Privacy Amplification by Iteration for ADMM with (Strongly) Convex Objective Functions

T-H. Hubert Chan, Hao Xie, Mengshi Zhao

The University of Hong Kong

Abstract

We examine a private ADMM variant for (strongly) convex objectives which is a primal-dual iterative method. Each iteration has a user with a private function used to update the primal variable, masked by Gaussian noise for local privacy, without directly adding noise to the dual variable. Privacy amplification by iteration explores if noises from later iterations can enhance the privacy guarantee when releasing final variables after the last iteration.

Our main result is that the privacy guarantee for the gradient ADMM variant can be amplified proportionally to the number of iterations. For strongly convex objective functions, this amplification exponentially increases with the number of iterations. These amplification results align with the previously studied special case of stochastic gradient descent.

Problem

- **Objective Function**
  - A distribution $\mathcal{D}$.
  - Objective function $\mathfrak{f} = E_{f\leftarrow\mathcal{D}}(f)$.
  - Distributed Setting: T users, each of the users samples $f_1, f_2, \ldots, f_T$.
- **Target:**

$$\min_{x,y} \quad \mathfrak{f}(x) + g(y)$$
$$\text{s.t.} \quad Ax + By = c \in \mathcal{R}^m$$
$$x \in \mathcal{R}^n, y \in \mathcal{R}^\ell$$

ADMM

- Optimize over x:

$$x_{t+1} = \arg\min \mathfrak{f}(x_t) + \langle \nabla \mathfrak{f}_{t+1}(x_t), x - x_t\rangle + g(y_t) + \frac{1}{2\eta}\|x - x_t\|^2 + \mathcal{L}(x, y_t, \lambda_t)$$
$$\tilde{x}_{t+1} = x_{t+1} + \mathcal{N}(0, \sigma^2 I)$$

- Optimize over y:

$$y_{t+1} = \arg\min \mathfrak{f}(\tilde{x}_{t+1}) + g(y) + \mathcal{L}(\tilde{x}_{t+1}, y, \lambda_t)$$

- Update λ:

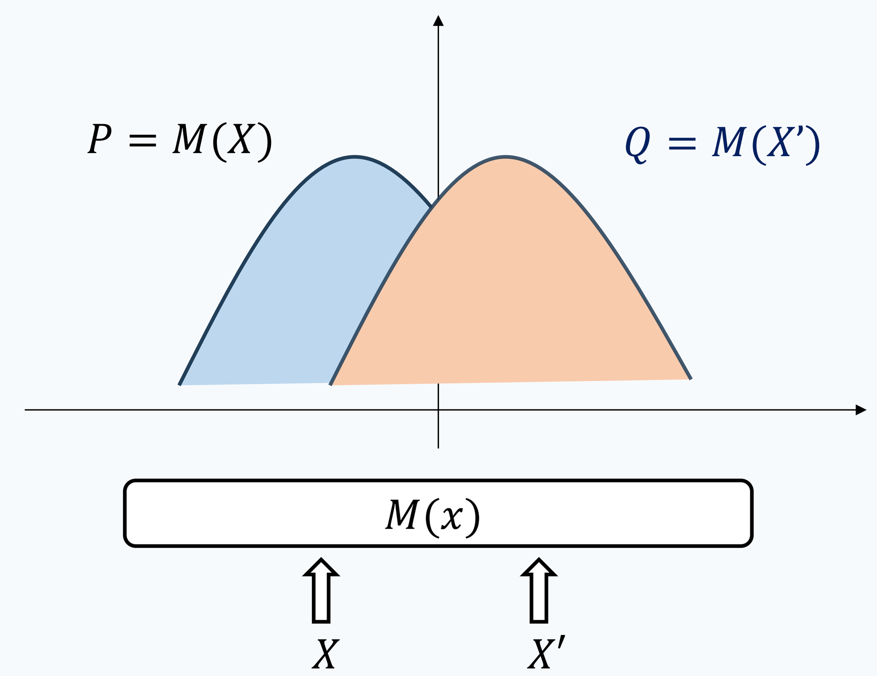
$$\lambda_{t+1} = \lambda_t - \beta(A\tilde{x}_{t+1} + By_{t+1} + c)$$
$$\mathcal{L}(x, y, \lambda) = \frac{\beta}{2}\|Ax + By - c\|^2 - \langle \lambda, Ax + By - c\rangle$$

Our Contribution

We show that from the perspective of the user from the first iteration, the final variables after T noisy ADMM iterations achieve privacy amplification in the sense that the Rényi divergence is proportional to $\frac{1}{T}$.

For strongly convex objective functions, the privacy amplification is improved to $\frac{L^T}{T}$ for some $0 < L < 1$.

