Differentially Private Coordinate Descent for Composite Empirical Risk Minimization

Paul Mangold[1], Aurélien Bellet[1], Joseph Salmon[2] and Marc Tommasi[1]

[1] Inria Lille, [2] Univ. Montpellier

Problem Statement

DP-ERM (Chaudhuri et al., 2011):

$$w^* \in \arg\min_{w\in\mathbb{R}^p} F(w) = \frac{1}{n}\sum_{i=1}^{n} \ell(w; d_i) + \psi(w)$$

Under (ϵ, δ)-differential-privacy

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: $|\nabla_j \ell(w;\cdot) - \nabla_j \ell(w + te_j;\cdot)| \leq M_j |t|$

and $\psi(w) = \sum_{j=1}^p \psi_j(w_j)$ is convex, separable

DP-CD Algorithm

Init. w^0; $\sigma_j > 0$; $T, K > 0$; $\gamma_j > 0$

for $t = 0, \dots, T-1$ **do**

  $\theta^0 = \bar{w}^t$

  for $k = 0, \dots, K-1$ **do**

    $j \sim \mathcal{U}(\{1, \dots, p\})$

    $\theta_j^{k+1} = \mathrm{prox}_{\gamma_j \psi_j}(\theta_j^k - \gamma_j(\nabla_j f(\theta^k) + \eta_j))$ with $\eta_j \sim \mathcal{N}(0, \sigma_j^2)$

  $\bar{w}^{t+1} = \frac{1}{K}\sum_{k=1}^K \theta^k$

return $w^{priv} = \bar{w}^T$

Privacy

- Gaussian Mechanism (Dwork and Roth, 2013) with $\sigma_j^2 = \frac{8TL_j^2 \log(1/\delta)}{n^2\epsilon^2}$ + RDP Composition

Utility Results, with norm $\|X\|_{M^{-1}}^2 = \sum_{j=1}^p M_j^{-1} X_j^2$

- Convex with $R_M = \max(\|w^0 - w^*\|_{M^{-1}}, F(w^0) - F^*)$, set $T = 1, K > 0$

$$\mathbb{E}\big[F(w^{priv}) - F^*\big] \leq \frac{3pR_M^2}{2K} + \frac{12K\log(1/\delta)\,\|L\|_{M^{-1}}^2}{n^2\epsilon^2} = \widetilde{O}\left(\frac{\sqrt{p} R_M \|L\|_{M^{-1}}}{n\epsilon}\right)$$

- Strongly-Convex (i.e., $F(w) \geq F(v) + \langle \nabla F(v), w - v\rangle + \frac{\mu_M}{2}\|w - v\|_M^2$), set $T > 0, K = \widetilde{O}(p/\mu_M)$

$$\mathbb{E}\big[F(w^{priv}) - F^*\big] \leq \frac{F(w^0) - F(w^*)}{2^T} + \frac{48pT\log(1/\delta)\,\|L\|_{M^{-1}}^2}{\mu_M n^2\epsilon^2} = \widetilde{O}\left(\frac{p\,\|L\|_{M^{-1}}^2}{\mu_M n^2\epsilon^2}\right)$$

Lower Bounds, assuming $\sum_{j\in\mathcal{S}} L_j^2 = \Omega(\|L\|^2)$ for any $|\mathcal{S}| \geq \lfloor \frac{p}{75} \rfloor$

Convex: $\mathbb{E}\big[F(w^{priv}) - F^*\big] = \widetilde{\Omega}\left(\frac{\sqrt{p}\|w^*\|_2\|L\|_2}{n\epsilon}\right)$

Strongly-Convex: $\mathbb{E}\big[F(w^{priv}) - F^*\big] = \widetilde{\Omega}\left(\frac{p\|L\|_2^2}{\mu_I n^2\epsilon^2}\right)$

Take home!

Private Coordinate Descent Algorithm:

- Adapts to coordinate-wise regularity
- Works with privately estimated constants
- No amplification required
- Nearly matched lower bounds

Coordinate-wise Clipping

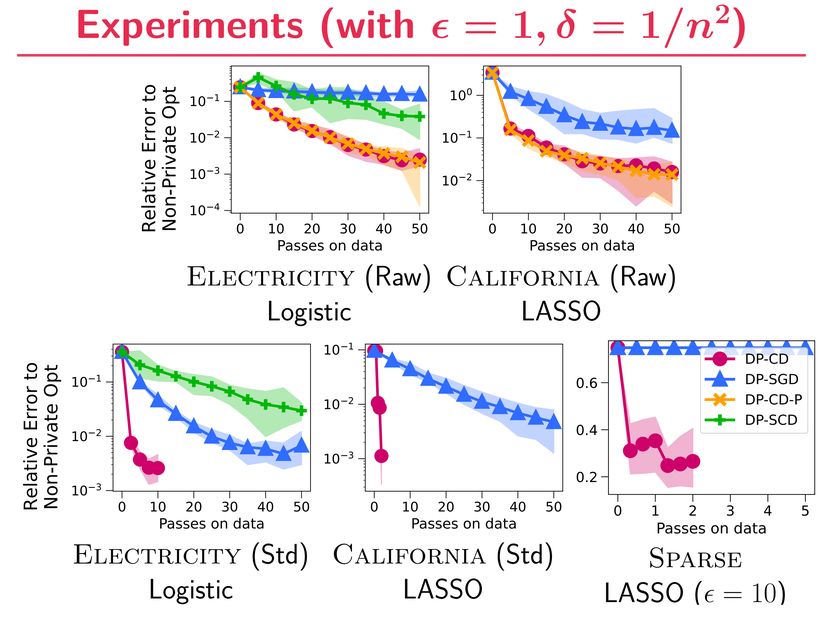
$$\nabla_j f(w)^{clipped} = \frac{1}{n}\sum_{i=1}^n \mathrm{clip}\left(\nabla_j \ell(w; d_i), C_j\right)$$

where for $X \in \mathbb{R}$, $\mathrm{clip}(X, C_j) = \mathrm{sign}(X)\min(|X|, C_j)$

Guarantees $\left|\nabla_j f(w)^{clipped}\right| \leq C_j$

$\rightarrow$ choose $C_j = \sqrt{\frac{M_j}{\mathrm{tr}(M)}} C$ so that $\|C_j\|_{M^{-1}} = C$

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



Estimating Constants Privately

$$M_j = \frac{1}{n}\sum_{i=1}^n M_j(\ell(\cdot; d_i)) \text{ thus } M_j^{priv} = M_j + \mathrm{Lap}\left(\frac{B_j}{n\epsilon}\right)$$

For linear models, $M_j(\cdot; d_i) = X_{i,j}^2 \leq B_j$ for some $B_j > 0$

$\rightarrow$ clip to enforce a rough upper bound B_j (known *a priori*)