MITSUBISHI ELECTRIC
Changes for the Better

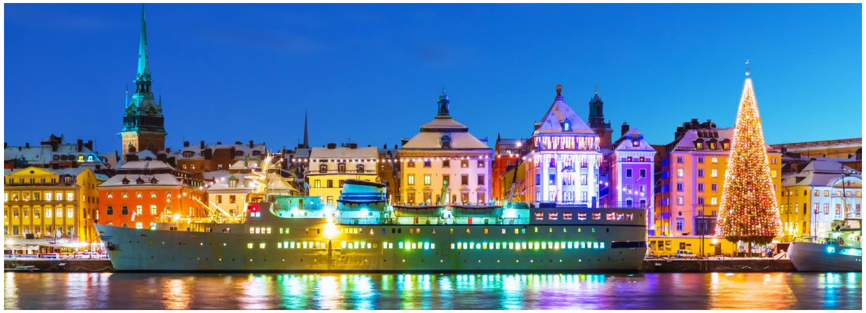
for a greener tomorrow

Game Theoretic Optimization via Gradient-based Nikaido-Isoda Function

Arvind U Raghunathan, Anoop Cherian, Devesh Jha

Mitsubishi Electric Research Labs (MERL), Cambridge, MA

ICML



Problem Statement

2-player Game*

Find $\boldsymbol{x}^* = (x_1^*, x_2^*)$ such that

$$x_1^* = \arg \min_{x_1 \in R^{n_1}} f_1(x_1, x_2^*)$$
$$x_2^* = \arg \min_{x_2 \in R^{n_2}} f_2(x_1^*, x_2)$$

Where do they arise?

- Economics , Mechanism Design
- Generative Adversarial Networks (GANs)
  - Min-Max optimization can be cast as 2-player game
  - Adversarial/robust problem formulations

Our Goal: Algorithms for the solution of N-player Games

* - Can be generalized to the N-player setting

Characterization of Solutions

2-player Game*

Find $\boldsymbol{x}^* = (x_1^*, x_2^*)$ such that

$$x_1^* = \arg\min_{x_1 \in R^{n_1}} f_1(x_1, x_2^*)$$
$$x_2^* = \arg\min_{x_2 \in R^{n_2}} f_2(x_1^*, x_2)$$

- Nash Equilibrium: $\mathcal{S}^{NE} = \{\boldsymbol{x}^* \mid above\ holds\}$
  - x_i^* solves player-i 's problem

- Stationary Nash Points: $\mathcal{S}^{SNP} = \{\boldsymbol{x}^* \mid \nabla_i f_i(\boldsymbol{x}^*) = 0\}$
  - x_i^* is the first-order stationary point for player-i 's problem

f_i **are nonconvex** $\rightarrow \mathcal{S}^{SNP}$ **are likely to be limit points of algorithms**

* - Can be generalized to the N-player setting

Our Key Contributions

1. We propose re-formulations of the game objectives using merit functions, namely the Nikaido-Isoda (NI) function
2. As optimization of NI functions is difficult, we introduce *Gradient-based NI functions (GNI)*, which is cheap and converges to Stationary Nash Points of the games.
3. We explore theoretical properties of GNI, providing error bounds and convergence results under various game settings.
   - Specifically, under certain conditions, we show that the solutions converge linearly.
4. We empirically demonstrate the usefulness of our formulations on synthetic datasets.

