Minimax Optimization

Nonsmooth Composite Nonconvex-Concave

Jiajin Li

Department of Management Science and Engineering
Stanford University

ICCOPT, July 2022

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

Our Focus

We are interested in studying nonconvex concave minimax problems of the form

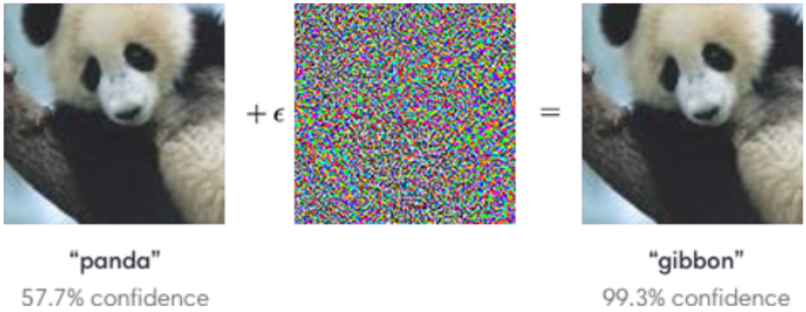
$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} F(x, y), \tag{1}$$

where $F : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ is nonconvex in x but concave in y, $\mathcal{X} \subseteq \mathbb{R}^n$ is closed convex and $\mathcal{Y} \subseteq \mathbb{R}^d$ is convex compact.

Applications

Problem (1) has attracted intense attention across both optimization and machine learning communities.

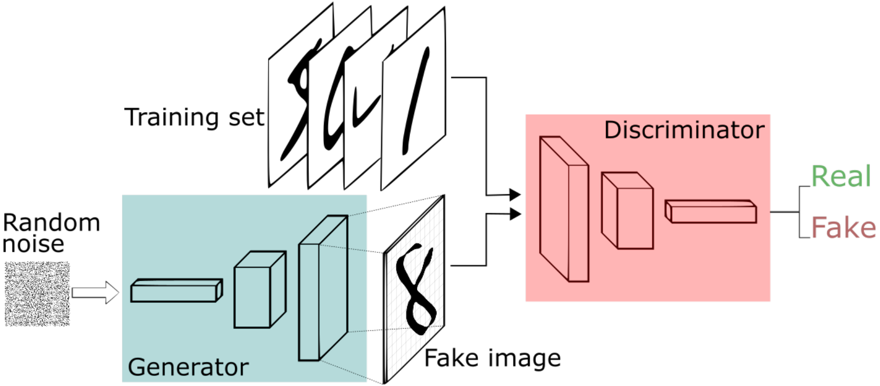
- Adversarial Training:



Applications

Problem (1) has attracted intense attention across both optimization and machine learning communities.

- Generative Adversarial Network:



Applications

- Distributionally Robust Optimization (DRO):

$$\min_{x \in \mathcal{X}} \max_{Q \in \mathcal{U}(\mathbb{P}_N)} \mathbb{E}_{\xi \sim Q}[f(x;\xi)]$$

  - $\mathbb{P}_N$: empirical distribution;
  - $\mathcal{U}(\mathbb{P}_N)$: ambiguity set defined by a host of probablity metrics, e.g., f-divergence, Wasserstein, etc

$$\mathcal{U}(\mathbb{P}_N) = \{Q : d(Q, \mathbb{P}_N) \leqslant r\}.$$

Gradient Descent Ascent (GDA)

- $$x^{k+1} = x^k - \alpha_k \nabla_x F(x^k, y^k),$$
  $$y^{k+1} = y^k + \tau_k \nabla_y F(x^{k+1}, y^k),$$
  where α_k and τ_k are the step sizes.
- Strongly-Concave [Lin et al. 2020]:
  GDA can generate an ϵ-stationary solution with iteration complexity $\mathcal{O}(\epsilon^{-2})$ — matching the optimal!
- Concave: GDA suffers from oscillation — diminishing step size strategies $\mathcal{O}(\epsilon^{-6})$ [Lin et al. 2020], smoothing $\mathcal{O}(\epsilon^{-4})$ [Zhang et al. 2020] $\cdots$

Gradient Descent Ascent (GDA)

