Monotone Near-Zero-Sum Games: A Generalization of Convex-Concave Minimax

Ruichen Luo, Sebastian Stich, and Krishnendu Chatterjee

rluo@ista.ac.at stich@cispa.de krishnendu.chatterjee@ista.ac.at

Institute of Science and Technology Austria

CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

Optimization Problem

The Two-Player *Nash Equilibrium Problem* in which

Player 1 maximizes $u_1(\mathbf{z}) = (-g - h)(\mathbf{z})$ over $\mathbf{x} \in X$,
Player 2 maximizes $u_2(\mathbf{z}) = (-g + h)(\mathbf{z})$ over $\mathbf{y} \in Y$,

or the *Variational Inequality Problem* of

$$\mathcal{F}(\mathbf{z}) = \nabla g(\mathbf{z}) + \mathcal{H}(\mathbf{z}), \quad \mathbf{z} \in X \times Y, \qquad (1)$$

where $\mathcal{H}(\mathbf{z}) \equiv [\nabla_{\mathbf{x}} h(\mathbf{z}), -\nabla_{\mathbf{y}} h(\mathbf{z})]$, $\mathbf{z} \equiv (\mathbf{x}, \mathbf{y})$, and

1. X and Y are compact and convex, with respective diameters D_X and D_Y;
2. ∇g is δ-Lipschitz continuous and $g(\cdot,\cdot)$ is jointly convex;
3. $\mathcal{H}$ is L-Lipschitz continuous ($\delta \le L$) and h is μ-strongly convex-ν-strongly concave.

Goal: Find an ε-approximate Nash equilibrium.

Special cases:

- Coupling part $g = 0$: $\min_{\mathbf{x}\in X} \max_{\mathbf{y}\in Y} h(\mathbf{x}, \mathbf{y})$;
- Zero-sum part $h = 0$: $\min_{\mathbf{z}\in X\times Y} g(\mathbf{x}, \mathbf{y})$.

Prior Results

General-sum games ($\delta = L$):

- Extragradient [Tse95]: $\widetilde{\mathcal{O}}\left(\frac{L}{\min\{\mu,\nu\}} \cdot \log\left(\frac{1}{\varepsilon}\right)\right)$.

Acceleration in zero-sum games ($\delta = 0$):

- Bilinear cases (Primal-Dual Method [CP11]):

$$\widetilde{\mathcal{O}}\left(\frac{L}{\sqrt{\mu\nu}} \cdot \log\left(\frac{1}{\varepsilon}\right)\right);$$

- General cases (Catalyst [KG22]):

$$\widetilde{\mathcal{O}}\left(\frac{L}{\sqrt{\mu\nu}} \cdot \log\left(\frac{1}{\varepsilon}\right)\right).$$

- Matching lower bound:

$$\widetilde{\Omega}\left(\frac{L}{\sqrt{\mu\nu}} \cdot \log\left(\frac{1}{\varepsilon}\right)\right).$$

Question:

CAN WE ACCELERATE IN NON-ZERO-SUM GAMES?

REFERENCES:

[BN11] Adam M Brandenburger and Barry J Nalebuff. *Co-opetition*. 2011.

[CP11] Antonin Chambolle and Thomas Pock. "A first-order primal-dual algorithm for convex problems with applications to imaging". In: (2011).

[KG22] Dmitry Kovalev and Alexander Gasnikov. "The first optimal algorithm for smooth and strongly-convex-strongly-concave minimax optimization". In: (2022).

[Tse95] Paul Tseng. "On linear convergence of iterative methods for the variational inequality problem". In: (1995).

Motivating Applications

- **MATRIX GAMES WITH TRANSACTION FEE**
  - Player 1 maximizes $u_1 = \langle \mathbf{A}\mathbf{x}, \mathbf{y}\rangle + \mathcal{R}(\mathbf{x}, \mathbf{y})$;
    Player 2 maximizes $u_2 = \langle \mathbf{B}\mathbf{x}, \mathbf{y}\rangle - \mathcal{R}(\mathbf{x}, \mathbf{y})$.
  - $\|\mathbf{A}\|_2 \le L$, $\|\mathbf{B}\|_2 \le L$; $\left\|\frac{\mathbf{A}+\mathbf{B}}{2}\right\|_2 \le \beta$;
  - $\mathcal{R}$ is L-smooth and μ-strongly concave-ν-strongly convex.
  - **Examples:** transaction fee, tax rates.
  - **Assume:** $\beta \le \frac{1}{2}\sqrt{\mu\nu}$.
  - Classic rates: $\widetilde{\mathcal{O}}\left(\frac{L}{\min\{\mu,\nu\}} \cdot \log\left(\frac{1}{\varepsilon}\right)\right)$.
  - Our rates:

$$\widetilde{\mathcal{O}}\left(\left(\frac{L}{\sqrt{\mu\nu}} + \frac{L}{\min\{\mu,\nu\}} \cdot \frac{\beta}{\sqrt{\mu\nu}}\right) \cdot \log^2\left(\frac{1}{\varepsilon}\right)\right).$$

- **SEMI-COOPERATION**
  - Player 1 maximizes $u_1 = -g(\mathbf{z}) - h(\mathbf{z})$;
    Player 2 maximizes $u_2 = -g(\mathbf{z}) + h(\mathbf{z})$.
  - h is the **competition payoff**: L-smooth and μ-strongly convex-ν-strongly concave;
  - g is the **additional incentive**: β-smooth.
  - **Example:** co-opetition [BN11].
  - **Assume:** $\beta = o\left(\min\{\frac{\mu}{2}, \frac{\nu}{2}\}\right)$.
  - Classic rates: $\widetilde{\mathcal{O}}\left(\frac{L}{\min\{\mu,\nu\}} \cdot \log\left(\frac{1}{\varepsilon}\right)\right)$.
  - Our rates:

$$\widetilde{\mathcal{O}}\left(\frac{L}{\sqrt{\mu\nu}} \cdot \log^2\left(\frac{1}{\varepsilon}\right)\right).$$

Algorithm and Convergence Analysis

Potential Function: For all $\mathbf{z} = (\mathbf{x}, \mathbf{y}) \in X \times Y$,

$$\Delta(\mathbf{z}) \equiv \max_{\widetilde{\mathbf{z}}=(\widetilde{\mathbf{x}},\widetilde{\mathbf{y}})\in X\times Y} \underbrace{g(\mathbf{z}) - g(\widetilde{\mathbf{z}})}_{\text{jointly convex coupling}} + \underbrace{h(\mathbf{x}, \widetilde{\mathbf{y}}) - h(\widetilde{\mathbf{x}}, \mathbf{y})}_{\text{convex-concave zero-sum}}.$$

Proposition 1. *For all $\mathbf{z} = (\mathbf{x}, \mathbf{y}) \in X \times Y$, we have $\Delta(\mathbf{z}) \ge 0$ and*

$$2\Delta(\mathbf{z}) \ge \max_{\widetilde{\mathbf{z}}=(\widetilde{\mathbf{x}},\widetilde{\mathbf{y}})\in X\times Y} u_1(\widetilde{\mathbf{x}}, \mathbf{y}) - u_1(\mathbf{x}, \mathbf{y}) + u_2(\mathbf{x}, \widetilde{\mathbf{y}}) - u_2(\mathbf{x}, \mathbf{y}).$$

Proposition 2. *Let $\mathbf{z}^* \in X \times Y$. Then, $\mathbf{z}^*$ is the Nash equilibrium if and only if $\Delta(\mathbf{z}^*) = 0$.*

Technical insights. At every iteration t, we linearize the coupling part in the potential function at $\mathbf{z}_t$:

$$\min_{\mathbf{z}\in X\times Y} \Delta(\mathbf{z}) \rightsquigarrow \min_{\mathbf{z}\in X\times Y} \max_{\widetilde{\mathbf{z}}\in X\times Y} \langle \nabla g(\mathbf{z}_t), \mathbf{z} - \widetilde{\mathbf{z}}\rangle + \frac{1}{2\eta_t}\left(\|\mathbf{z} - \mathbf{z}_t\|_2^2 - \|\widetilde{\mathbf{z}} - \mathbf{z}_t\|_2^2\right) + h(\mathbf{x}, \widetilde{\mathbf{y}}) - h(\widetilde{\mathbf{x}}, \mathbf{y}),$$

and note that this minimax optimization can be fully decomposed into two separate problems:

$$\min_{\mathbf{x}\in X} \max_{\widetilde{\mathbf{y}}\in Y} \langle \nabla g(\mathbf{z}_t), (\mathbf{x}, -\widetilde{\mathbf{y}})\rangle + h(\mathbf{x}, \widetilde{\mathbf{y}}) + \frac{1}{2\eta_t}\left(\|\mathbf{x} - \mathbf{x}_t\|_2^2 - \|\widetilde{\mathbf{y}} - \mathbf{y}_t\|_2^2\right) = \varphi_t(\mathbf{x}, \widetilde{\mathbf{y}}), \text{ and}$$

$$\min_{\mathbf{y}\in Y} \max_{\widetilde{\mathbf{x}}\in X} \langle \nabla g(\mathbf{z}_t), (-\widetilde{\mathbf{x}}, \mathbf{y})\rangle - h(\widetilde{\mathbf{x}}, \mathbf{y}) + \frac{1}{2\eta_t}\left(-\|\widetilde{\mathbf{x}} - \mathbf{x}_t\|_2^2 + \|\mathbf{y} - \mathbf{y}_t\|_2^2\right) = -\varphi_t(\widetilde{\mathbf{x}}, \mathbf{y}).$$

Moreover, by Sion's minimax theorem, (after simple substitutions) these two separate problems unify into a *single* saddle point problem of $\min_{\mathbf{x}\in X} \max_{\mathbf{y}\in Y} \varphi_t(\mathbf{x}, \mathbf{y})$.

Thus, we obtain a clean, black-box reduction that
reduces near-zero-sum games to a sequence of zero-sum subproblems.

Algorithm 1 Iterative Coupling Linearization (ICL)

Require: $\mathbf{x}_0 \in X$, $\mathbf{y}_0 \in Y$.

1: **for** $t = 0, 1, \cdots, T-1$ **do**

2: Let $\varphi_t(\mathbf{x}, \mathbf{y}) \equiv \langle \nabla_{\mathbf{x}} g(\mathbf{x}_t, \mathbf{y}_t), \mathbf{x}\rangle + \frac{1}{2\eta_t}\|\mathbf{x}_t - \mathbf{x}\|_2^2 + h(\mathbf{x}, \mathbf{y}) - \langle \nabla_{\mathbf{y}} g(\mathbf{x}_t, \mathbf{y}_t), \mathbf{y}\rangle - \frac{1}{2\eta_t}\|\mathbf{y}_t - \mathbf{y}\|_2^2$.

3: Find an **inexact saddle point** $\mathbf{z}_{t+1} \in X \times Y$ of φ_t s.t.

$$\langle \nabla_{\mathbf{x}} \varphi_t(\mathbf{z}_{t+1}), \mathbf{x}_{t+1} - \mathbf{x}\rangle - \langle \nabla_{\mathbf{y}} \varphi_t(\mathbf{z}_{t+1}), \mathbf{y}_{t+1} - \mathbf{y}\rangle \le \varepsilon_t, \quad \text{for all } \mathbf{z} \in X \times Y.$$

4: **end for**

Theorem 1 (Main theoretical result). *Denote $\eta = \min\left\{\frac{1}{\delta}, \frac{1}{\min\{\mu,\nu\}}\right\}$ and $\theta = \frac{\min\{\mu,\nu\}}{\eta^{-1}+\min\{\mu,\nu\}}$. Let $\eta_t = \eta$ and $\varepsilon_t = \frac{\theta\varepsilon}{4\eta}$, for all $t \in [0, T-1] \cap \mathbb{Z}$. For strongly monotone δ-near-zero-sum games, for $T \ge \frac{1}{\theta}\log\frac{2D^2}{\varepsilon}$, the outer loop iterates of Algorithm 1 converge to an ε-accurate Nash equilibrium with the number of gradient queries bounded by*

$$\mathcal{O}\left(\left(\frac{L}{\sqrt{\mu\nu}} + \frac{L}{\min\{\mu,\nu\}} \cdot \min\left\{1, \sqrt{\frac{\delta}{\mu+\nu}}\right\}\right) \cdot \log\left(\frac{LD^2}{\min\{\mu,\nu\}\cdot\varepsilon}\right) \log\left(\frac{D^2}{\varepsilon}\right)\right).$$

Acceleration conditioning: $\delta + \min\{\mu,\nu\} \ll \mu + \nu$. [NEAR-ZERO-SUM]