Our Gradient-based Nikaido-Isoda (GNI) Function

$$V(\boldsymbol{x}) = \sum_{i=1}^{2} V_i(\boldsymbol{x})$$

$$V_1(\boldsymbol{x}) = f_1(x_1, x_2) - f_1(y_1(\boldsymbol{x}), x_2)$$
$$V_2(\boldsymbol{x}) = f_2(x_1, x_2) - f_2(x_1, y_2(\boldsymbol{x}))$$

Nikaido-Isoda Function

$$y_1(\boldsymbol{x}) = \inf_{y_1} f_1(y_1, x_2)$$
$$y_2(\boldsymbol{x}) = \inf_{y_2} f_2(x_1, y_2)$$

GNI Function

$$y_1(\boldsymbol{x}) = x_1 - \nabla_1 f_1(x_1, x_2)$$
$$y_2(\boldsymbol{x}) = x_2 - \nabla_2 f_2(x_1, x_2)$$

- $V_i(\boldsymbol{x}) \geq 0$ *for all* $\boldsymbol{x}$, $\{\boldsymbol{x} \mid V(\boldsymbol{x}) = 0\} = \mathcal{S}^{SNP}$
- Minimize $V(\boldsymbol{x})$ instead of NI function
- Evaluation of $V_i(\boldsymbol{x})$ requires only a gradient of f_i w.r.t x_i

Properties of GNI Function

$$V(\boldsymbol{x}) = \sum_{i=1}^{2} V_i(\boldsymbol{x})$$

$$V_1(\boldsymbol{x}) = f_1(x_1, x_2) - f_1(y_1(\boldsymbol{x}), x_2)$$
$$V_2(\boldsymbol{x}) = f_2(x_1, x_2) - f_2(x_1, y_2(\boldsymbol{x}))$$

$$y_1(\boldsymbol{x}) = x_1 - \eta \nabla_1 f_1(x_1, x_2)$$
$$y_2(\boldsymbol{x}) = x_2 - \eta \nabla_2 f_2(x_1, x_2)$$

f_i has L_f–Lipschitz gradient. If $\eta \leq \frac{1}{L_f}$ then

$$\frac{\eta}{2} \|\nabla_i f_i(\boldsymbol{x})\|^2 \leq V_i(\boldsymbol{x}) \leq \frac{3\eta}{2} \|\nabla_i f_i(\boldsymbol{x})\|^2$$

$V(\boldsymbol{x}^*) = 0$ if and only if $\boldsymbol{x}^* \in \mathcal{S}^{SNP}$

Gradient Descent on GNI Function

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^{k} - \rho \nabla V(\boldsymbol{x})$$

$$\nabla V_1(\boldsymbol{x}) = \nabla f_1(\boldsymbol{x}) - (I - \eta E_1 \nabla^2 f_1(\boldsymbol{x})) \nabla f_1(y_1(\boldsymbol{x}), x_2)$$
$$\nabla V_2(\boldsymbol{x}) = \nabla f_2(\boldsymbol{x}) - (I - \eta E_2 \nabla^2 f_2(\boldsymbol{x})) \nabla f_2\big(x_1, y_2(\boldsymbol{x})\big)$$

V has L_V–Lipschitz gradient. If $\rho \leq \frac{1}{L_V}$ then $\{\boldsymbol{x}^k\}$ converges sublinearly to $\boldsymbol{x}^*: \nabla V(\boldsymbol{x}^*) = 0$

If in addition, $V(\boldsymbol{x}^*) = 0$ then $\boldsymbol{x}^* \in \mathcal{S}^{SNP}$

Further, if V satisfies Polyak-Lojasiewicz inequality, then $\{\boldsymbol{x}^k\}$ converges linearly to $\boldsymbol{x}^*: V(\boldsymbol{x}^*) = 0$.
e.g., Satisfied for Quadratic Games

MITSUBISHI ELECTRIC
Changes for the Better

for a greener tomorrow

Modified Gradient Descent on GNI Function

$$x^{k+1} = x^k - \rho \nabla \hat{V}(x)$$

$$\nabla \hat{V}_1(x) = \nabla f_1(x) - (I - \eta E_1 \nabla^2 f_1(x)) \nabla f_1(y_1(x), x_2)$$

$$\nabla \hat{V}_2(x) = \nabla f_2(x) - (I - \eta E_2 \nabla^2 f_2(x)) \nabla f_2\big(x_1, y_2(x)\big)$$