- $$x^{k+1} = x^k - \alpha_k \nabla_x F(x^k, y^k),$$
  $$y^{k+1} = y^k + \tau_k \nabla_y F(x^{k+1}, y^k),$$
  where α_k and τ_k are the step sizes.
- Strongly-Concave [Lin et al. 2020]:
  GDA can generate an ϵ-stationary solution with iteration complexity $\mathcal{O}(\epsilon^{-2})$ — matching the optimal!
- Concave: GDA suffers from oscillation — diminishing step size strategies $\mathcal{O}(\epsilon^{-6})$ [Lin et al. 2020], smoothing $\mathcal{O}(\epsilon^{-4})$ [Zhang et al. 2020] $\cdots$

Gradient Descent Ascent (GDA)

- $$x^{k+1} = x^k - \alpha_k \nabla_x F(x^k, y^k),$$
  $$y^{k+1} = y^k + \tau_k \nabla_y F(x^{k+1}, y^k),$$
  where α_k and τ_k are the step sizes.
- Strongly-Concave [Lin et al. 2020]:
  GDA can generate an ϵ-stationary solution with iteration complexity $\mathcal{O}(\epsilon^{-2})$ — matching the optimal!
- Concave: GDA suffers from oscillation — diminishing step size strategies $\mathcal{O}(\epsilon^{-6})$ [Lin et al. 2020], smoothing $\mathcal{O}(\epsilon^{-4})$ [Zhang et al. 2020] $\cdots$

Smoothed GDA

- Iterative Scheme:

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha_k[\nabla_x F(x^k, y^k) + \gamma(x^k - z^k)], \\
y^{k+1} &= \mathrm{proj}_{\mathcal{Y}}(y^k + \tau_k \nabla_y F(x^{k+1}, y^k)), \\
z^{k+1} &= z^k + \beta(x^{k+1} - z^k),
\end{aligned}
$$

where α_k and τ_k are the step sizes, β is the extrapolation parameter.

- PŁ Condition [Yang et al. 2022]:
Smoothed GDA can generate an ϵ-stationary solution with iteration complexity $\mathcal{O}(\epsilon^{-2})$ — matching the optimal!

Smoothed GDA

- Iterative Scheme:

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha_k[\nabla_x F(x^k, y^k) + \gamma(x^k - z^k)], \\
y^{k+1} &= \mathrm{proj}_{\mathcal{Y}}(y^k + \tau_k \nabla_y F(x^{k+1}, y^k)), \\
z^{k+1} &= z^k + \beta(x^{k+1} - z^k),
\end{aligned}
$$

  where α_k and τ_k are the step sizes, β is the extrapolation parameter.

- PŁ Condition [Yang et al. 2022]:
  Smoothed GDA can generate an ϵ-stationary solution with iteration complexity $\mathcal{O}(\epsilon^{-2})$ — matching the optimal!

Limitations and Motivations

- (Smoothed) GDA relies on the gradient Lipschitz condition.
- Proximally guided stochastic subgradient method [Rafique et al. 2021] has been proposed for general **nonsmooth** weakly convex-concave problems but suffers from the slow iteration complexity $\mathcal{O}(\epsilon^{-6})$.

Can we design a provably efficient algorithm to address nonsmooth nonconvex-concave problems, which matches the lower bound $\mathcal{O}(\epsilon^{-2})$?

Limitations and Motivations

- (Smoothed) GDA relies on the gradient Lipschitz condition.
- Proximally guided stochastic subgradient method [Rafique et al. 2021] has been proposed for general nonsmooth weakly convex-concave problems but suffers from the slow iteration complexity $\mathcal{O}(\epsilon^{-6})$.

Can we design a provably efficient algorithm to address nonsmooth nonconvex-concave problems, which matches the lower bound $\mathcal{O}(\epsilon^{-2})$?

Limitations and Motivations

- (Smoothed) GDA relies on the gradient Lipschitz condition.
- Proximally guided stochastic subgradient method [Rafique et al. 2021] has been proposed for general nonsmooth weakly convex-concave problems but suffers from the slow iteration complexity $\mathcal{O}(\epsilon^{-6})$.

