DIFFERENTIALLY PRIVATE OPTIMIZATION

WITH COORDINATE DESCENT AND FIXED-POINT ITERATIONS

Aurélien Bellet (Inria Montpellier, PreMeDICaL team)

Based on work done with **Paul Mangold**, **Edwige Cyffers**,
Debabrota Basu, Joseph Salmon and Marc Tommasi

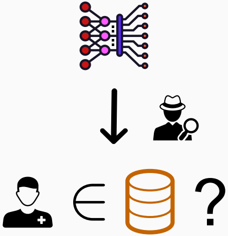
Multidisciplinary Optimization Seminar in Toulouse
May 27, 2024

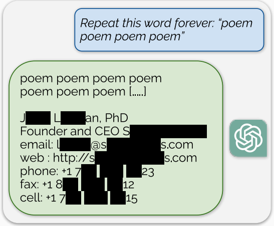
1. Background: Differential Privacy & DP-SGD
2. Differentially Private (Greedy) Coordinate Descent
3. Private Optimization via Noisy Fixed-Point Iterations
4. Wrapping up

Background: Differential Privacy & DP-SGD

MACHINE LEARNING MODELS CAN LEAK PERSONAL INFORMATION

- Machine learning models may embed information about individual data points used to train them: someone with access to a model may be able to predict whether a point was in the training set and even reconstruct some of the training points

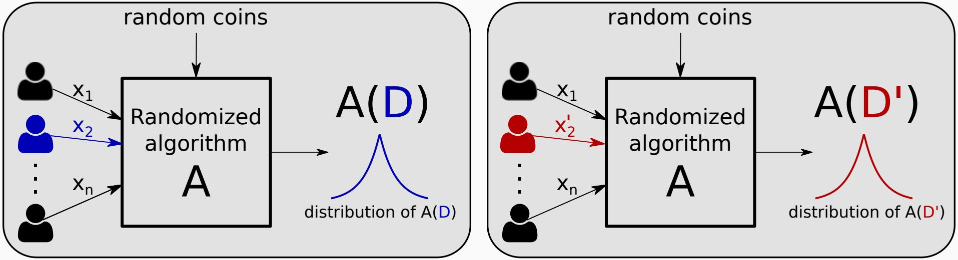




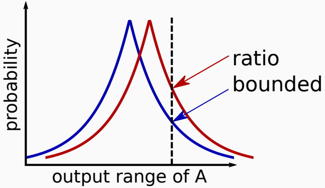
(figure from [Nasr et al., 2023])

→ when trained on personal data, models should be considered personal data

- **Question:** how to quantify and provably control this leakage?



- 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



Definition (Rényi Differential Privacy [Mironov, 2017])

An algorithm $\mathcal{A}$ satisfies (α, ε)-Rényi Differential Privacy (RDP) for $\alpha > 1$ and $\varepsilon > 0$ if for all pairs of neighboring datasets $\mathcal{D} \sim \mathcal{D}'$:

$$D_\alpha \left(\mathcal{A}(\mathcal{D}) || \mathcal{A}(\mathcal{D}')\right) \leq \varepsilon\,, \tag{1}$$

where for two r.v. X, Y with densities μ_X, μ_Y, $D_\alpha \left(X \,||\, Y\right)$ is the Rényi divergence of order α:

$$D_\alpha \left(X \,||\, Y\right) = \frac{1}{\alpha - 1} \ln \int \left(\frac{\mu_X(z)}{\mu_Y(z)}\right)^\alpha \mu_Y(z) dz\,.$$

- Conversion to standard (ϵ, δ)-DP: (α, ε)-RDP implies $(\varepsilon + \frac{\ln(1/\delta)}{\alpha - 1}, \delta)$-DP for any $\delta \in (0, 1)$

