High-Dimensional Private Empirical Risk Minimization by Greedy Coordinate Descent

Paul Mangold[1], Aurélien Bellet[1], Joseph Salmon[2], Marc Tommasi[1]
[1]Inria Lille, France [2]Univ. Montpellier, France

Problem Statement

DP-ERM [1]: find (ϵ, δ)-DP approximation of

$$w^* \in \arg\min_{w \in \mathbb{R}^p} \left\{ f(w) := \frac{1}{n} \sum_{i=1}^{n} \ell(w; d_i) \right\}$$

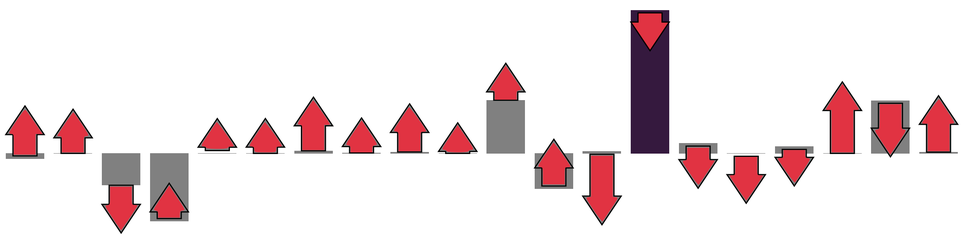
Where $\ell(\cdot, d_i)$ is, for all $w, v \in \mathbb{R}^p$,

- Convex: $\ell(w;\cdot) \geq \ell(v;\cdot) + \langle \nabla \ell(v;\cdot), w - v \rangle$
- Lipschitz: $|\nabla_j \ell(w;\cdot)| \leq L_j$
- Smooth: $|\nabla_j \ell(w;\cdot) - \nabla_j \ell(w + te_j;\cdot)| \leq M_j |t|$

Notation: $L_{\min} = \min L_j$, $L_{\max} = \max L_j$ (same for $M_{\min}$, $M_{\max}$)

Private Greedy Coordinate Selection

Report-noisy-max [2]: add noise and select largest entry



$\Rightarrow$ and noise scale does not depend on the dimension!

Reminder: General Utility Lower Bounds [3]

- Convex loss

$$f(w^T) - f(w^*) = O\left(\frac{\sqrt{p}}{n\epsilon} \|L\|_2 \|w^*\| \right)$$

- Strongly-Convex loss (*i.e.*, when $f - \frac{\mu}{2}\|\cdot\|_2^2$ is convex)

$$f(w^T) - f(w^*) = O\left(\frac{p}{\mu n^2 \epsilon^2} \|L\|_2^2 \right)$$

Problem: dependence on dimension p is high

Idea: exploit structure to outperform lower bounds!

Algorithm: Private Greedy Coordinate Descent

Initialization $w^0 = 0$, $T > 0$, $\gamma_j = \frac{1}{M_j}$

for $t = 0$ to $T - 1$ **do**

$$j = \arg\max_{j' \in [p]} \left\{ \frac{1}{\sqrt{M_{j'}}} \left| \nabla_{j'} f(w^t) + \chi_{j'}^t \right| \right\} \quad \chi_{j'}^t \sim \text{Lap}\left(\frac{8 L_j \sqrt{T \log(1/\delta)}}{n\epsilon}\right)$$

$$w_j^{t+1} = w_j^t - \gamma_j (\nabla_j f(w^t) + \eta_j^t) \quad \eta_j^t \sim \text{Lap}\left(\frac{8 L_j \sqrt{T \log(1/\delta)}}{n\epsilon}\right)$$

return w^T.

Private greedy coordinate descent naturally exploits problem structure to achieve near dimension independent utility in DP-ERM.

General Utility Results

Scaled norm $\|w - v\|_{M,q} = \left(\sum_{j=1}^{p} M_j^{\frac{q}{2}} |w_j - v_j|^q \right)^{\frac{1}{q}}$, for $q \in \{1, 2\}$

With probability $1 - \zeta$:

- Convex loss

$$f(w^T) - f(w^*) = O\left(\frac{\log(p/\zeta)}{n^{2/3}\epsilon^{2/3}} \cdot \frac{R_{M,1}^{4/3} L_{\max}^{2/3}}{M_{\min}^{1/3}} \right)$$

where $R_{M,1} = \max_{w \in \mathbb{R}^p} \min_{w^* \in \mathcal{W}^*} \left\{ \|w - w^*\|_{M,1} \mid f(w) \leq f(w^0) \right\}$