Differential Privacy



- The distribution of the output P on input X has nearly the similar distribution as the output Q on input X'.
- **Rényi divergence** of order $\alpha > 1$:

$$\mathrm{D}_\alpha(P\|Q) := \frac{1}{\alpha-1}\ln E_{x\leftarrow Q}\left(\frac{P(x)}{Q(x)}\right)^\alpha.$$

- **(α, ϵ) *Rényi* differential privacy:** For all neighbouring inputs X and X',

$$\mathrm{D}_\alpha(M(X)\|M(X')) \le \epsilon.$$

- **Divergence for Gaussian noise:**

$$\mathrm{D}_\alpha(\mathcal{N}(x, \sigma^2 I_n)\|\mathcal{N}(x', \sigma^2 I_n)) = \frac{\|x - x'\|^2}{2\alpha\sigma^2}.$$

- Smaller *Rényi*-divergence means more similar.

Privacy Amplification By Iteration

Privacy Amplification loosely refers to the improvement of privacy analysis for a user using extra sources of randomness other than the noise used for achieving its local privacy.

Privacy amplification has been proposed to analyze an iterative procedure in which some noise is sampled in each iteration to achieve local privacy for the user in that iteration. The improved privacy analysis is from the perspective of the user from the **first** iteration.

Suppose the iteration operators $K_1, K_2, \ldots, K_T$ satisfies the following conditions.

- Non-Expansion. For any K_i and input z and z', the Wasserstein distance $W(K(z), K(z')) \le \| z - z' \|$.
- One-step Privacy. There exists a constant $C > 0$ such that for any K_i and input z and z', it holds that:

$$\mathrm{D}_\alpha(K_i(z) \parallel K_i(z')) \le C \parallel z - z' \parallel^2.$$

Then there exists a constant for any input z and z',

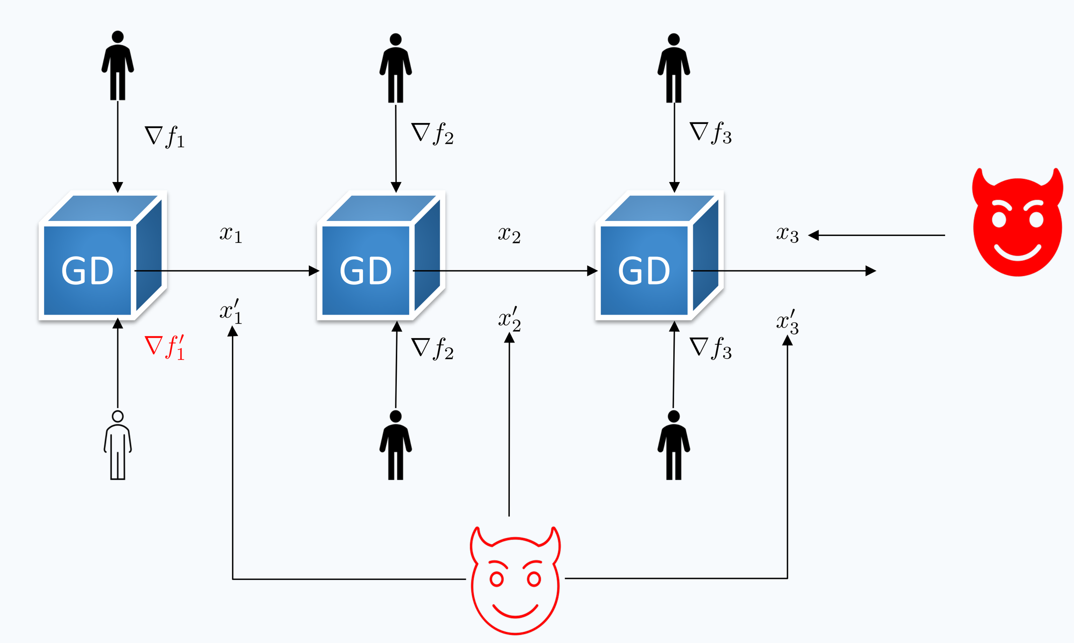
$$\mathrm{D}_\alpha(K_T K_{T-1} \ldots K_1(z) \parallel K_T K_{T-1} \ldots K_1(z')) \le \frac{C}{T} \parallel z - z' \parallel^2.$$

Warm-up: Noisy Stochastic Gradient Descent

- **Problem**
  - A distribution $\mathcal{D}$.
  - Objective function $\mathfrak{f} = E_{f\leftarrow\mathcal{D}}(f)$.
  - Distributed Setting: T users, each of the users samples $f_1, f_2, \ldots, f_T$.
  - Target: $\min \mathfrak{f}(x)$.
- Noisy Stochastic Gradient Descent

$$x_{t+1} \leftarrow \Pi_K(x_t - \eta(\nabla f_{t+1}(x_t) + N(0, \sigma^2 I_d)))$$

