models using modern auto-differentiation frameworks.

Example: Training Machine Learning Models

Consider two player adversarial game with $\mathcal{G} = (\mathcal{L}, -\mathcal{L})$ on (G, D) using the vanilla GAN objective. We demonstrate implementations of

- **Jacobian-vector products** using forward- and reverse- mode;
- **Inverse-Jacobian-vector products** with regularization using fixed iteration linear solvers (e.g. conjugate gradient methods).

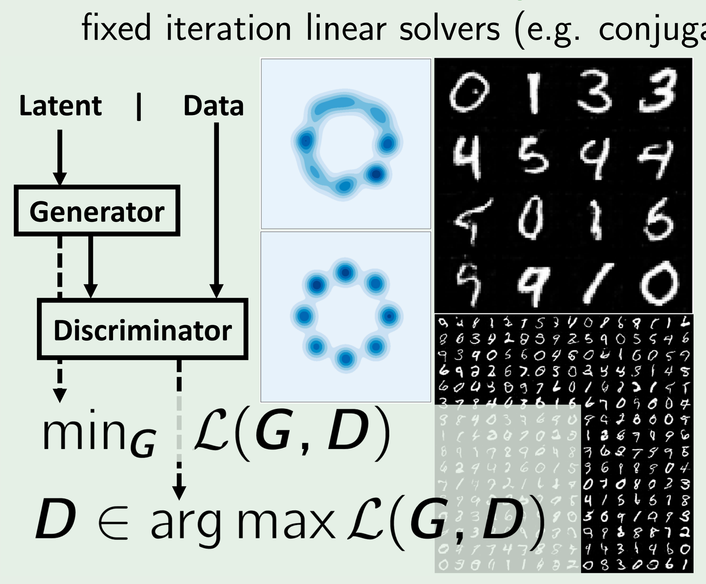


Figure: Generative adversarial network converges to non-Nash Stackelberg equilibrium: generator with implicit conjecture, follower with Nash conjecture.