Can we design a provably efficient algorithm to address nonsmooth nonconvex-concave problems, which matches the lower bound $\mathcal{O}(\epsilon^{-2})$?

Limitations and Motivations

- (Smoothed) GDA relies on the gradient Lipschitz condition.
- Proximally guided stochastic subgradient method [Rafique et al. 2021] has been proposed for general nonsmooth weakly convex-concave problems but suffers from the slow iteration complexity $\mathcal{O}(\epsilon^{-6})$.

Can we design a provably efficient algorithm to address nonsmooth nonconvex-concave problems, which matches the lower bound $\mathcal{O}(\epsilon^{-2})$?

Nonsmooth Composite

Nonconvex-Concave Minimax

Main Results

Table 1: Comparison of the iteration complexities of smoothed PLDA proposed in this paper and other related methods under different settings for solving $\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} F(x, y)$.

	Primal Func.	**Dual Func.**	**Iter. Compl.**[1]	**Add. Asm.**
GDA	L-smooth	concave	$\mathcal{O}(\epsilon^{-6})$	$\mathcal{X} = \mathbb{R}^n$
Smoothed GDA	L-smooth	concave	$\mathcal{O}(\epsilon^{-4})$	—
PG-SMD	weakly-convex	concave	$\mathcal{O}(\epsilon^{-6})$	$\mathcal{X}$ bounded
This paper	nonsmooth composite	concave	$\mathcal{O}(\epsilon^{-4})$	—
GDA	L-smooth	strongly-concave	$\mathcal{O}(\epsilon^{-2})$	$\mathcal{X} = \mathbb{R}^n$
Smoothed GDA	L-smooth	PŁ condition	$\mathcal{O}(\epsilon^{-2})$	$\mathcal{Y} = \mathbb{R}^d$
This paper	nonsmooth composite	KŁ exponent $\theta = \frac{1}{2}$	$\mathcal{O}(\epsilon^{-2})$	—

Problem Setup

- (**Primal Function**) $F(\cdot, y) := h_y \circ c_y$, where $c_y : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable with L_c-Lipschitz continuous Jacobian map for all $y \in \mathcal{Y}$ on $\mathcal{X}$:

$$\|\nabla c_y(x) - \nabla c_y(x')\| \leqslant L_c \|x - x'\| \quad \text{for all } x, x' \in \mathcal{X},$$

  and $h_y : \mathbb{R}^m \to \mathbb{R}$ for any $y \in \mathcal{Y}$ is a convex and L_h-Lipschitz continuous function satisfying

$$|h_y(z) - h_y(z')| \leqslant L_h \|z - z'\|, \quad \text{for all } z, z' \in \mathbb{R}^m.$$

- For example, $h_y = \|\cdot\|_p$ where $p = \{1, 2, +\infty\}$.

Problem Setup

- (**Dual Function**) $F(x, \cdot)$ is concave and continuously differentiable on $\mathcal{Y}$ with $\nabla_y F(\cdot, \cdot)$ being L-Lipschitz continuous on $\mathcal{X} \times \mathcal{Y}$, i.e.,

$$\|\nabla_y F(x, y) - \nabla_y F(x', y')\| \leqslant L\|(x, y) - (x', y')\|$$

for all $(x, y), (x', y') \in \mathcal{X} \times \mathcal{Y}$.

Smoothed Proximal Linear Descent Ascent (PLDA)

Smoothed PLDA

Due to the composite structure $h_y \circ c_y$, there is no available gradient information to rely on. Instead, it is natural to invoke the proximal linear scheme for the primal update.

- Potential function:

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

- Proximal linear update:

$$x^{k+1} = \arg\min_{x \in \mathcal{X}} h_{y^k}\left(c_{y^k}(x^k) + \nabla c_{y^k}(x^k)^\top (x - x^k)\right) + \frac{\lambda}{2}\|x - x^k\|^2 + \frac{r}{2}\|x - z^k\|^2.$$

Convergence Analysis

Lyapunov Function

Define a Lyapunov function function as

$$\Phi_r(x, y, z) := \underbrace{F_r(x, y, z) - d_r(y, z)}_{\text{Primal Descent}} + \underbrace{p_r(z) - d_r(y, z)}_{\text{Dual Ascent}} + \underbrace{p_r(z)}_{\text{Proximal Descent}} .$$

