

L6. NC-PL: convergence of two-timescale GDA

$$\min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m} L(x, y) \quad \kappa := \frac{1}{\mu} \geq 1$$

(A1) L is C^1 on $\mathbb{R}^n \times \mathbb{R}^m$ and L -smooth

(A2) $\forall x$, $L(x, \cdot)$ is μ -PL in y and $y^*(x) \exists!$

$$\text{i.e. } y^*(x) = \arg \max_y L(x, y) \exists!, \quad \Phi(x) := \max_y L(x, y)$$

$$\frac{1}{2} \|\nabla_y L(x, y)\|^2 \geq \mu (\Phi(x) - L(x, y)), \quad \forall y$$

By (PL) \equiv (EB), we also have

$$(EB) \quad \|\nabla_y L(x, y)\| \geq \mu \|y - y^*(x)\|, \quad \forall y, \forall x$$

A-GDA (z, η) two-timescale

$$\begin{cases} x_{k+1} \leftarrow x_k - \tau \nabla_x L(x_k, y_k) \\ y_{k+1} \leftarrow y_k + \eta \nabla_y L(x_{k+1}, y_k) \end{cases}$$

Thm For $\tau \sim \frac{1}{\sqrt{k}^2}$, $\eta \sim \frac{1}{k}$ to be specified in proof. Suppose

$\Phi(x) = \max_y L(x, y)$ has finite lower bound. Then, A-GDA

finds $\|\nabla \Phi(x_k)\| \leq \varepsilon$ in $k \leq T = O(\frac{k^2}{\varepsilon^2})$ steps, and

x_k is an $(\sqrt{2}\varepsilon, \varepsilon/k)$ -stationary pt of L .

Def. We say (\hat{x}, \hat{y}) is an $(\varepsilon, \varepsilon')$ -stationary pt of L if

$$\|\nabla_x L(\hat{x}, \hat{y})\| \leq \varepsilon, \quad \|\nabla_y L(\hat{x}, \hat{y})\| \leq \varepsilon'.$$

To prove the Thm., we first establish Lip-const. of Φ

lemma (Lip of Φ) $\kappa := \frac{1}{\mu}$

- descent of $L(x_k, y_k)$

in the y -step, ℓ -smoothness $\eta \nabla_y L(x_{k+1}, y_k)$

$$L(x_{k+1}, y_{k+1}) - L(x_{k+1}, y_k) \stackrel{\downarrow}{\geq} \langle \nabla_y L(x_{k+1}, y_k), \overbrace{y_{k+1} - y_k}^{\eta \nabla_y L(x_{k+1}, y_k)} \rangle$$

$$- \frac{\ell}{2} \|y_{k+1} - y_k\|^2$$

$$= \eta \|\nabla_y L(x_{k+1}, y_k)\|^2 - \frac{\ell}{2} \eta^2 \|\nabla_y L(x_{k+1}, y_k)\|^2$$

$$\geq \frac{\eta}{2} \|\nabla_y L(x_{k+1}, y_k)\|^2 \dots \textcircled{1} \quad \text{Assume: } \underline{\ell \eta \leq 1}$$

in the x -step,

$$L(x_{k+1}, y_k) - L(x_k, y_k) \geq \langle \nabla_x L(x_k, y_k), \overbrace{x_{k+1} - x_k}^{-\tau \nabla_x L(x_k, y_k)} \rangle$$

$$- \frac{\ell}{2} \|x_{k+1} - x_k\|^2$$

$$= -\tau \|\nabla_x L(x_k, y_k)\|^2 - \frac{\ell}{2} \tau^2 \|\nabla_x L(x_k, y_k)\|^2$$

already have $\ell \tau \leq \frac{1}{2}$

$$\geq -\frac{5}{4} \tau \|\nabla_x L(x_k, y_k)\|^2 \dots \textcircled{2}$$

$\textcircled{1}, \textcircled{2} \Rightarrow$

$$L(x_{k+1}, y_{k+1}) - L(x_k, y_k) \geq \frac{\eta}{2} \|\nabla_y L(x_{k+1}, y_k)\|^2 - \frac{5}{4} \tau \|\nabla_x L(x_k, y_k)\|^2$$

