Composite ERM

$$\min_{w \in \mathbb{R}^p} F(w) = \underbrace{\frac{1}{n} \sum_{i=1}^{n} \ell(w; d_i)}_{f(w)} + \underbrace{\sum_{j=1}^{p} \psi_j(w_j)}_{\psi(w)}$$

ℓ is convex, smooth, $\quad$ ψ is convex

Composite ERM

$$\min_{w \in \mathbb{R}^p} F(w) = \underbrace{\frac{1}{n} \sum_{i=1}^{n} \ell(w; d_i)}_{f(w)} + \underbrace{\sum_{j=1}^{p} \psi_j(w_j)}_{\psi(w)}$$

ℓ is convex, smooth, ψ is convex

Goal: (ϵ, δ)-DP solver?

The Classical: DP-SGD

Choose $i \sim_u \{1, \ldots, n\}$

Update

$$w^{t+1} = \text{prox}_{\psi}\left(w^t - \eta\left(\nabla \ell(w^t; d_i) + \boldsymbol{\mathcal{N}(\sigma^2 \mathbb{1}_p)}\right)\right)$$

The Classical: DP-SGD

Choose $i \sim_u \{1, \ldots, n\}$

Update

$$w^{t+1} = \text{prox}_\psi \left(w^t - \eta \left(\nabla \ell(w^t; d_i) + \boldsymbol{\mathcal{N}(\sigma^2 \mathbb{1}_p)} \right) \right)$$

With $\boldsymbol{\sigma^2} \propto \frac{T}{n^2 \epsilon^2} \boldsymbol{\Lambda^2}$ assuming $\boldsymbol{|\nabla \ell(w; \cdot)| \leq \Lambda}$

The Classical: DP-SGD

C

U

w))

Two drawbacks:

- Relies on privacy amplification
- Oblivious to imbalanced gradients

With $\boldsymbol{\sigma^2} \propto \dfrac{T}{n^2\epsilon^2}\boldsymbol{\Lambda^2}$ assuming $|\boldsymbol{\nabla \ell(w;\cdot)}| \leq \boldsymbol{\Lambda}$

Our method: DP-CD

Choose $j \sim_u \{1, \ldots, p\}$

Update

$$w_j^{t+1} = \mathrm{prox}_{\eta_j \psi_j}\left(w_j^t - \eta_j\left(\nabla_j f(w^t) + \boldsymbol{\mathcal{N}(\sigma_j^2)}\right)\right)$$

Our method: DP-CD

Choose $j \sim_u \{1, \ldots, p\}$

Update

$$w_j^{t+1} = \text{prox}_{\eta_j \psi_j}\left(w_j^t - \eta_j\left(\nabla_j f(w^t) + \boldsymbol{\mathcal{N}(\sigma_j^2)}\right)\right)$$

With $\boldsymbol{\sigma_j^2} \propto \frac{T}{n^2\epsilon^2}\boldsymbol{L_j^2}$ assuming $\boldsymbol{|\nabla_j f(w)| \leq L_j}$

Utility: $\mathbb{E}[F(w) - F^*] \leq$

$$\widetilde{O}\left(\frac{\sqrt{p}}{n\epsilon}\|\boldsymbol{L}\|_{M^{-1}}\left\|\boldsymbol{w}^0 - \boldsymbol{w}^*\right\|_{M}\right)$$

Utility: $\mathbb{E}[F(w) - F^*] \leq$

$$\widetilde{O}\left(\frac{\sqrt{p}}{n\epsilon}\|\boldsymbol{L}\|_{M^{-1}}\left\|\boldsymbol{w^0} - \boldsymbol{w^*}\right\|_{\boldsymbol{M}}\right)$$

Where F is convex and

- $\|\boldsymbol{L}\|_{M^{-1}} = \sqrt{\sum_{j=1}^{\boldsymbol{p}} \boldsymbol{L_j^2}/\boldsymbol{M_j}}$
- $\left\|\boldsymbol{w^0} - \boldsymbol{w^*}\right\|_{M^{-1}} = \sqrt{\sum_{j=1}^{\boldsymbol{p}} \boldsymbol{M_j}(\boldsymbol{w_j^0} - \boldsymbol{w_j^*})^2}$

$$\left(\begin{array}{lll} \text{And } f \text{ is} & \textit{Lipschitz} & |f(w) - f(w + te_j)| \leq L_j\,|t| \\ & \textit{Smooth} & |\nabla f(w) - \nabla f(w + te_j)| \leq M_j\,|t| \end{array}\right)$$

Utility: $\mathbb{E}[F(w) - F^*] \leq$

$$\widetilde{\mathcal{O}}\left(\sqrt{p}\cdots\right)$$

We also prove:

- Results for strongly-convex F
- Corresponding lower bounds

Whe

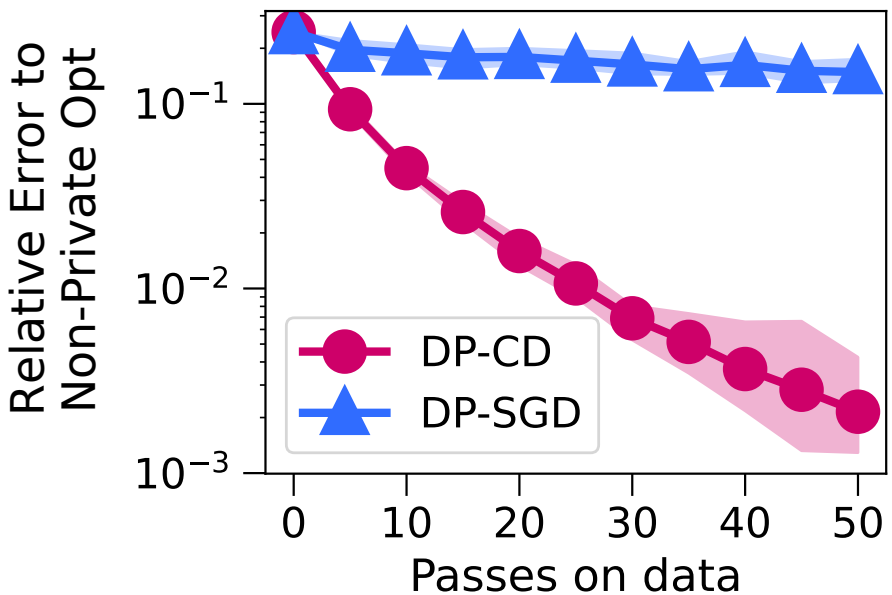
- $\|$
- $\|$

$$\|M - \sqrt{\textstyle\sum_{j=1}} \cdots$$

$$\left(\begin{array}{lll} \text{And } f \text{ is} & \textit{Lipschitz} & |f(w) - f(w + te_j)| \leq L_j |t| \\ & \textit{Smooth} & |\nabla f(w) - \nabla f(w + te_j)| \leq M_j |t| \end{array}\right)$$

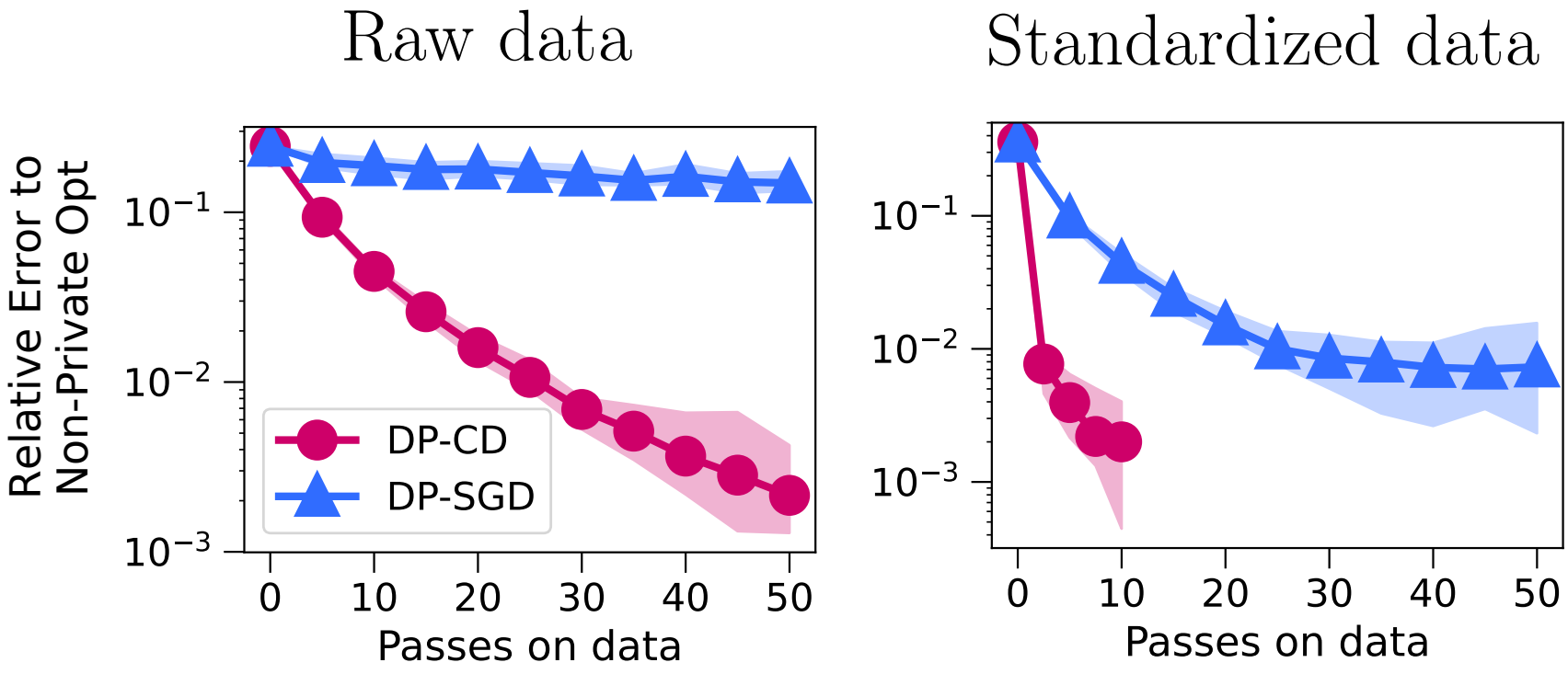
Logistic Regression, ELECTRICITY dataset
$(n = 45k, p = 8, \epsilon = 1, \delta = \frac{1}{n^2})$

Raw data



Logistic Regression, ELECTRICITY dataset

$(n = 45k, p = 8, \epsilon = 1, \delta = \frac{1}{n^2})$



Come to our poster to discuss:

- non-asymptotic results

Come to our poster to discuss:

- non-asymptotic results
- lower bounds

Come to our poster to discuss:

- non-asymptotic results
- lower bounds
- practical implementation
  - clipping thresholds
  - estimation of the constants

Come to our poster to discuss:

- non-asymptotic results
- lower bounds
- practical implementation
  - clipping thresholds
  - estimation of the constants
- more experiments