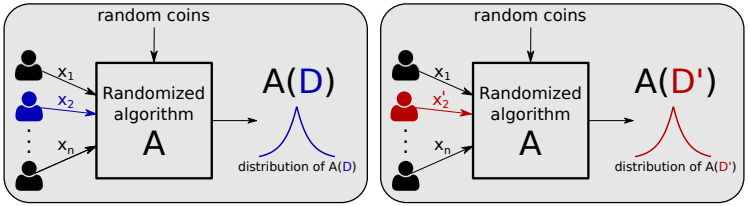
From Noisy Fixed Point Iterations to Private ADMM for Centralized and Federated Learning

ICML 2023

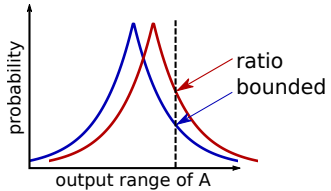
Edwige Cyffers, Aurelien Bellet, Debabrota Basu

Univ. Lille, Inria, CNRS, Centrale Lille

Differential Privacy



- Neighboring datasets $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$ and $\mathcal{D}' = \{x_1, x'_2, x_3, \ldots, x_n\}$
- **Requirement**: $\mathcal{A}(\mathcal{D})$ and $\mathcal{A}(\mathcal{D}')$ should have similar distributions
- Here, we use Renyi DP, which requires $D_\alpha\left(\mathcal{A}(\mathcal{D})||\mathcal{A}(\mathcal{D}')\right) \leq \varepsilon$



Noise Injection and DP-SGD

- Given $\mathcal{D} = (d_1, \ldots, d_n)$, find $u \in \mathbb{R}^p$ minimizing $f(u; \mathcal{D}) = \frac{1}{n} \sum_{i=1}^{n} f(u; d_i)$ under DP

Noise Injection and DP-SGD

- Given $\mathcal{D} = (d_1, \ldots, d_n)$, find $u \in \mathbb{R}^p$ minimizing $f(u; \mathcal{D}) = \frac{1}{n} \sum_{i=1}^{n} f(u; d_i)$ under DP

Algorithm 2 Differentially Private SGD (DP-SGD) **Bassily2014a**; **Abadi2016**

Initialize $u_0 \in \mathbb{R}^p$ (independent of $\mathcal{D}$)
for $t = 0, \ldots, T-1$ **do**
  Pick $i_t \in \{1, \ldots, n\}$ uniformly at random
  $u_{t+1} \leftarrow u_t - \gamma^{(t)} \left(\nabla f(u_t; d_{i_t}) + \eta_{t+1} \right)$ where $\eta_{t+1} \sim \mathcal{N}(0, \sigma^2 \Delta^2 \mathbb{I}_p)$
Return u_T

- Utility analysis: same as non-private SGD (with additional noise due to privacy)

Noise Injection and DP-SGD

- Given $\mathcal{D} = (d_1, \ldots, d_n)$, find $u \in \mathbb{R}^p$ minimizing $f(u; \mathcal{D}) = \frac{1}{n} \sum_{i=1}^{n} f(u; d_i)$ under DP

Algorithm 3 Differentially Private SGD (DP-SGD) **Bassily2014a**; **Abadi2016**

Initialize $u_0 \in \mathbb{R}^p$ (independent of $\mathcal{D}$)
for $t = 0, \ldots, T-1$ **do**
  Pick $i_t \in \{1, \ldots, n\}$ uniformly at random
  $u_{t+1} \leftarrow u_t - \gamma^{(t)} \left(\nabla f(u_t; d_{i_t}) + \eta_{t+1} \right)$ where $\eta_{t+1} \sim \mathcal{N}(0, \sigma^2 \Delta^2 \mathbb{I}_p)$
Return u_T

- Utility analysis: same as non-private SGD (with additional noise due to privacy)
- Privacy analysis: DP-SGD is $(\alpha, \frac{\alpha T}{2n^2\sigma^2})$ by subsampled Gaussian mechanism and composition property of RDP over T iterations