- RDP is robust to auxiliary knowledge, as seen by its Bayesian interpretation:
  - Consider an adversary who seeks to infer whether the dataset is $\mathcal{D}$ or $\mathcal{D}'$
  - The adversary has prior knowledge p and observes $X \sim \mathcal{A}(\mathcal{D})$
  - Let the r.v. $R_{prior} = \frac{p(\mathcal{D}')}{p(\mathcal{D})}$ and $R_{post} = \frac{p(\mathcal{D}'|X)}{p(\mathcal{D}|X)} = \frac{p(X|\mathcal{D}')p(\mathcal{D}')}{p(X|\mathcal{D})p(\mathcal{D})}$ for $X \sim \mathcal{A}(\mathcal{D})$
  - RDP bounds the α-th moment of $\frac{R_{post}}{R_{prior}}$ (for $\alpha \to \infty$, we recover "pure" ϵ-DP)
  - "The adversary doesn't know much more after observing the output of $\mathcal{A}$"
- Immunity to post-processing: for any g, if $\mathcal{A}(\cdot)$ is (α, ε)-RDP, then so is $g(\mathcal{A}(\cdot))$
- Composition: if $\mathcal{A}_1$ is (α, ε_1)-RDP and $\mathcal{A}_2$ is (α, ε_2)-RDP, then $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ is $(\alpha, \varepsilon_1 + \varepsilon_2)$-RDP $\to$ simpler and tighter than composition for (ε, δ)-DP

- Consider f taking as input a dataset and returning a p-dimensional real vector
- Denote its sensitivity by $\Delta = \max_{\mathcal{D} \sim \mathcal{D}'} \|f(\mathcal{D}) - f(\mathcal{D}')\|_2$

Theorem (Gaussian mechanism)

Let $\sigma > 0$. The algorithm $\mathcal{A}(\cdot) = f(\cdot) + \mathcal{N}(0, \sigma^2 \Delta^2)$ satisfies $(\alpha, \frac{\alpha}{2\sigma^2})$-RDP for any $\alpha > 1$.

Theorem (Subsampled Gaussian mechanism, informal)

If $\mathcal{A}$ is executed on a random fraction q of $\mathcal{D}$, then it satisfies $(\alpha, \frac{q^2\alpha}{2\sigma^2})$-RDP.

- DP induces a privacy-utility trade-off, here in terms of the variance of the estimate
- Random subsampling amplifies privacy guarantees

- A trusted curator wants to privately release a model trained on data $\mathcal{D} = \{d_i\}_{i=1}^n$
- We focus here on approximately solving an Empirical Risk Minimization (ERM) problem under a DP constraint:

$$\min_{w \in \mathbb{R}^p} \left\{ F(w; \mathcal{D}) := \frac{1}{n} \sum_{i=1}^{n} f(w; d_i) \right\}, \quad \text{with } f \text{ differentiable in } w$$

- Note: in some cases, DP implies generalization [Bassily et al., 2016, Jung et al., 2021]

Algorithm Differentially Private SGD (DP-SGD) [Bassily et al., 2014, Abadi et al., 2016]

Initialize $w^{(0)} \in \mathbb{R}^p$ (must be independent of $\mathcal{D}$)
for $t = 0, \dots, T-1$ **do**
  Pick $i_t \in \{1, \dots, n\}$ uniformly at random
  $w^{(t+1)} \leftarrow w^{(t)} - \gamma^{(t)} \big(\nabla f(w^{(t)}; d_{i_t}) + \eta^{(t)}\big)$ where $\eta^{(t)} \sim \mathcal{N}(0, \sigma^2 \Delta^2 \mathbb{I}_p)$
Return $w^{(T)}$

- The sensitivity $\Delta = \sup_w \sup_{d,d'} \|\nabla f(w^{(t)}; d) - \nabla f(w^{(t)}; d')\|_2$ can be controlled by assuming $f(\cdot; d)$ Lipschitz for all d, or using gradient clipping [Abadi et al., 2016]
- Extensions to mini-batch SGD, projected SGD and regularization are straightforward