- **Two types of randomness:**
  - Sampling: for convergence of the algorithm.
  - Gaussian noise: For achieving local DP.
- **Two types of adversaries:**
  - Observe all intermediate x_t. (Local DP)
  - Only obverse the final output x_T. (Privacy Amplification)



- Result for noisy stochastic gradient descent:
  - Two scenarios.
  - Same input X_0.
  - T iterations of noisy private GD.
  - In the two scenarios, the first functions f_1 and f_1' are different. Other functions are the same.
  - Then

$$D_\alpha(X_T\|X_T') \le O(\frac{1}{T}).$$

Our Methods

- **Two Challenges:** In order to apply the method, those are required.
- Non-expansion for ADMM? One iteration without noise is non-expansion.

$$\|(x_{t+1} - x_{t+1}', \lambda_{t+1} - \lambda_{t+1}')\|^2 \le \|(x_t - x_t', \lambda_t - \lambda_t')\|^2.$$

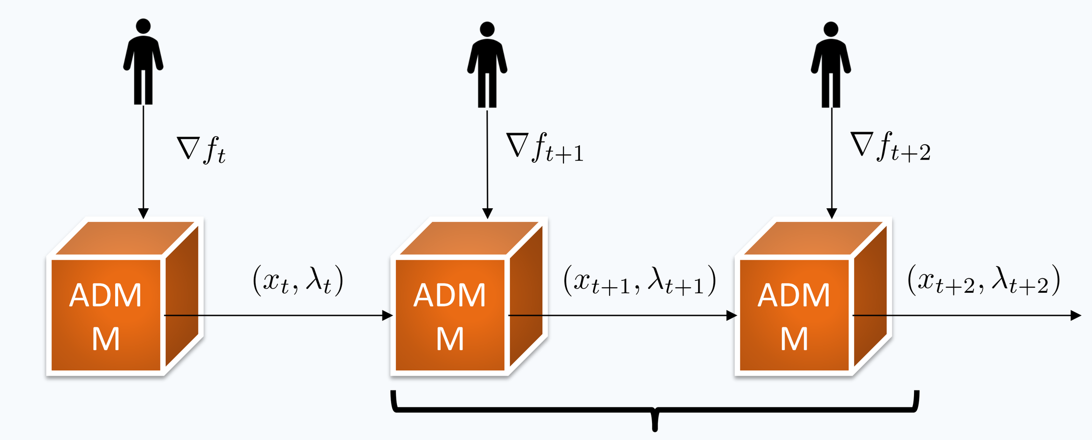
- Usual Norm may not be non-expansion, this is because one iteration signifies a transition in the (x,λ)-space.
- We resolve the problem by using a customize norm:

$$\|(x, \lambda)\|_*^2 := \|x\|^2 + \frac{\eta}{\beta}\cdot\|\lambda - \beta A x\|^2.$$

- Given two different input (x_t, λ_t) and (x_t', λ_t') ,

$$\|(x_{t+1} - x_{t+1}', \lambda_{t+1} - \lambda_{t+1}')\|_*^2 \le \|(x_t - x_t', \lambda_t - \lambda_t')\|_*^2.$$

- One-step Privacy for ADMM? There is an upper bound for the divergence for noisy stochastic ADMM.
- If we only consider one iteration, then noises added on x_t and λ_t are not independent, making the divergence be infinity. So we consider two iterations of ADMM.



$$\mathrm{D}_\alpha\left((\tilde{x}_{t+2}, \lambda_{t+2})\|(\tilde{x}_{t+2}', \lambda_{t+2}')\right) \le C_K \|(\tilde{x}_t - \tilde{x}_t', \lambda_t - \lambda_t')\|_*^2.$$

Results

- Main theorem:
  - Two scenarios.
  - Same input (X_0, λ_0).
  - T iterations of noisy stochastic ADMM.
  - In the two scenarios, the first functions f_1 and f_1' are different. Other functions are the same.
  - Then

$$\mathrm{D}_\alpha((\tilde{x}_T, \lambda_T)\|(\tilde{x}_T', \lambda_T')) \le O(\tfrac{1}{T}).$$

- For strongly convex function: there exits an $L \in (0,1)$, such that

$$\mathrm{D}_\alpha((\tilde{x}_T, \lambda_T)\|(\tilde{x}_T', \lambda_T')) \le O(\tfrac{L^T}{T}).$$