- $d_r(y, z) := \min_{x \in \mathcal{X}} F_r(x, y, z);$
- $p_r(z) := \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} F_r(x, y, z);$

Lipschitz-type Primal Error Bound Condition

Main Technical Results I

For any $k \geqslant 0$, it holds that

$$\|x^{k+1} - x_r(y^k, z^k)\| \leqslant \zeta \|x^k - x^{k+1}\|, \qquad (2)$$

where $\zeta := \frac{2(r-L)^{-1}+(\lambda+L)^{-1}}{(\lambda+L)^{-1}} \left(\sqrt{\frac{2L}{\lambda+L}} + 1 \right)$ and $x_r(y, z) := \operatorname{argmin}_{x \in \mathcal{X}} F_r(x, y, z)$.

- Smooth case: Luo-Tseng error bound condition

$$\|x^{k+1} - x_r(y^k, z^k)\| \leqslant \zeta \|x^k - \underbrace{\text{proj}_{\mathcal{X}}(x^k - c\nabla_x F_r(x^k, y^k, z^k))}_{x^{k+1}}\|,$$

Lipschitz-type Primal Error Bound Condition

Main Technical Results I

For any $k \geqslant 0$, it holds that

$$\|x^{k+1} - x_r(y^k, z^k)\| \leqslant \zeta \|x^k - x^{k+1}\|, \tag{2}$$

where $\zeta := \frac{2(r-L)^{-1}+(\lambda+L)^{-1}}{(\lambda+L)^{-1}}\left(\sqrt{\frac{2L}{\lambda+L}}+1\right)$ and $x_r(y,z) := \operatorname{argmin}_{x\in\mathcal{X}} F_r(x,y,z)$.

- Smooth case: Luo-Tseng error bound condition

$$\|x^{k+1} - x_r(y^k, z^k)\| \leqslant \zeta \|x^k - \underbrace{\operatorname{proj}_{\mathcal{X}}(x^k - c\nabla_x F_r(x^k, y^k, z^k))}_{x^{k+1}}\|,$$

Sufficient Decrease Property

Proposition

$r \geqslant 3L$, $\lambda \geqslant L$, $\beta \leqslant \min\left\{\frac{1}{28}, \frac{(r-L)^2}{32\alpha r(r+L)^2}\right\}$ *and* $\alpha \leqslant \min\left\{\frac{1}{10L}, \frac{1}{4L\zeta^2}\right\}$.
Then for any $k \geqslant 0$,

$$\Phi_r^k - \Phi_r^{k+1} \geqslant \frac{\lambda}{16}\|x^k - x^{k+1}\|^2 + \frac{1}{8\alpha}\|y^k - y_+^k(z^k)\|^2 + \frac{4r}{7\beta}\|z^k - z^{k+1}\|^2 - 28r\beta\|x_r^*(z^k) - x_r(y_+^k(z^k), z^k)\|^2,$$

where $y_+(z) := \text{proj}_{\mathcal{Y}}\left(y + \alpha\nabla_y F_r(x_r(y,z), y, z)\right)$ *and*
$x_r^*(z) := \operatorname*{argmin}_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} F_r(x,y,z)$.

KŁ Exponent θ for the Dual Function

Motivation: explicitly control the trade-off between the decrease in the primal and the increase in the dual.

Kurdyka-Łojasiewicz (KŁ) Exponent

For any fixed $x \in \mathcal{X}$, the problem $\max_{y \in \mathcal{Y}} F(x, y)$ has a nonempty solution set and a finite optimal value. There exist $\mu > 0$ and $\theta \in [0, 1)$ such that

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

for any $x \in \mathcal{X}, y \in \mathcal{Y}$.