ADMM

- Alternating Direction Method of Multipliers (ADMM) aims to solve:

$$\begin{aligned} \underset{x,z}{\text{minimize}} \quad & f(x;\mathcal{D}) + g(z) \\ \text{subject to} \quad & Ax + Bz = c \end{aligned}$$

Algorithm 4 ADMM algorithm

Input: initial point u_0, step size $\lambda \in (0, 1]$, Lagrange parameter $\gamma > 0$
for $k = 0$ to $K - 1$ **do**
$z_{k+1} = \text{argmin}_z \left\{ g(z) + \frac{1}{2\gamma} \|Bz + u_k\|^2 \right\}$
$x_{k+1} = \text{argmin}_x \left\{ f(x;\mathcal{D}) + \frac{1}{2\gamma} \|Ax + 2Bz_{k+1} + u_k - c\|^2 \right\}$
$u_{k+1} = u_k + 2\lambda (Ax_{k+1} + Bz_{k+1} - c)$
Return z^K

- How can we make ADMM private and analyze its utility?

Noisy Fix-point Operators

- We study the general noisy fixed-point iteration

Algorithm 5 Noisy fixed-point iteration

Input: non-expansive operator $R = (R_1, \dots, R_B)$ over $1 \leq B \leq p$ blocks,step sizes $(\lambda_k)_{k\in\mathbb{N}} \in (0, 1]$, active blocks $(\rho_k)_{k\in\mathbb{N}} \in \{0, 1\}^B$, errors $(e_k)_{k\in\mathbb{N}}$, noise variance $\sigma^2 \geq 0$

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

  for $b = 1, \dots, B$ **do**

    $u_{k+1,b} = u_{k,b} + \rho_{k,b}\lambda_k(R_b(u_k) + e_{k,b} + \eta_{k+1,b} - u_{k,b})$ with $\eta_{k+1,b} \sim \mathcal{N}(0, \sigma^2\mathbb{I}_p)$

- This general algorithm applies a λ_k-averaged operator with Gaussian noise, with possibly randomized, inexact and block-wise updates
- This setup is easy to combine with amplification by iteration and by subsampling

Noisy Fix-point Operators

- We study the general noisy fixed-point iteration

Algorithm 6 Noisy fixed-point iteration

Input: non-expansive operator $R = (R_1, \dots, R_B)$ over $1 \leq B \leq p$ blocks,step sizes $(\lambda_k)_{k\in\mathbb{N}} \in (0,1]$, active blocks $(\rho_k)_{k\in\mathbb{N}} \in \{0,1\}^B$, errors $(e_k)_{k\in\mathbb{N}}$, noise variance $\sigma^2 \geq 0$

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

  for $b = 1, \dots, B$ **do**

    $u_{k+1,b} = u_{k,b} + \rho_{k,b}\lambda_k(R_b(u_k) + e_{k,b} + \eta_{k+1,b} - u_{k,b})$ with $\eta_{k+1,b} \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_p)$

- This general algorithm applies a λ_k-averaged operator with Gaussian noise, with possibly randomized, inexact and block-wise updates
- This setup is easy to combine with amplification by iteration and by subsampling
- We recover DP-SGD with $R(u) = u - \frac{2}{\beta}\nabla f(u; \mathcal{D})$, $B = 1$, $e_k = \frac{2}{\beta}(\nabla f(u_k; \mathcal{D}) - \nabla f(u_k; d_{i_k}))$

General utility analysis

Theorem (Utility guarantees for noisy fixed-point iterations Cyffers2023a)

Assume that R is τ-contractive with fixed point u^*. Let $P[\rho_{k,b} = 1] = q$ for some $q \in (0, 1]$. Then there exists a learning rate $\lambda_k = \lambda \in (0, 1]$ such that the iterates satisfy:

$$\mathbb{E}\left(\|u_{k+1} - u^*\|^2\right) \leqslant \left(1 - \frac{q^2(1-\tau)}{8}\right)^k D + 8\left(\frac{\sqrt{p}\sigma + \zeta}{\sqrt{q}\,(1-\tau)} + \frac{p\sigma^2 + \zeta^2}{q^3(1-\tau)^3}\right) \tag{1}$$

where $D = \|u_0 - u^*\|^2$, p is the dimension of u, and $\mathbb{E}[\|e_k\|^2] \leq \zeta^2$ for some $\zeta \geq 0$.