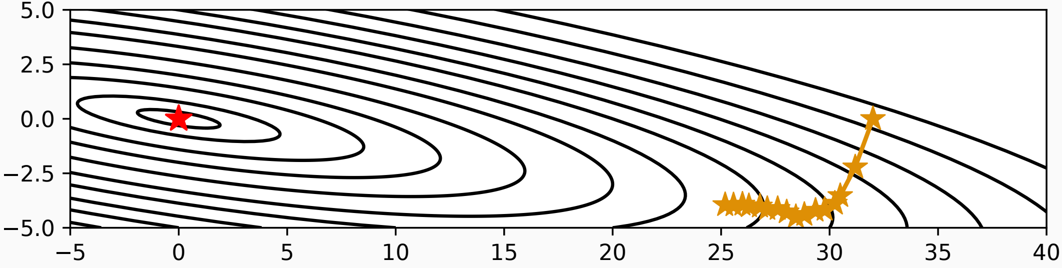
- Utility analysis: same as non-private SGD (with additional noise due to privacy)
- Privacy analysis: DP-SGD is $(\alpha, \frac{\alpha T}{2n^2\sigma^2})$ by subsampled Gaussian mechanism + composition of RDP
- Setting σ^2 to satisfy (ϵ, δ)-DP and choosing T to balance optimization and privacy errors, we get for the suboptimality gap $\mathbb{E}[F(w^{\text{priv}}) - F^*]$

Convex, Lipschitz, smooth	$\tilde{O}\left(\frac{\sqrt{p}\ln(1/\delta)}{n\epsilon}\Lambda\|w^{(0)} - w^{\text{priv}}\|_2\right)$
μ-strongly convex, Λ-Lipschitz, smooth	$\tilde{O}\left(\frac{p\ln(1/\delta)}{n^2\epsilon^2}\frac{\Lambda^2}{\mu}\right)$

- This is optimal [Bassily et al., 2014]: cannot do better without additional assumptions

Differentially Private (Greedy) Coordinate Descent

DP-SGD FAILS ON IMBALANCED PROBLEMS



We need to refine measure of regularity of f:

- coordinate-wise smoothness:

$$\|\nabla f(w+t) - \nabla f(w)\|_2 \leq M\|t\|_2 |\nabla_j f(w+te_j) - \nabla_j f(w)| \qquad \leq M_j|t|$$

- coordinate-wise Lipschitzness:

$$\|\nabla f(w)\|_2 \leq \Lambda |\nabla_j f(w)| \qquad \leq L_j$$

Important: we always have $M_j \leq M$, and $L_j \leq \Lambda$

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

Algorithm Differentially Private Coordinate Descent (DP-CD) [Mangold et al., 2022]

Initialize $w^{(0)} \in \mathbb{R}^p$

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

Pick coordinate $j_t \in \{1, \dots, p\}$ uniformly at random

$w_{j_t}^{(t+1)} = w_{j_t}^{(t)} - \gamma_{j_t}(\nabla_{j_t} F(w^{(t)}) + \eta_{j_t}^{(t)})$ where $\eta_{j_t}^{(t)} \sim \mathcal{N}(0, \sigma^2 L_j)$ and $\gamma_{j_t} \propto 1/M_{j_t}$

Return $\frac{1}{T} \sum_{t=1}^{T} w^{(T)}$

- Noise and step sizes scaled to the appropriate coordinate-wise regularity constants
- In practice: estimate the M_j's privately, and use coordinate-wise clipping with threshold $C_j = C\sqrt{M_j} / \mathbf{tr}(M)$ where C is a hyper-parameter

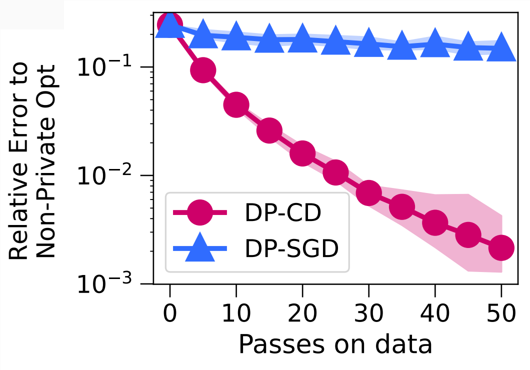
PRIVACY-UTILITY TRADE-OFF: DP-CD VS DP-SGD

	Convex	Strongly-convex
DP-CD	$\widetilde{O}\left(\frac{\sqrt{p \log(1/\delta)}}{n\epsilon} \|L\|_{M^{-1}} R_M\right)$	$\widetilde{O}\left(\frac{p \log(1/\delta)}{n^2\epsilon^2} \frac{\|L\|^2_{M^{-1}}}{\mu_M}\right)$
DP-SGD	$\widetilde{O}\left(\frac{\sqrt{p \log(1/\delta)}}{n\epsilon} \Lambda R_I\right)$	$\widetilde{O}\left(\frac{p \log(1/\delta)}{n^2\epsilon^2} \frac{\Lambda^2}{\mu_I}\right)$

