Minimax Optimization

Nonsmooth Composite Nonconvex Concave

Jiajin Li (Stanford)

Joint work with Linglingzhi Zhu (CUHK) and Anthony Man-Cho So (CUHK)

Our Focus

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} F(x, y) \qquad \text{(NC-C)}$$

- $F(x, y)$ is **nonconvex** in x but concave in y
- $\mathcal{X}$ is closed convex
- $\mathcal{Y}$ is convex compact

Our Focus

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} F(x, y) \qquad \text{(NC-C)}$$

- $F(x,y)$ is **nonconvex** in x but concave in y
- $\mathcal{X}$ is closed convex
- $\mathcal{Y}$ is convex compact

No gradient information?

Our Focus

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} F(x, y) \qquad \text{(NC-C)}$$

- $F(x, y)$ is **weakly convex** in x but concave in y
- $\mathcal{X}$ is closed convex
- $\mathcal{Y}$ is convex compact

Subgradient?

Our Focus

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} F(x, y) \qquad \text{(NC-C)}$$

- $F(x, y)$ is **weakly convex** in x but concave in y
- $\mathcal{X}$ is closed convex
- $\mathcal{Y}$ is convex compact

Subgradient --- $\mathcal{O}(\epsilon^{-6})$

[R-Liu-Yang-Lin, 2020, Lin-Jin-Jordan, 2022]

Motivation and Limitation

Can we design an algorithm for nonsmooth NC-C problems, which match the smooth case or the lower bound $\mathcal{O}(\epsilon^{-2})$?

[Zhang et al. NeurIPS 2020]

Motivation and Limitation

Can we design an algorithm for nonsmooth NC-C problems, which match the smooth case or the lower bound $\mathcal{O}(\epsilon^{-2})$?

Yes! Composite structure can help us!

[Zhang et al. NeurIPS 2020]

Nonsmooth Composite

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} F(x, y) \quad \text{(NC-C)}$$

- **Composite**: $F(\cdot, y) = h_y \circ c_y$.
- c_y is continuous differentiable with Lipschitz continuous Jacobian map ("***L*-smooth**").
- h_y is **convex Lipschitz**, e.g., $||\cdot||_1$.
- $F(x, \cdot)$ is standard ***L*-smooth**.

Smoothed GDA

- Potential function:

$$F_r(x, y, z) := F(x, y) + \frac{r}{2}||x - z||^2.$$

- Smoothed gradient descent ascent:

$$x^{k+1} = \mathrm{Proj}_{\mathcal{X}}\left(x^k - \tau \nabla_x F_r\left(x^k, y^k, z^k\right)\right)$$

$$y^{k+1} = \mathrm{Proj}_{\mathcal{Y}}\left(x^k + \alpha \nabla_y F\left(x^{k+1}, y^k\right)\right)$$

$$z^{k+1} = z^k + \beta(x^{k+1} - z^k)$$

Stabilized Sequence!

[Zhang et al. NeurIPS 2020]

Smoothed GDA

- Potential function:

$$F_r(x, y, z) := F(x, y) + \frac{r}{2}||x - z||^2.$$

- S

…zed Sequence!

1. Achieve the optimal rate $\mathcal{O}(\epsilon^{-2})$ when the dual function is **polyhedral**!
2. For general NC-C, improve vanilla GDA $\mathcal{O}(\epsilon^{-6})$ to $\boldsymbol{\mathcal{O}(\epsilon^{-4})}$!

Our Contributions:

- Extend smoothed GDA to **nonsmooth composite NC-C.**
- Develop a **new analysis framework** based on Kurdyka-Łojasiewicz exponent $\theta \in [0,1)$ of the dual function.
- **Optimal** convergence rate when $\theta \in [0, ½]$.