- Design the Lyapunov function

$\alpha > 0$ TBD

$$V_k = V(x_k, y_k) := \Phi(x_k) + \alpha (\Phi(x_k) - L(x_k, y_k))$$

$$V_k - V_{k+1} = (1 + \alpha) (\Phi(x_k) - \Phi(x_{k+1})) - \alpha (L(x_k, y_k) - L(x_{k+1}, y_{k+1}))$$

$$\geq (1 + \alpha) \left(\frac{\tau}{2} \|\nabla \Phi(x_k)\|^2 - \frac{\tau}{2} \|\delta_k\|^2 \right)$$

$$+ \alpha \left(\frac{\eta}{2} \|\nabla_y L(x_{k+1}, y_k)\|^2 - \frac{5}{4} \tau \underbrace{\|\nabla_x L(x_k, y_k)\|^2}_{\|\nabla \Phi(x_k) + \delta_k\|^2} \right)$$

The two negative terms are trouble

a) $\nabla_x L(x_k, y_k) \approx \nabla \Phi(x_k)$ up to δ_k , so handled by $\|\nabla \Phi(x_k)\|$;

b) $\|\delta_k\| \leq \kappa \|\nabla_y L(x_k, y_k)\|$ by $\textcircled{1}$, not quite $\nabla_y L(x_{k+1}, y_k)$

but maybe close to.

$$\underbrace{\| \nabla_y L(x_{k+1}, y_k) - \nabla_y L(x_k, y_k) \|}_{=: e_k} \leq l \|x_{k+1} - x_k\|$$

$$= l \tau \| \nabla_x L(x_k, y_k) \| \quad \text{--- (5)}$$

again controlled by a).

For a): $\| \nabla_x L(x_k, y_k) \|^2 = \| \nabla \Phi(x_k) + \delta_k \|^2 \leq 2 (\| \nabla \Phi(x_k) \|^2 + \| \delta_k \|^2)$ --- (3)

For b): $\| \nabla_y L(x_{k+1}, y_k) \|^2 = \| \nabla_y L(x_k, y_k) + e_k \|^2$

$$\geq \frac{1}{2} \| \nabla_y L(x_k, y_k) \|^2 - \| e_k \|^2$$

$$(\| a+b \|^2 \geq \frac{1}{2} \| a \|^2 - \| b \|^2)$$

$$\Rightarrow V_k - V_{k+1} \geq \tau \frac{1+\alpha}{2} \| \nabla \Phi(x_k) \|^2 - \tau \frac{1+\alpha}{2} \| \delta_k \|^2 - \tau \frac{5\alpha}{4} \| \nabla_x L(x_k, y_k) \|^2$$

$$+ \eta \frac{\alpha}{4} \| \nabla_y L(x_k, y_k) \|^2 - \eta \frac{\alpha}{2} \| e_k \|^2$$

by (5) $\rightarrow \leq l^2 \tau^2 \| \nabla_x L(x_k, y_k) \|^2$

$$\eta l^2 \tau^2 \leq \frac{\tau}{2} \quad (\eta l \leq 1, l \tau \leq \frac{1}{2})$$

$$\geq \tau \frac{1+\alpha}{2} \| \nabla \Phi(x_k) \|^2 - \tau \frac{1+\alpha}{2} \| \delta_k \|^2 - \tau \frac{3\alpha}{2} \| \nabla_x L(x_k, y_k) \|^2$$

$$+ \eta \frac{\alpha}{4} \| \nabla_y L(x_k, y_k) \|^2$$

by (3)

$$\geq \tau \left(\frac{1+\alpha}{2} - 3\alpha \right) \| \nabla \Phi(x_k) \|^2 - \tau \left(\frac{1+\alpha}{2} + 3\alpha \right) \| \delta_k \|^2$$

$$+ \eta \frac{\alpha}{4} \| \nabla_y L(x_k, y_k) \|^2$$

$$= \tau \frac{1-5\alpha}{2} \| \nabla \Phi(x_k) \|^2 - \tau \frac{1+7\alpha}{2} \| \delta_k \|^2 + \eta \frac{\alpha}{4} \| \nabla_y L(x_k, y_k) \|^2$$

by (4) $\rightarrow \geq \left(\eta \frac{\alpha}{4} - \tau \frac{1+7\alpha}{2} k^2 \right) \| \nabla_y L(x_k, y_k) \|^2$

if $\eta \frac{\alpha}{4} - \tau \frac{1+7\alpha}{2} k^2 > 0$, then ≥ 0

let $\alpha = \frac{1}{8}$, $\eta \frac{1}{32} \geq \tau k^2$ Assume.