- Strongly-Convex loss (*i.e.*, when $f - \frac{\mu_{M,q}}{2}\|\cdot\|_{M,q}^2$ is convex)

$$f(w^T) - f(w^*) = O\left(\frac{\log(p/\zeta)}{n^2\epsilon^2} \cdot \frac{L_{\max}^2}{M_{\min} \mu_{M,1}^2} \right)$$

Better Utility in (Quasi) Sparse Problems

(α, τ)-quasi-sparse solution: at most τ values greater than α

With α small enough, assuming iterates remain $(0, s)$-sparse:

$$f(w^T) - f(w^*) = O\left(\frac{(s+\tau)^2 \log(2p/\zeta)}{n^2\epsilon^2} \frac{L_{\max}^2}{M_{\min} \mu_{M,2}} \right)$$

DP-GCD is fast in first iterations [4] $\Rightarrow$ iterates remain sparse

Proximal Variant for Composite Problems

Initialization $w^0 = 0$, $T > 0$, $\gamma_j = \frac{1}{M_j}$

for $t = 0$ to $T - 1$ **do**

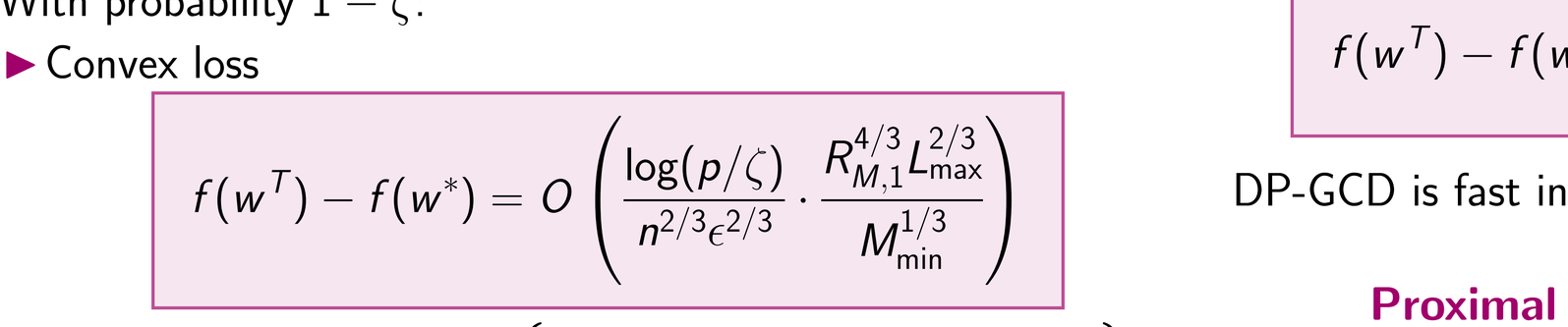
$$j = \arg\max_{j \in [p]} \left\{ \min_{\xi_j \in \partial \psi_j(w_j)} \left\{ \frac{1}{\sqrt{M_j}} \left| \nabla_j f(w^t) + \chi_j^t + \xi_j \right| \right\} \right\} \quad \chi_{j'}^t \sim \text{Lap}\left(\frac{8 L_j \sqrt{T \log(1/\delta)}}{n\epsilon}\right)$$

$$w_j^{t+1} = \text{prox}_{\gamma_j \psi_j}(w^t - \gamma_j (\nabla_j f(w^t) + \eta_j^t)) \quad \eta_{j'}^t \sim \text{Lap}\left(\frac{8 L_j \sqrt{T \log(1/\delta)}}{n\epsilon}\right)$$

return w^T.

Works in practice, but analysis is an open problem!

Experiments (with $\epsilon = 1, \delta = 1/n^2$)



Logistic + L2 $n = 1000, p = 100$, $w^* \sim$ lognormal(1)

Logistic + L2 $n = 1000, p = 100$, $w^* \sim$ lognormal(2)

Logistic + L1 $n = 800, p = 88,119$ `dorothea`

Logistic + L1 $n = 2600, p = 501$ `madelon`

References

[1] K. Chaudhuri, C. Monteleoni, and A. D. Sarwate. "Differentially Private Empirical Risk Minimization". In: *JMLR* (2011).

[2] C. Dwork and A. Roth. "The Algorithmic Foundations of Differential Privacy". In: *Foundations and Trends® in Theoretical Computer Science* (2013).

[3] R. Bassily, A. Smith, and A. Thakurta. "Private Empirical Risk Minimization: Efficient Algorithms and Tight Error Bounds". In: *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*. 2014.

[4] H. Fang et al. "Greed Meets Sparsity: Understanding and Improving Greedy Coordinate Descent for Sparse Optimization". In: *AISTATS*. 2020.