Dual Error Bound Condition

Main Technical Results II

- KŁ exponent $\theta \in (0,1)$:

$$\|x_r^*(z) - x_r(y_+(z), z)\| \leqslant \omega \|y - y_+(z)\|^{\frac{1}{2\theta}},$$

- KŁ exponent $\theta = 0$:

$$\|x_r^*(z) - x_r(y_+(z), z)\| \leqslant \omega' \|y - y_+(z)\|.$$

Stationarity Concept

Definition

The pair $(x, y) \in \mathcal{X} \times \mathcal{Y}$ is an **ϵ-game stationary point (ϵ-GS)** if

$$\|\nabla_x d_r(y, x)\| \leqslant \epsilon \quad \text{and} \quad \mathrm{dist}(0, -\nabla_y F(x, y) + \partial \iota_{\mathcal{Y}}(y)) \leqslant \epsilon.$$

With the aid of our newly developed dual error bound condition, we can clarify the relationship among various stationarity concepts both conceptually and quantitatively.

Stationarity Concept

Definition

The pair $(x, y) \in \mathcal{X} \times \mathcal{Y}$ is an ϵ-game stationary point (ϵ-GS) if

$$\|\nabla_x d_r(y, x)\| \leqslant \epsilon \quad \text{and} \quad \mathrm{dist}(0, -\nabla_y F(x, y) + \partial \iota_{\mathcal{Y}}(y)) \leqslant \epsilon.$$

With the aid of our newly developed dual error bound condition, we can clarify the relationship among various stationarity concepts both conceptually and quantitatively.

Main Theorem — Iteration Complexity

Suppose that $r \geqslant 3L$, $\lambda \geqslant L$, $\beta \leqslant \min\left\{\frac{1}{28}, \frac{(r-L)^2}{32\alpha r(r+L)^2}\right\}$ and $\alpha \leqslant \min\left\{\frac{1}{10L}, \frac{1}{4L\zeta^2}\right\}$. Then for any $k \geqslant 0$,

- General concave: there exists a $k \in [K]$ such that (x^{k+1}, y^{k+1}) is an $\mathcal{O}(K^{-\frac{1}{4}})$-game stationary if $\beta \leqslant K^{-\frac{1}{2}}$.
- KŁ exponent $\theta \in (\frac{1}{2}, 1)$: there exists a $k \in [K]$ such that (x^{k+1}, y^{k+1}) is an $\mathcal{O}(K^{-\frac{1}{4\theta}})$-game stationary if $\beta \leqslant K^{-\frac{2\theta-1}{2\theta}}$.
- KŁ exponent $\theta \in [0, \frac{1}{2}]$: there exists a $k \in [K]$ such that (x^{k+1}, y^{k+1}) is an $\mathcal{O}(K^{-\frac{1}{2}})$-game stationary if $\beta = \mathcal{O}(1)$.

Main Theorem — Iteration Complexity

Suppose that $r \geqslant 3L$, $\lambda \geqslant L$, $\beta \leqslant \min\left\{\frac{1}{28}, \frac{(r-L)^2}{32\alpha r(r+L)^2}\right\}$ and $\alpha \leqslant \min\left\{\frac{1}{10L}, \frac{1}{4L\zeta^2}\right\}$. Then for any $k \geqslant 0$,

- General concave: there exists a $k \in [K]$ such that (x^{k+1}, y^{k+1}) is an $\mathcal{O}(K^{-\frac{1}{4}})$-game stationary if $\beta \leqslant K^{-\frac{1}{2}}$.
- KŁ exponent $\theta \in (\frac{1}{2}, 1)$: there exists a $k \in [K]$ such that (x^{k+1}, y^{k+1}) is an $\mathcal{O}(K^{-\frac{1}{4\theta}})$-game stationary if $\beta \leqslant K^{-\frac{2\theta-1}{2\theta}}$.
- KŁ exponent $\theta \in [0, \frac{1}{2}]$: there exists a $k \in [K]$ such that (x^{k+1}, y^{k+1}) is an $\mathcal{O}(K^{-\frac{1}{2}})$-game stationary if $\beta = \mathcal{O}(1)$.