Replace with
Secant Approximation

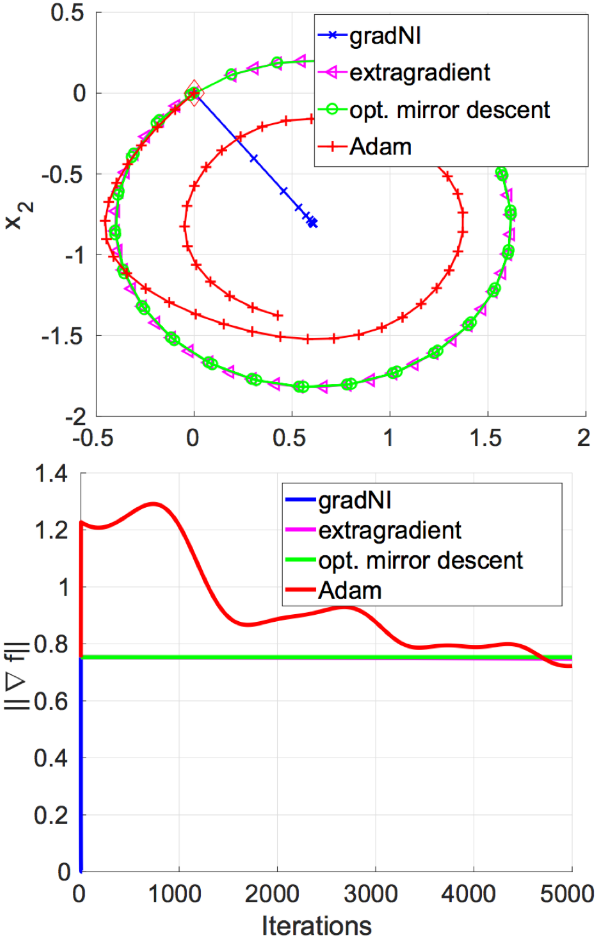
Under additional assumptions on the approximation,
we recover previous results

Experiments on Two-Player Games

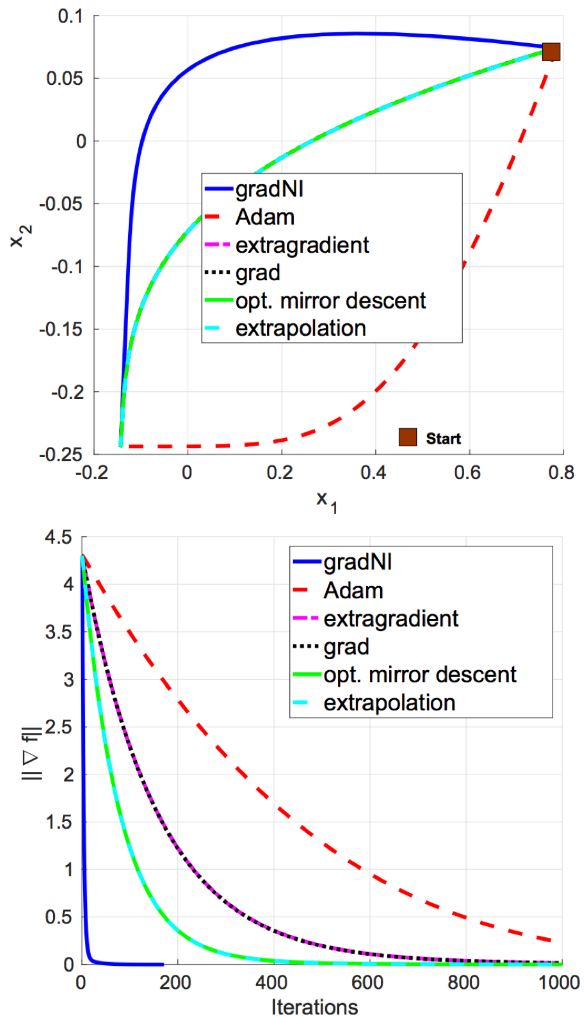
1. Bilinear: $f_1(x) = x_1^T Q x_2 + q_1^T x_1 + q_2^T x_2 = -f_2(x)$
2. Quadratic: $f_i(x) = \frac{1}{2} x^T Q_i x + r_i^T x, \text{ for } i = 1, 2$
3. Delta GAN: $$f_1 = \log(1 + \exp(\theta x_1)) + \log(1 + \exp(x_1 x_2))$$ $$f_2 = -\log(1 + \exp(x_1 x_2)),$$
4. Linear GAN:

$$f_1 = -\mathbb{E}_{\theta \sim P_r} \log\left(x_1^T \theta\right) - \mathbb{E}_{z \sim P_z} \log\left(1 - x_1^T \operatorname{diag}\left(x_2\right) z\right)$$
$$f_2 = -\mathbb{E}_{z \sim P_z} \log\left(x_2^T \operatorname{diag}\left(x_1\right) z\right)$$

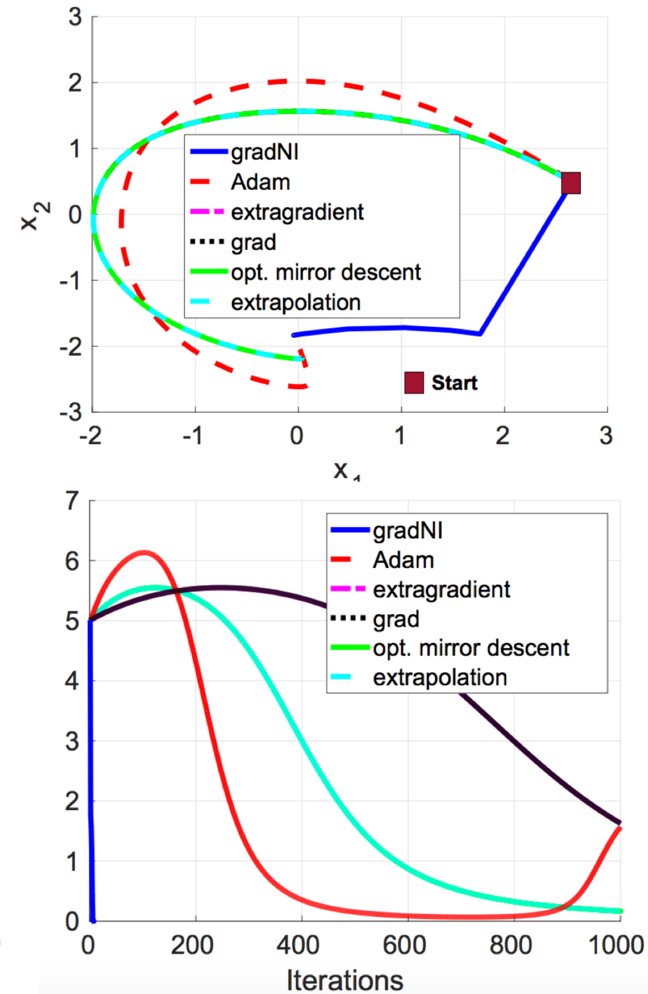
Trajectories and Convergence Using the GNI Function



Bilinear Game



Quadratic Game



Delta GAN Game

MITSUBISHI ELECTRIC
Changes for the Better

for a greener tomorrow

Conclusions and Future Work

- We presented a novel surrogate function -- Gradient-based Nikaido Isoda function for reformulating N-player games that:
  - Vanishes only at the first-order Nash points of the original games
  - Provide error-bounds for various popular game settings
- We presented empirical results comparing the convergence of GNI to zeros of GNI function (Stationary Nash Points).

- Future work will explore
  - Convergence of GNI in a stochastic setting
  - Constrained game payoffs
  - Applications to standard Generative adversarial networks