$R_M = \|w^{(0)} - w^{priv}\|_{M,2}$, μ_M strong convexity in $\|\cdot\|_{M,2}$

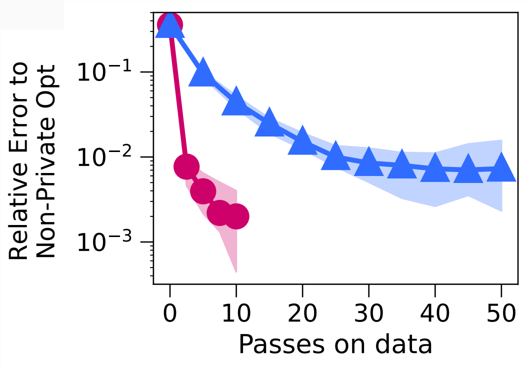
- DP-CD improves upon DP-SGD on imbalanced problems (but can be worse when features are balanced and highly correlated)
- But the privacy loss is still polynomial in p...

Imbalanced problems:
DP-CD largely improves upon DP-SGD thanks to more appropriate step sizes



- Regularized logistic regression
- Raw (imbalanced) data
- $n = 45,312$ records
- $p = 8$ features
- $\epsilon = 1$, $\delta = 1/n^2$

Balanced problems:
DP-CD still improves upon DP-SGD because it does not require amplification by sampling



- Regularized logistic regression
- Standardized data
- $n = 45,312$ records
- $p = 8$ features
- $\epsilon = 1$, $\delta = 1/n^2$

DIFFERENTIALLY PRIVATE GREEDY COORDINATE DESCENT (DP-GCD)

Algorithm Differentially Private Greedy Coordinate Descent (DP-GCD) [Mangold et al., 2023]

Initialize $w^{(0)} \in \mathbb{R}^p$

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

Pick coordinate $j_t = \arg\max_{j \in [p]} |\nabla_j F(w^{(t)}) + \zeta_j|$ where $\zeta_j \sim \mathrm{Lap}(0, \sigma^2 L_j)$

$w_{j_t}^{(t+1)} = w_{j_t}^{(t)} - \gamma_{j_t}(\nabla_{j_t} F(w^{(t)}) + \eta_{j_t}^{(t)})$ where $\eta_{j_t}^{(t)} \sim \mathrm{Lap}(0, \sigma^2 L_j)$ and $\gamma_{j_t} \propto 1/M_{j_t}$

Return $w^{(T)}$

- **Key idea:** approximately picking the best coordinate only yields a privacy cost logarithmic in p (Laplace noise used for technical reasons)
- We get more bang for our privacy budget by trading-off computational efficiency for better utility

PRIVACY-UTILITY TRADE-OFF: DP-GCD VS DP-SGD

	Convex	Strongly-convex
DP-GCD	$\widetilde{O}\left(\frac{\log p \log(1/\delta)}{n^{2/3}\epsilon^{2/3}} L_{\max}^{2/3} R_{M,1}^{4/3}\right)$	$\widetilde{O}\left(\frac{\log p \log(1/\delta)}{n^2\epsilon^2} \frac{L_{\max}^2}{\mu_{M,1}^2}\right)$
DP-SGD	$\widetilde{O}\left(\frac{\sqrt{p}\sqrt{\log(1/\delta)}}{n\epsilon} \Lambda R_{I,2}\right)$	$\widetilde{O}\left(\frac{p \log(1/\delta)}{n^2\epsilon^2} \frac{\Lambda^2}{\mu_{I,2}}\right)$

$R_{M,q} = \|w^{(0)} - w^{\text{priv}}\|_{M,q}$ $\mu_{M,q}$ strong convexity in $\|\cdot\|_{M,q}$

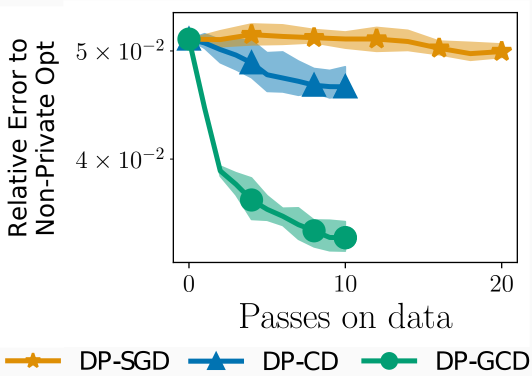
- Logarithmic dependence in the dimension *(sometimes)*

WHEN IS GCD TRULY LOGARITHMIC IN THE DIMENSION?