Main Theorem — Iteration Complexity

Suppose that $r \geqslant 3L$, $\lambda \geqslant L$, $\beta \leqslant \min\left\{\frac{1}{28}, \frac{(r-L)^2}{32\alpha r(r+L)^2}\right\}$ and $\alpha \leqslant \min\left\{\frac{1}{10L}, \frac{1}{4L\zeta^2}\right\}$. Then for any $k \geqslant 0$,

- General concave: there exists a $k \in [K]$ such that (x^{k+1}, y^{k+1}) is an $\mathcal{O}(K^{-\frac{1}{4}})$-game stationary if $\beta \leqslant K^{-\frac{1}{2}}$.
- KŁ exponent $\theta \in (\frac{1}{2}, 1)$: there exists a $k \in [K]$ such that (x^{k+1}, y^{k+1}) is an $\mathcal{O}(K^{-\frac{1}{4\theta}})$-game stationary if $\beta \leqslant K^{-\frac{2\theta-1}{2\theta}}$.
- KŁ exponent $\theta \in [0, \frac{1}{2}]$: there exists a $k \in [K]$ such that (x^{k+1}, y^{k+1}) is an $\mathcal{O}(K^{-\frac{1}{2}})$-game stationary if $\beta = \mathcal{O}(1)$.

Numerical Results

Variation Regularized Wasserstein DRO

$$\min_{\theta} g(\theta) := \mathbb{E}_{\mathbb{P}_N}\left[\ell(y, f_\theta(x))\right] + \rho \max_{i \in [N]} \|\nabla_x \ell(y_i, f_\theta(x_i))\|_p. \tag{3}$$

- $\ell : \mathbb{R} \to \mathbb{R}$ is the loss function;
- $f_\theta : \mathbb{R}^d \to \mathbb{R}$ is the feature mapping;
- $\{(x_i, y_i)\}_{i=1}^N$ is the training dataset and $p = \{1, 2, +\infty\}$;
- closed connection with the Lipschitz constant of deep neural networks;

Variation Regularized Wasserstein DRO

$$\min_{\theta} g(\theta) := \mathbb{E}_{\mathbb{P}_N} [\ell(y, f_\theta(x))] + \rho \max_{i \in [N]} \|\nabla_x \ell(y_i, f_\theta(x_i))\|_p. \qquad (3)$$

- $\ell : \mathbb{R} \to \mathbb{R}$ is the loss function;
- $f_\theta : \mathbb{R}^d \to \mathbb{R}$ is the feature mapping;
- $\{(x_i, y_i)\}_{i=1}^N$ is the training dataset and $p = \{1, 2, +\infty\}$;
- closed connection with the Lipschitz constant of deep neural networks;

Key Difficulties

- It is super challenging for calculating the subdifferential set of the pointwise supremum of an arbitrary family (possibly not differentiable) of (weakly) convex functions.

- Minimax reformulation technique:

$$\min_{\theta} \max_{w \in \Delta_N} \mathbb{E}_{\mathbb{P}_N} [\ell(y, f_\theta(x))] + \rho \sum_{i=1}^{N} w_i \|\nabla_x \ell(y_i, f_\theta(x_i))\|_p, \quad (4)$$

which can be recast into the general form (1) that we investigated in this talk.

Key Difficulties

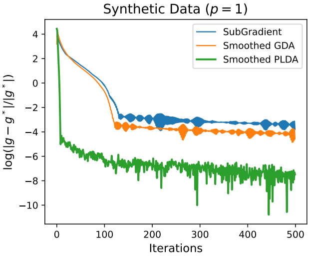
- It is super challenging for calculating the subdifferential set of the pointwise supremum of an arbitrary family (possibly not differentiable) of (weakly) convex functions.
- Minimax reformulation technique:

$$\min_{\theta} \max_{w \in \Delta_N} \mathbb{E}_{\mathbb{P}_N} [\ell(y, f_\theta(x))] + \rho \sum_{i=1}^{N} w_i \|\nabla_x \ell(y_i, f_\theta(x_i))\|_p, \quad (4)$$

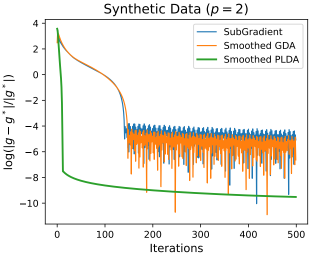
which can be recast into the general form (1) that we investigated in this talk.