$$\geq \frac{\tau}{8} \| \nabla \Phi(x_k) \|^2 + \left(\frac{1}{32} - \tau k^2 \right) \| \nabla_y L(x_k, y_k) \|^2 \quad (6)$$

We have shown $\frac{\tau}{8} \| \nabla \Phi(x_k) \|^2 \leq V_k - V_{k+1}$

... $\tau \frac{1}{32} \| \nabla_y L(x_k, y_k) \|^2 \leq V_k - V_{k+1}$

By telescoping, $\sum_{k=0}^T \|\nabla \Phi(x_k)\| \leq V_0 - V_T$

$$V_T = \Phi(x_T) + \alpha \left(\overbrace{\Phi(x_T) - L(x_T, y_0)}^{\geq 0} \right) \geq \Phi(x_T) \geq \Phi^*$$

$$\Phi^* = \min_z \Phi(z) \text{ assumed to be finite}$$

$$\Rightarrow V_0 - V_T \leq V_0 - \Phi^* = \Phi(x_0) - \Phi^* + \frac{1}{8} (\Phi(x_0) - L(x_0, y_0))$$

$$=: \Delta \text{ const. depending on initial value } (x_0, y_0)$$

$$\Rightarrow \sum_{k=0}^T \|\nabla \Phi(x_k)\|^2 \leq 8 \frac{\Delta}{\epsilon}$$

$$\Rightarrow \exists k \leq T \text{ s.t. } \|\nabla \Phi(x_k)\|^2 \leq 8 \frac{\Delta}{\epsilon T}$$

Overall, we have assumed

$$\begin{cases} \eta L \leq 1 \\ \tau L(HK) \leq 1 \\ \tau K^2 \leq \frac{\eta}{32} \end{cases}$$

This can be fulfilled by setting $\eta = \frac{1}{L}$, $\tau = \frac{1}{32} \frac{1}{\epsilon K^2}$

Then,

$$\|\nabla \Phi(x_k)\|^2 \leq \frac{\Delta}{\epsilon T} \sim \frac{LK^2\Delta}{T}$$

One can use the refined estimate (*) to show that

$$\|\nabla_x L(x_k, y_k)\| \leq \sqrt{2}\epsilon, \quad \|\nabla_y L(x_k, y_k)\| \leq \epsilon/K.$$

$$V_k - V_{k+1} \geq \frac{\epsilon}{8} \|\nabla \Phi(x_k)\|^2 + \underbrace{\left(\frac{\eta}{32} - \tau K^2 \right)}_{= \frac{\epsilon}{4} K^2 \text{ if } \frac{\eta}{32} = \frac{\epsilon}{4} \tau K^2} \|\nabla_y L(x_k, y_k)\|^2$$

$$= \frac{\epsilon}{8} (\|\nabla \Phi(x_k)\|^2 + 2K^2 \|\nabla_y L(x_k, y_k)\|^2)$$

By summing $k=0, \dots, K-1$, $\exists k \leq T \sim \frac{LK^2\Delta}{T}$ s.t.

$$\|\nabla \Phi(x_k)\|^2 + 2K^2 \|\nabla_y L(x_k, y_k)\|^2 \leq \epsilon^2 \quad (*)$$

This x_k satisfies that $\|\nabla \Phi(x_k)\| \leq \epsilon$.

We now derive an upper bound of $\|\nabla_x L(x_k, y_k)\| : \forall x, y$,

$$\|\nabla_x L(x, y) - \nabla_x L(x, y^*(x))\| \leq L \|y - y^*(x)\| \leq K \|\nabla_y L(x, y)\|$$

then

$$\|\nabla \Phi(x)\|$$

$$\|\nabla_x L(x_k, y_k)\| \leq \underbrace{\|\nabla_x L(x_k, y_k) - \nabla \Phi(x_k)\|}_{\leq \kappa \|\nabla_y L(x_k, y_k)\|} + \|\nabla \Phi(x_k)\|$$

$$\Rightarrow \|\nabla_x L(x_k, y_k)\|^2 \leq (\|\nabla \Phi(x_k)\| + \kappa \|\nabla_y L(x_k, y_k)\|)^2 \\ \leq 2(\|\nabla \Phi(x_k)\|^2 + \kappa^2 \|\nabla_y L(x_k, y_k)\|^2)$$

Back to (*), we have

$$\Sigma^2 \geq \|\nabla \Phi(x_k)\|^2 + 2\kappa^2 \|\nabla_y L(x_k, y_k)\|^2 \\ \geq \kappa^2 \|\nabla_y L(x_k, y_k)\|^2 + \frac{1}{2} \|\nabla_x L(x_k, y_k)\|^2$$

Then, $\kappa \|\nabla_y L(x_k, y_k)\|, \frac{1}{\sqrt{2}} \|\nabla_x L(x_k, y_k)\| \leq \Sigma$. \square

KK. The proof for GDA is similar.