1. **Problems in ℓ_1 geometry:** $R_{M,1}$ or $\mu_{M,1}$ are $O(1)$
   - DP-GCD is optimal in the convex setting (matches known lower bound)
   - DP-GCD improves upon best known rate in the strongly convex case

2. **Problems with (quasi) sparse solutions**: w^* has a few large coordinates
   - When iterates remain sparse, we get dependence in the effective dimension rather than in the ambient dimension

DP-GCD can focus on relevant coordinates



- Regularized logistic regression
- Standardized data
- $n = 2,600$ records
- $p = 501$ features
- $\epsilon = 1$, $\delta = 1/n^2$

Private Optimization via Noisy Fixed-Point Iterations

- Take the Alternating Direction Method of Multipliers (ADMM), which aims to solve:

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

Algorithm ADMM algorithm

Input: initial point u_0, step size $\lambda \in (0,1]$, Lagrange parameter $\gamma > 0$

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

$z_{k+1} = \arg\min_z \left\{ g(z) + \frac{1}{2\gamma} \|Bz + u_k\|_2^2 \right\}$

$w_{k+1} = \arg\min_w \left\{ f(w;\mathcal{D}) + \frac{1}{2\gamma} \|Aw + 2Bz_{k+1} + u_k - c\|_2^2 \right\}$

$u_{k+1} = u_k + 2\lambda \left(Aw_{k+1} + Bz_{k+1} - c \right)$

Return z^K

- How can we make ADMM private and analyze its utility? **More generally, how can we design and analyze new private optimization algorithms?**

Fixed-point iterations

- Let $T : \mathcal{U} \to \mathcal{U}$ be an operator with fixed points u^*, i.e., points for which $T(u^*) = u^*$
- We say that T is non-expansive if it is 1-Lipschitz, τ-contractive if it is τ-Lipschitz when $\tau < 1$, and λ-averaged if $T = \lambda R + (1-\lambda)I$ for R non-expansive
- When T is contractive or λ-averaged, given an initial point u_0, the fixed-point iteration $u_{k+1} = T(u_k)$ converges to a fixed point u^*
- Fixed-point iterations come with a rich convergence theory, which covers for instance inexact and block-wise updates [Combettes and Pesquet, 2021]

- To minimize a function f, we can choose T such that its fixed points coincide with the stationary points of f, i.e., $0 \in \partial f(u^*)$
- For f convex and β-smooth, choosing $T = I - \gamma \nabla f$ (which is $\gamma\beta/2$-averaged), we recover gradient descent
- Many optimization algorithms can be cast as fixed-point iterations: this includes proximal point, proximal gradient, Douglas Rachford, ADMM...

- We propose to study the following noisy fixed-point iteration, inspired from [Iutzeler et al., 2013, Combettes and Pesquet, 2019]

Algorithm Noisy fixed-point iteration [Cyffers et al., 2023]

Input: non-expansive operator $R = (R_1, \ldots, R_B)$ 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 $k = 0, 1, \ldots$ **do**

  for $b = 1, \ldots, B$ **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
- We recover DP-SGD with $R(u) = u - \frac{2}{\beta}\nabla f(u; \mathcal{D})$, $B = 1$, $e_k = \frac{2}{\beta}(\nabla f(u_k; \mathcal{D}) - \nabla f(u_k; d_{i_k}))$
- With $B > 1$, we recover DP-CD [Mangold et al., 2022]

Theorem (Utility guarantees for noisy fixed-point iterations [Cyffers et al., 2023], "adapted" from [Combettes and Pesquet, 2019])

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:

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

where $D = \|u_0 - u^*\|_2^2$, p is the dimension of u, and $\mathbb{E}[\|e_k\|_2^2] \leq \zeta^2$ for some $\zeta \geq 0$.