Our Contributions:

- Extend smoothed GDA to **nonsmooth composite NC-C.**
- Develop a **new analysis framework** based on Kurdyka-Łojasiewicz exponent $\theta \in [0,1)$ of the dual function.
- **Optimal** convergence rate when $\theta \in [0, \frac{1}{2}]$.

Algorithm Design:

$$x^{k+1} = \arg\min_{x \in \mathcal{X}} \boldsymbol{F}_{x^k,\lambda}(\boldsymbol{x}) + \frac{r}{2}\left|\left|x - z^k\right|\right|^2$$

$$y^{k+1} = \mathrm{Proj}_{\mathcal{Y}}\left(x^k + \alpha \nabla_y F\left(x^{k+1}, y^k\right)\right)$$

$$z^{k+1} = z^k + \beta\left(x^{k+1} - z^k\right)$$

Proximal Linear Scheme

$$\boldsymbol{F}_{x^k,\lambda}(\boldsymbol{x}) := h_{y^k}\left(c_{y^k}\left(x^k\right) + \nabla c_{y^k}\left(x^k\right)^T\left(x - x^k\right)\right) + \frac{\lambda}{2}\left|\left|x - x^k\right|\right|^2$$

[Drusvyatskiy-Paquette MP 2019]

Technical Difficulty:

Theorem 1. [**Primal error bound condition** (without gradient Lip)]

$$\mathrm{dist}\left(x^{k+1}, x(y^k, z^k)\right) \leq \xi ||x^{k+1} - x^k||,$$

[optimality residual]

where $x(y^k, z^k) := \arg\min_{x \in \mathcal{X}} F(x, y^k) + \frac{r}{2}\left|\left|x - z^k\right|\right|^2$.

Technical Difficulty:

Theorem 1. [**Primal error bound condition** (without gradient Lip)]

$$\text{dist}\left(x^{k+1}, x(y^k, z^k)\right) \leq \xi ||x^{k+1} - x^k||,$$

[optimality residual]

Explicit bound for ξ?

[Drusvyatskiy-Lewis MOR 2018]

Our Contributions:

- Extend smoothed GDA to **nonsmooth composite NC-C.**
- Develop a **new analysis framework** based on Kurdyka-Łojasiewicz exponent $\theta \in [0,1)$ of the dual function.
- **Optimal** convergence rate when $\theta \in [0, ½]$.

Dual Error Bound:

Explicitly control the trade-off between the decrease in the primal and the increase in the dual.

Definition (KL exponent of the dual function): there exists a constant $\mu > 0$ such that

$$\mathrm{dist}\Big(0, -\nabla_y F(x,y) + \partial I_{\mathcal{Y}}(y)\Big) \geq \mu \left(\max_{y' \in \mathcal{Y}} F(x,y') - F(x,y)\right)^{\theta}.$$

Dual Error Bound:

Theorem 2. If $\theta \in (0,1)$,

$$||x^\star(z) - x(y_+(z), z)|| \leq \omega ||y - y_+(z)||^{\frac{1}{2\theta}}.$$

[primal update] **[dual update]**

Otherwise ($\theta = 0$),

$$||x^\star(z) - x(y_+(z), z)|| \leq \omega ||y - y_+(z)||.$$

$$x(y, z) := \arg\min_{x \in \mathcal{X}} F(x, y) + \frac{r}{2}||x - z||^2$$

$$x^\star(z) := \arg\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} F(x, y) + \frac{r}{2}||x - z||^2$$

$$y_+(z) := \mathrm{Proj}_{\mathcal{Y}}(y + \alpha \nabla y\, F(x(y, z), y, z)$$

Dual Error Bound:

Theorem 2. If $\theta \in (0,1)$,

$$||x^\star(z) - x(y_+(z), z)|| \le \omega \big||y - y_+(z)|\big|^{\frac{1}{2\theta}}.$$

[primal update] [dual update]

Otherwise ($\theta = 0$),

$$||x^\star(z) - x(y_+(z), z)|| \le \omega \big||y - y_+(z)|\big|.$$

$\mathcal{O}\left(\epsilon^{-2\max(2\theta,1)}\right)$!