• proximal step S-GDA

$$\begin{cases} x_{k+1} \leftarrow x_k - \tau \left(\nabla_x L(x_k, y_k) + \underbrace{\beta(x_k - w_k)}_{\nabla_x K(x_k, y_k, w_k)} \right) \\ y_{k+1} \leftarrow y_k + \eta \nabla_y L(x_{k+1}, y_k) \quad \nabla_y K(x_{k+1}, y_k, w_k) \\ w_{k+1} \leftarrow w_k + \underbrace{\rho(x_{k+1} - w_k)}_{\nabla_w K(x_{k+1}, y_k, w_k)} \end{cases}$$

$$\min_x \Phi(x) + \frac{\rho}{2} \|x - w\|^2 = \min_x \max_y \underbrace{L(x, y) + \frac{\rho}{2} \|x - w\|^2}_{K(x, y, w)}$$

$$\text{set } \rho = 2\ell, \beta \sim \eta\mu \Rightarrow \frac{\beta}{\rho} \sim \frac{\eta\mu}{\ell} = \frac{\eta}{\kappa}$$

Thm. For $\tau \sim \frac{1}{\ell}$, $\eta \sim \frac{1}{\ell}$, $\rho = 2\ell$, $\beta \sim \eta\mu$, S-GDA finds

an $(\varepsilon, \frac{\Sigma}{\sqrt{\kappa}})$ -stationary pt of L in $O(\frac{\ell\kappa}{\varepsilon^2})$ steps.

ref. Thm 4.1 [YOLH 2022]

KK: Comparing to GD, the step size is less restricted by condition number κ , and etc. complexity is also better, i.e. $\sim \kappa$ instead of κ^2 .

Generally, proximal step improves "stability" and allows for larger step size in GD.

$$\min_x f(x) \quad f_s(x) = \min_u f(u) + \frac{1}{2s} \|u - x\|^2$$

$$\text{GD: } x_{k+1} = x_k - s \nabla f(x_k) \quad (\text{fwd Euler})$$

$$\begin{aligned} \text{PPM: } x_{k+1} &= \text{Prox}_{s,f}(x_k) \quad (\text{bwd Euler}) \\ &= x_k - s \nabla f_s(x_k) \quad \text{as if GD on } f_s \end{aligned}$$

Suppose f is L -smooth, l -W.C., $\kappa = \frac{L}{l} \geq 1$

- GD needs $s < \frac{2}{L}$, typically $s = \frac{1}{L}$, then \underline{L} shows

$$\frac{1}{T} \sum_{k=0}^{T-1} \|\nabla f(x_k)\|^2 \lesssim \frac{\Delta}{T}, \quad \Delta = f(x_0) - f^*$$

- f_s need $s < \frac{1}{L}$, typically $s = \frac{1}{2L}$,

then f_s is \bar{L} -smooth, $\bar{L} = \max\{\frac{1}{s}, \frac{1}{\frac{1}{L} - s}\} = 2L$.

for GD on f_s , need $s < \frac{2}{\bar{L}} = \frac{1}{L}$ satisfied.

$$\Rightarrow \frac{1}{T} \sum_{k=0}^{T-1} \underbrace{\|\nabla f_s(x_k)\|^2}_\downarrow \leq \frac{\bar{\Delta}}{T}, \quad \bar{\Delta} = f_s(x_0) - f_s^* \leq \Delta$$

$$\nabla f((x_k)_+) = \nabla f(x_{k+1})$$

Thus, in GD, $s \sim \frac{1}{L}$, $T \sim \frac{L}{\epsilon^2}$,

in PPM, $s \sim \frac{1}{L}$, $T \sim \frac{L}{\epsilon^2}$,

both are $O(\kappa)$ better.