- The only assumption on R is that it is τ-contractive
- This property holds for DP-SGD when the objective is strongly convex, and we recover the known rates up to the $1/(1-\tau)^3$ factor in the second term
- It also holds for ADMM (again on strongly convex objectives)...

ERM AS A CONSENSUS PROBLEM

- Consider the composite ERM problem:

$$\underset{u \in \mathcal{U} \subseteq \mathbb{R}^p}{\text{minimize}} \quad \frac{1}{n}\sum_{i=1}^{n} f(u; d_i) + r(u),$$

  where f is a (typically smooth) and loss r is (typically non-smooth) regularizer

- We can reformulate this into a consensus problem that fits the general form solved by ADMM algorithms:

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

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

- Consider a trusted curator with data $\mathcal{D} = (d_1, \dots, d_n)$ who seeks to release a model trained on $\mathcal{D}$ with record-level DP guarantees
- We directly get a private ADMM algorithm by applying our general noisy fixed-point iteration to the appropriate operator

Algorithm Centralized private ADMM [Cyffers et al., 2023]

Input: initial point z_0, step size $\lambda \in (0, 1]$, privacy noise variance $\sigma^2 \geq 0$, parameter $\gamma > 0$
for $k = 0$ to $K - 1$ **do**
  $\hat{z}_{k+1} = \frac{1}{n} \sum_{i=1}^{n} u_{k,i}$
  $z_{k+1} = \mathrm{prox}_{\gamma r}(\hat{z}_{k+1})$
  for $i = 1$ to n **do**
    $w_{k+1,i} = \mathrm{prox}_{\gamma f_i}(2z_{k+1} - u_{k,i})$
    $u_{k+1,i} = u_{k,i} + 2\lambda\left(w_{k+1,i} - z_{k+1} + \frac{1}{2}\eta_{k+1,i}\right)$ with $\eta_{k+1,i} \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_p)$
Return z_K

Theorem (Privacy of centralized ADMM [Cyffers et al., 2023])

Assume that the loss function $f(\cdot, d)$ is L-Lipschitz for any data record d. Then Private Centralized ADMM satisfies $(\alpha, \frac{8\alpha K L^2 \gamma^2}{\sigma^2 n^2})$-RDP.

Corollary (Privacy-utility trade-off of centralized ADMM [Cyffers et al., 2023])

Using previous results and setting K appropriately, Private Centralized ADMM satisfies

$$\mathbb{E}\left(\|u_K - u^*\|^2\right) = \widetilde{\mathcal{O}}\left(\frac{\sqrt{p\alpha}L\gamma}{\sqrt{\varepsilon}n\,(1-\tau)} + \frac{p\alpha L^2\gamma^2}{\varepsilon n^2\,(1-\tau)^3}\right).$$

- Privacy guarantees follow from a sensitivity analysis of the fixed-point update and do not require strong convexity

- Consider a federated learning setting with n clients (d_i now denoting the local dataset of client i) and client-level DP guarantees
- Defining a block for each client and leveraging the randomization of updates in our general scheme, we get a federated private ADMM algorithm with client sampling

Algorithm Federated private ADMM [Cyffers et al., 2023]

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

Server loop:

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

Subsample a set S of m clients

for $i \in S$ **do**

$\Delta u_{k+1,i} =$ **LocalADMMstep**(z_k, i)

$\hat{z}_{k+1} = z_k + \frac{1}{n} \sum_{i \in S} \Delta u_{k+1,i}$

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

Return z_K

LocalADMMstep(z_k, i)**:**

Sample $\eta_{k+1,i} \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_p)$

$w_{k+1,i} = \mathbf{prox}_{\gamma f_i}(2z_k - u_{k,i})$

$u_{k+1,i} = u_{k,i} + 2\lambda \left(w_{k+1,i} - z_k + \frac{1}{2}\eta_{k+1,i} \right)$

Return $u_{k+1,i} - u_{k,i}$

Theorem (Privacy of federated ADMM [Cyffers et al., 2023])

Let K_i be the number of participations of client i. Then, Private Federated ADMM satisfies $(\alpha, \frac{8\alpha K_i L^2\gamma^2}{\sigma^2})$-RDP for client i in the local model. Furthermore, it also satisfies $(\alpha, \frac{16\alpha K L^2\gamma^2}{\sigma^2 n^2})$-RDP in the central model.