$$x(y, z) := \arg\min_{x \in \mathcal{X}} F(x, y) + \frac{r}{2}||x - z||^2$$

$$x^\star(z) := \arg\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} F(x, y) + \frac{r}{2}||x - z||^2$$

$$y_+(z) := \mathrm{Proj}_{\mathcal{Y}}(y + \alpha \nabla y\, F(x(y, z), y, z)$$

Our Contributions:

- Extend smoothed GDA to **nonsmooth composite NC-C**;
- Develop a **new analysis framework** based on Kurdyka-Łojasiewicz exponent $\theta \in [0,1)$ of the dual function;
- **Optimal** convergence rate when $\theta \in [0, ½]$

When $\theta = 0$, the dual function is **polyhedral**.
The technique developed in [zhang et al, 2020] cannot handle the nonsmooth primal function!

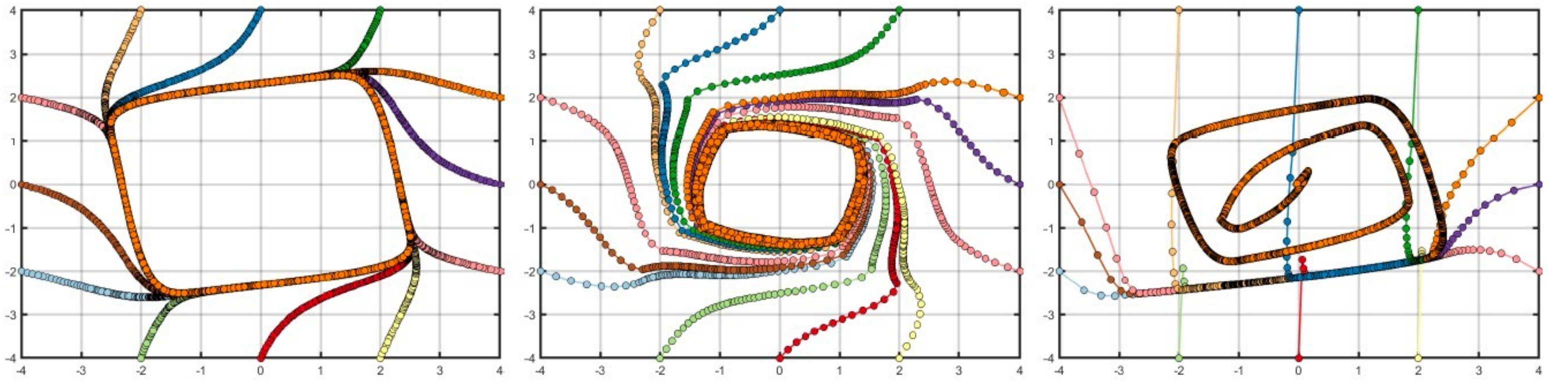
Our Contributions:

- Extend smoothed GDA to **nonsmooth composite NC-C**;
- Develop a **new analysis framework** based on Kurdyka-Łojasiewicz exponent $\theta \in [0,1)$ of the dual function;
- **Optimal** convergence rate when $\theta \in [0, 1/2]$

When $\theta = 1/2$, the dual function satisfies **PL inequaility.** The technique developed in [Yang et al, 2022] only works for the unconstrained case.

New Results on NC-NC

We develop the first provably convergent algorithm, which gets rid of the limiting cycle without requiring any additional conditions (e.g., PL, Weak Minty VI, Dominance condition).



(*d*) Damped EG (*e*) CurvatureEG+ (*f*) Double Smoothed GDA

Taoli Zheng, Linglingzhi Zhu, Anthony Man-Cho So, Jose Blanchet, **Jiajin Li**
Escape from the Limit Cycle: Double Smoothed GDA for Nonconvex-Nonconcave Minimax Optimization