- The only assumption on R is that it is τ-contractive
- We roughly recover DP-SGD rate for strongly convex objective
- Let's apply it to ADMM

ADMM as fix point for ERM

ADMM can be written as Lions Mercier operator

$$T = \lambda R_{\gamma p_1} R_{\gamma p_2} + (1 - \lambda) I$$

with $R_{\gamma p} = 2\,\mathrm{prox}_{\gamma p} - I$.

The consensus problem fits the general form solved by ADMM algorithms:

$$\begin{array}{ll} \underset{x \in \mathbb{R}^{np},\, z \in \mathbb{R}^p}{\text{minimize}} & \frac{1}{n} \sum_{i=1}^{n} f(x_i; d_i) + r(z) \\ \text{subject to} & x - I_{n(p \times p)} z = 0, \end{array}$$

where each data item d_i has its own parameter $x_i \in \mathbb{R}^p$

Privacy-utility for centralized, federated and decentralized ADMM

Algorithm 7 Private ADMM

Input: initial point z_0, step size $\lambda \in (0, 1]$, privacy noise variance $\sigma^2 \geq 0$, parameter $\gamma > 0$, number of sampled users $1 \leq m \leq n$

for $k = 0$ to $K - 1$ **do**

$\hat{z}_{k+1} = \frac{1}{n} \sum_{i=1}^{n} u_{k,i}$

$z_{k+1} = \text{prox}_{\gamma r}(\hat{z}_{k+1})$

for $i = 1$ to n **do**

$x_{k+1,i} = \text{prox}_{\gamma f_i}(2z_{k+1} - u_{k,i})$

$u_{k+1,i} = u_{k,i} + 2\lambda\big(x_{k+1,i} - z_{k+1} + \frac{1}{2}\eta_{k+1,i}\big)$ with $\eta_{k+1,i} \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_p)$

Return z_K

	Centralized	Federated	Decentralized
Privacy loss $\mathbb{E}(\|u^K - u^*\|^2)$ (in $\mathcal{O}(\cdot)$)	$\frac{\sqrt{p\alpha}L\gamma}{\sqrt{\varepsilon}n(1-\tau)} + \frac{\frac{8\alpha K L^2\gamma^2}{\sigma^2 n^2}}{\frac{p\alpha L^2\gamma^2}{\varepsilon n^2(1-\tau)^3}}$	$\frac{\sqrt{p\alpha}L\gamma}{\sqrt{\varepsilon r}n(1-\tau)} + \frac{\frac{16\alpha K L^2\gamma^2}{\sigma^2 n^2}}{\frac{p\alpha L^2\gamma^2}{\varepsilon r^2 n^2(1-\tau)^3}}$	$\frac{\sqrt{p\alpha}L\gamma}{\sqrt{\varepsilon}n(1-\tau)} + \frac{\frac{8\alpha K_i L^2\gamma^2 \ln n}{\sigma^2 n}}{\frac{p\alpha L^2\gamma^2}{\varepsilon n(1-\tau)^3}}$

Conclusion

- We provide a unifying view of private optimization algorithms by framing them as noisy fixed-point iterations, and prove general utility guarantees
- Our framework can be used to derive and analyze new private algorithms by instantiating our general scheme with particular fixed-point operators
- We illustrate this by designing private ADMM algorithms for centralized and federated learning; in contrast, prior work used ad-hoc algorithmic modifications and customized analysis with many privacy parameters