Linear Regression

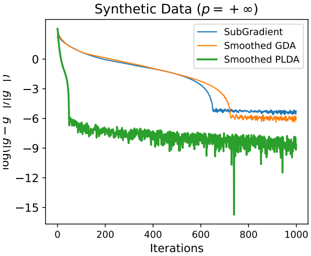
Consider a simple case — the quadratic loss function with linear feature mapping, i.e., $\ell(y, f_\theta(x)) = \frac{1}{2}(y - \theta^\top x)^2$



(a) $p = 1$



(b) $p = 2$

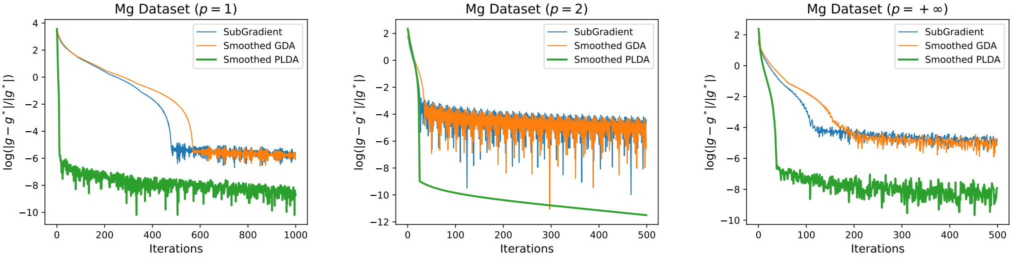


(c) $p = +\infty$

Figure: Compare the convergence behaviours of smoothed PLDA with subgradient and smoothed GDA on both synthetic and real world datasets.

Linear Regression

Consider a simple case — the quadratic loss function with linear feature mapping, i.e., $\ell(y, f_\theta(x)) = \frac{1}{2}(y - \theta^\top x)^2$

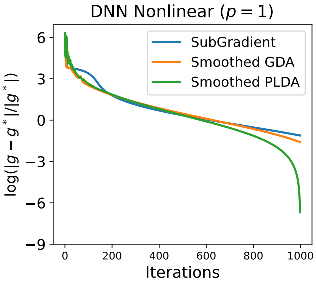


(a) $p = 1$ (b) $p = 2$ (c) $p = +\infty$

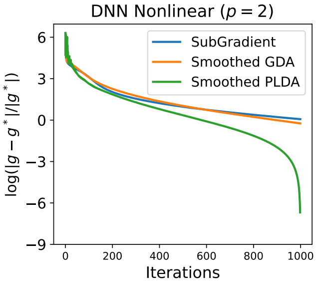
Figure: Compare the convergence behaviours of smoothed PLDA with subgradient and smoothed GDA on both synthetic and real world datasets.

Deep Neural Network

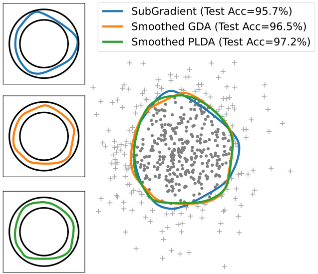
Here, $\ell(\cdot,\cdot)$ is the cross-entropy loss and $f_\theta(\cdot)$ is the feature mapping generated by a neural network with 2 hidden layers of size 5 and use the exponential linear unit (ELU) as the activation function.



(a) $p = 1$



(b) $p = 2$



(c) Decision boundary

Take Home Message

- The proposed smoothed PLDA can achieve the optimal iteration complexity of $\mathcal{O}(\epsilon^{-2})$ when the dual function satisfies the KŁ condition with the exponent $\theta \in [0, \frac{1}{2}]$.
- To the best of our knowledge, this is the first provably efficient algorithm for nonsmooth nonconvex-concave problems, which can achieve the same results as the smooth case.

Take Home Message

- The proposed smoothed PLDA can achieve the optimal iteration complexity of $\mathcal{O}(\epsilon^{-2})$ when the dual function satisfies the KŁ condition with the exponent $\theta \in [0, \frac{1}{2}]$.
- To the best of our knowledge, this is the first provably efficient algorithm for nonsmooth nonconvex-concave problems, which can achieve the same results as the smooth case.

Reference

- Jiajin Li, Linglingzhi Zhu, and Anthony Man-Cho So. Nonsmooth Composite Nonconvex-Concave Minimax Optimization. **Submitted**.

Thank you for listening! Q&A?

Jiajin Li

jiajinli@stanford.edu

https://gerrili1996.github.io/