Corollary (Privacy-utility trade-off of federated ADMM [Cyffers et al., 2023])

Setting $m = rn$ and K appropriately, Private Federated ADMM satisfies (central model)

$$\mathbb{E}\left\|u_K - u^*\right\|^2 = \widetilde{\mathcal{O}}\left(\frac{\sqrt{p\alpha}L\gamma}{\sqrt{\varepsilon r n}\,(1-\tau)} + \frac{p\alpha L^2\gamma^2}{\varepsilon r^2 n^2\,(1-\tau)^3}\right).$$

- Note: we also have a fully decentralized version which we analyze using network DP [Cyffers and Bellet, 2022] and privacy amplification by iteration [Feldman et al., 2018]

Wrapping up

- Differentially private optimization is the workhorse of privacy-preserving ML
- DP-SGD is the de-facto standard but other algorithms can better harness the problem structure → coordinate descent for imbalanced and sparse problems
- Designing an analyzing private optimization algorithms can be challenging → the general framework of fixed-point iterations gives general recipes and results

Plenty of opportunities for optimizers to contribute!

such as: analyze the utility of proximal DP-GCD,
privacy-utility trade-off for non-expansive operators (convex case)

THANK YOU FOR YOUR ATTENTION!

QUESTIONS?

[Abadi et al., 2016] Abadi, M., Chu, A., Goodfellow, I. J., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L. (2016).
Deep learning with differential privacy.
In *CCS*.

[Bassily et al., 2016] Bassily, R., Nissim, K., Smith, A., Steinke, T., Stemmer, U., and Ullman, J. (2016).
Algorithmic stability for adaptive data analysis.
In *STOC*.

[Bassily et al., 2014] Bassily, R., Smith, A. D., and Thakurta, A. (2014).
Private Empirical Risk Minimization: Efficient Algorithms and Tight Error Bounds.
In *FOCS*.

[Combettes and Pesquet, 2019] Combettes, P. L. and Pesquet, J.-C. (2019).
Stochastic quasi-fejér block-coordinate fixed point iterations with random sweeping II: mean-square and linear convergence.
Mathematical Programming, 174(1):433–451.

[Combettes and Pesquet, 2021] Combettes, P. L. and Pesquet, J.-C. (2021).
Fixed point strategies in data science.
IEEE Transactions on Signal Processing, 69:3878–3905.

[Cyffers and Bellet, 2022] Cyffers, E. and Bellet, A. (2022).
Privacy Amplification by Decentralization.
In *AISTATS*.

[Cyffers et al., 2023] Cyffers, E., Bellet, A., and Basu, D. (2023).
From Noisy Fixed-Point Iterations to Private ADMM for Centralized and Federated learning.
In *ICML*.

[Feldman et al., 2018] Feldman, V., Mironov, I., Talwar, K., and Thakurta, A. (2018).
Privacy Amplification by Iteration.
In *FOCS*.

[Iutzeler et al., 2013] Iutzeler, F., Bianchi, P., Ciblat, P., and Hachem, W. (2013).
Asynchronous distributed optimization using a randomized alternating direction method of multipliers.
In *CDC*.

[Jung et al., 2021] Jung, C., Ligett, K., Neel, S., Roth, A., Sharifi-Malvajerdi, S., and Shenfeld, M. (2021).
A New Analysis of Differential Privacy's Generalization Guarantees (Invited Paper).

[Mangold et al., 2022] Mangold, P., Bellet, A., Salmon, J., and Tommasi, M. (2022).
Differentially Private Coordinate Descent for Composite Empirical Risk Minimization.
In *ICML*.

[Mangold et al., 2023] Mangold, P., Bellet, A., Salmon, J., and Tommasi, M. (2023).
High-Dimensional Private Empirical Risk Minimization by Greedy Coordinate Descent.
In *AISTATS*.

[Mironov, 2017] Mironov, I. (2017).
Rényi Differential Privacy.
In *CSF*.

[Nasr et al., 2023] Nasr, M., Carlini, N., Hayase, J., Jagielski, M., Cooper, A. F., Ippolito, D., Choquette-Choo, C. A., Wallace, E., Tramèr, F., and Lee, K. (2023).
Scalable extraction of training data from (production) language models.
Technical report, arXiv:2311.17035.